Aspects of Harmonic Analysis and Representation Theory

Jean Gallier and Jocelyn Quaintance
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: jean@seas.upenn.edu

April 13, 2023

Preface

The question that motivated writing this book is:

What is the Fourier transform?

We were quite surprised by how involved the answer is, and how much mathematics is needed to answer it, from measure theory, integration theory, some functional analysis, to some representation theory.

First we should be a little more precise about our question. We should ask two questions:

(1) What is the *input domain* of the Fourier transform?

(2) What is the *output domain* of the Fourier transform?

The answer to (1) is that the domain of the Fourier transform, denoted by $\mathcal{F}$, is a set of functions on a *group* G. Now in order for the Fourier transform to be useful, it should behave well with respect to *convolution* (denoted $f * g$) on the set of functions on G, which implies that these functions should be *integrable*.

This leads to the first subtopic, which is *what is integration on a group*? The technical answer involves the *Haar measure* on a locally compact group. Thus, any serious effort to understand what the Fourier transform is entails learning a certain amount of measure theory and integration theory, passing through versions of the Radon–Riesz theorem relating Radon functionals and Borel measures, and culminating with the construction of the Haar measure. The two candidates for the domain of the Fourier transform are the spaces $\mathrm{L}^1(G)$ and $\mathrm{L}^2(G)$. Unfortunately, convolution and the Fourier transform are not necessarily defined for functions in $\mathrm{L}^2(G)$, so the domain of the Fourier transform is $\mathrm{L}^1(G)$. Then the equation $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$ holds, as desired. If G is a compact group, $\mathrm{L}^2(G)$ is a suitable (and better) domain.

The answer to Question (2) is more complicated, and depends heavily on whether the group G is commutative or not. The answer is much simpler if G is commutative. In both cases, the output domain of the Fourier transform should be a set of functions from a space Y to a space Z.

If G is commutative, then we can pick $Z = \mathbb{C}$. However, the space Y is rarely equal to G (except when $G = \mathbb{R}$). It turns out that a good theory (which means that it covers all cases already known) is obtained by picking Y to be the group $\widehat{G}$, the *Pontrjagin dual* of G, which consists of the *characters* of the group G. A character of G is a continuous homomorphism $\chi\colon G \to \mathbf{U}(1)$ from G to the group of complex numbers of absolute value 1. For any function $f \in \mathrm{L}^1(G)$, the Fourier transform $\mathcal{F}(f)$ of f is then a function

$$\mathcal{F}(f)\colon \widehat{G} \to \mathbb{C}.$$

In general, $\widehat{G}$ is completely different from G, and this creates problems. For the familiar cases, $G = \mathbb{T} = \mathbf{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\}$, $G = \mathbb{Z}$, $G = \mathbb{R}$, and $G = \mathbb{Z}/n\mathbb{Z}$, the characters are well known, namely $\widehat{\mathbf{T}} = \mathbb{Z}$, $\widehat{\mathbb{Z}} = \mathbf{T}$, $\widehat{\mathbb{R}} = \mathbb{R}$, and $\widehat{\mathbb{Z}/n\mathbb{Z}} = \mathbb{Z}/n\mathbb{Z}$. The case $G = \mathbb{Z}/n\mathbb{Z}$ corresponds to the *discrete Fourier transform*.

For the groups listed above, we know that under some suitable restriction, we have *Fourier inversion*, which means that there is some transform $\overline{\mathcal{F}}$ (called *Fourier cotransform*) such that

$$f = \overline{\mathcal{F}}(\mathcal{F}(f)). \tag{$*$}$$

We have to be a bit careful because the domain of $\overline{\mathcal{F}}$ is $\mathrm{L}^1(\widehat{G})$, and not $\mathrm{L}^1(G)$, are they are usually very different beause in general G and $\widehat{G}$ are *not* isomorphic. Then (assuming that it makes sense), $\overline{\mathcal{F}}(\mathcal{F}(f))$ is a function with domain $\widehat{\widehat{G}}$, so there seems no hope, except in very special cases such as $G = \mathbb{R}$, that $(*)$ could hold. Fortunately, *Pontrjagin duality* asserts that G and $\widehat{\widehat{G}}$ are isomorphic, so $(*)$ holds (under suitable conditions) in the form

$$f = \overline{\mathcal{F}}(\mathcal{F}(f)) \circ \eta,$$

where $\eta\colon G \to \widehat{\widehat{G}}$ is a canonical isomorphism.

If G is a commutative abelian group, there is a beautiful and well understood theory of the Fourier transform based on results of Gelfand, Pontrjagin, and André Weil. In particular, even though the Fourier transform is not defined on $\mathrm{L}^2(G)$ in general, for any function $f \in \mathrm{L}^1(G) \cap \mathrm{L}^2(G)$, we have $\mathcal{F}(f) \in \mathrm{L}^2(\widehat{G})$, and by Plancherel's theorem, the Fourier transform extends in a unique way to an isometric isomorphism between $\mathrm{L}^2(G)$ and $\mathrm{L}^2(\widehat{G})$ (see Section 10.8). Furthermore, if we identify G and $\widehat{\widehat{G}}$ by Pontrjagin duality, then $\mathcal{F}$ and $\overline{\mathcal{F}}$ are mutual inverses (see Section 10.9).

If G is *not* commutative, things are a lot tougher. Characters no longer provide a good input domain, and instead one has to turn to *unitary representations*. A unitary representation is a homomorphism $U\colon G \to \mathbf{U}(H)$ satisfying a certain continuity property, where $\mathbf{U}(H)$ is the group of unitary operators on the Hilbert space H. Then $\widehat{G}$ is the set of equivalence classes of irreducible unitary representations of G, but it is no longer a group.

If G is compact, an important theorem due to Peter and Weyl gives a nice decomposition of $\mathrm{L}^2(G)$ as a Hilbert sum of finite-dimensional matrix algebras corresponding to the

irreducible unitary representations of G (see Theorem 13.2 and Theorem 13.6). As a consequence, there is good notion of Fourier transform, such that the Fourier transform $\mathcal{F}(f)$ is a function with domain $\widehat{G}$, but its output domain is no longer $\mathbb{C}$. Instead, it is a finite-dimensional hermitian space depending on the irreducible representation given as input (see Section 13.4). In general, it is very difficult to find the irreducible representations of a compact group, so this Fourier transform does not seem to be very useful in practice.

If the compact group is a *Lie group*, then the whole machinery of Lie algebras and Lie groups developed by Élie Cartan and Hermann Weyl involving weights and roots becomes available. In particular, if G is a connected *semisimple Lie group*, the finite-dimensional irreducible representations are determined by highest weights. There is a beautiful and extensive theory of representations of semisimple Lie groups, and many books have been written on the subject; see the end of Section 12.7 for some classical references.

A way to deal with noncommutativity due to Gelfand, is to work with pairs (G, K), where K is a compact subgroup of G. Then, instead of working with functions on G, which is "too big," we work with functions on the homogeneous space G/K, the space of left cosets. Then, under certain assumptions on G and K, which makes (G, K) a *Gelfand pair*, it is possible to consider a commutative algebra of functions on the set of double cosets KsK $(s \in G)$, so that some results from the commutative theory can be used (see Chapter 17). The domain of the Fourier transform is a set of functions called *spherical functions*, and this set happens to be homeomorphic to the set of characters on the commutative algebra mentioned above. There is a very nice theory of the Fourier transform and its inverse (see Section 17.7), but how useful it is in practice remains to be seen.

Acknowledgement: Many thanks to the participants of the "underground" Tuesday meetings, Christine Allen-Blanchette, Carlos Esteves, Stephen Phillips, and João Sedoc, for catching mistakes and for many helpful comments. We also thank Kostas Daniilidis for being a source of inspiration. Our debt to J. Dieudonné, G. Folland, E. Hewitt and K.A. Ross, A. Knapp, S. Lang, A.A. Kirillov, Laurent Schwartz, E. Stein and R. Shakarchi, and W. Rudin, is enormous. Every result in this manuscript is found in one form or another in their seminal books.

Contents

Chapter 1

Introduction

The main topic of this book is the Fourier transform and Fourier series, both understood in a broad sense.

Historically, trigonometric series were first used to solve equations arising in physics, such as the wave equation or the heat equation. D'Alembert used trigonometric series (1747) to solve the equation of a vibrating string, elaborated by Euler a year later, and then solved in a different way essentially using Fourier series by D. Bernoulli (1753). However it was Fourier who introduced and developed Fourier series in order to solve the heat equation, in a sequence of works on heat diffusion, starting in 1807, and culminating with his famous book, *Théorie analytique de la chaleur*, published in 1822.

Originally, the theory of Fourier series is meant to deal with periodic functions on the circle $\mathbb{T} = \mathbf{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\}$, say functions with period 2π. Remarkably the theory of Fourier series is captured by the following two equations:

$$f(\theta) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta}. \tag{1}$$

$$c_m = \int_{-\pi}^{\pi} f(\theta) e^{-im\theta} \, \frac{d\theta}{2\pi}. \tag{2}$$

Equation (1) involves a series, and Equation (2) involves an integral. There are two ways of interpreting these equations.

The first way consists of starting with a convergent series as given by the right-hand side of (1) (of course $c_n \in \mathbb{C}$), and to ask what kind of function is obtained. A second question is the following: Are the coefficients in (1) computable in terms of the formulae given by (2)?

The second way is to start with a periodic function f, apply Equation (2) to obtain the c_m, called *Fourier coefficients*, and then to consider Equation (1). Does the series $\sum_{m \in \mathbb{Z}} c_m e^{im\theta}$ (called *Fourier series*) converge at all? Does it converge to f?

Observe that the expression $f(\theta) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta}$ may be interpreted as a countably infinite superposition of elementary periodic functions (the harmonics), intuitively representing

simple wave functions, the functions $\theta \mapsto e^{im\theta}$. We can think of m as the frequency of this wave function.

The above questions were first considered by Fourier. Fourier boldly claimed that *every* function can be represented by a Fourier series. Of course, this is false, and for sereval reasons. First, one needs to define what is an integrable function. Second, it depends on the kind of convergence that are we dealing with. Remarkably, Fourier was almost right, because for every function f in $\mathrm{L}^2(\mathbb{T})$, a famous and deep theorem of Carleson states that its Fourier series converges to f almost everywhere in the L^2-norm.

Given a periodic function f, the problem of determining when f can be reconstructed as the Fourier series (Equation (1)) given by its Fourier coefficients c_m (Equation (2)) is called the problem of *Fourier inversion*. To discuss this problem, it is useful to adopt a more general point of view of the correspondence between functions and Fourier coefficients, and Fourier coefficients and Fourier series.

Given a function $f \in \mathrm{L}^1(\mathbb{T})$, Equation (2) yields the $\mathbb{Z}$-indexed sequence $(c_m)_{m\in\mathbb{Z}}$ of Fourier coefficients of f, with

$$c_m = \int_{-\pi}^{\pi} f(\theta)e^{-im\theta}\,\frac{d\theta}{2\pi},$$

which we call the *Fourier transform* of f, and denote by $\widehat{f}$, or $\mathcal{F}(f)$. We can view the Fourier transform $\mathcal{F}(f)$ of f as a function $\mathcal{F}(f)\colon \mathbb{Z} \to \mathbb{C}$ with domain $\mathbb{Z}$.

On the other hand, given a $\mathbb{Z}$-indexed sequence $c = (c_m)_{m\in\mathbb{Z}}$ of complex numbers c_m, we can define the Fourier series $\overline{\mathcal{F}}(c)$ associated with c, or *Fourier cotransform* of c, given by

$$\overline{\mathcal{F}}(c)(\theta) = \sum_{m\in\mathbb{Z}} c_m e^{im\theta}.$$

This time, $\overline{\mathcal{F}}(c)$ is a function $\overline{\mathcal{F}}(c)\colon \mathbb{T} \to \mathbb{C}$ with domain $\mathbb{T}$. Fourier inversion can be stated as the equation

$$f(\theta) = ((\overline{\mathcal{F}} \circ \mathcal{F})(f))(\theta).$$

Of course, there is an issue of convergence. Namely, in general, $\widehat{f} = \mathcal{F}(f)$ does not belong to $\ell^1(\mathbb{Z})$. There are special cases for which Fourier inversion holds, in particular, if $f \in \mathrm{L}^2(\mathbb{T})$.

Let us now consider the Fourier transform of (not necessarily periodic) functions defined on $\mathbb{R}$. For any function $f \in \mathrm{L}^1(\mathbb{R})$, the *Fourier transform* $\widehat{f} = \mathcal{F}(f)$ of f is the function $\mathcal{F}(f)\colon \mathbb{R} \to \mathbb{C}$ defined on $\mathbb{R}$ given by

$$\widehat{f}(x) = \mathcal{F}(f)(x) = \int_{\mathbb{R}} f(y)e^{-iyx}\,\frac{dx(y)}{\sqrt{2\pi}},$$

and the *Fourier cotransform* $\overline{\mathcal{F}}(f)$ of f is the function $\overline{\mathcal{F}}(f)\colon \mathbb{R} \to \mathbb{C}$ defined on $\mathbb{R}$ given by

$$\overline{\mathcal{F}}f(x) = \int_{\mathbb{R}} f(y)e^{iyx}\,\frac{dx(y)}{\sqrt{2\pi}}.$$

This time, the domain of the Fourier transform is the same as the domain of the Fourier cotransform, but this is an exceptional situation. Also, in general the Fourier transform $\widehat{f}$ is not integrable, so Fourier inversion only holds in special cases.

The preceding examples suggest two questions:

(1) What is the *input domain* of the Fourier transform?

(2) What is the *output domain* of the Fourier transform?

The answer to (1) is that the domain of the Fourier transform, denoted by $\mathcal{F}$, is a set of functions on a *group* G. In order for the Fourier transform to be useful, it should behave well with respect to an operation on the set of functions on G called *convolution* (denoted $f * g$), which implies that these functions should be *integrable*.

This leads to the first subtopic, which is: *what is integration on a group*? The technical answer involves the *Haar measure* on a locally compact group. Thus, any serious effort to understand what the Fourier transform is entails learning a certain amount of measure theory and integration theory, passing through versions of the Radon–Riesz theorem relating Radon functionals and Borel measures, and culminating with the construction of the Haar measure. This preliminary material is discussed in Chapters 2, 3, 4, 5, 7, and 8.

Chapter 2 gathers some basic results about function spaces, in particular, about different types of convergence (pointwise, uniform, compact). Some sophisticated notions cannot be avoided, such as equicontinuity, filters, topologies defined by semi-norms, and Fréchet spaces.

Chapter 3 provides a quick review of the Riemann integral and its generalization to regulated functions.

Chapter 4 is devoted to basics of measure theory: σ-algebras, semi-algebras, measurable spaces, monotone classes, (positive) measures, measure spaces, null sets, and properties holding almost everywhere. We also define outer measures and prove Carathéodory's theorem which gives a method for constructing a measure from an outer measure. We conclude by using Carathéodory's theorem to define the Lebesgue measure on $\mathbb{R}$ and $\mathbb{R}^n$ from the Lebesgue outer measure. Our presentation relies on Halmos [44], Rudin [79], Lang [62], and Schwartz [86].

Chapter 5 develops the theory of Lebesgue integration in a fairly general context, namely functions defined on a measure space taking values in a Banach space. This integral is usually known as the Bochner integral (developed independently by Dunford). We agree with Lang (Lang [62]) that the investment needed to deal with functions taking values in a Banach space rather than in $\mathbb{R}$ is minor, and that the reward is worthwhile. This approach is presented in detail in Dunford and Schwartz [30], and more recent (and easier to read) expositions of this method are given in Lang [62] and Marle [69].

After reading this chapter, the reader will know what are the spaces $\mathrm{L}^1(X)$, $\mathrm{L}^2(X)$, and $\mathrm{L}^\infty(X)$, which is essential to move on to the study of harmonic analysis. In this chapter, we provide some proofs.

Chapter 7 presents the theory of integration on locally compact spaces due to Radon and Riesz based on linear functionals on the space of continuous functionals with compact support. Although this material is well-known to analysts, it may be less familiar to other mathematicians, and most computer scientists have not been exposed to it. Our presentation relies heavily on Rudin [79] (Chapter 2), Lang [62] (Chapter IX), Folland [34] (Chapter 7), Marle [69], and Schwartz [86]. We also borrowed much from Dieudonné [24] (Chapter XIII).

We state the famous representation theorem of Radon and Riesz for positive linear functionals and certain types of positive Borel measures (Theorem 7.8 and Theorem 7.15). Here, inspired by Folland and Lang, we define a *σ-Radon measure* as a Borel measure which is outer regular, σ-inner regular, and finite on compact subsets. A *Radon measure* is a σ-Radon measure which is also inner regular. Linear functionals which are bounded on the space of continuous functions with support contained in a fixed compact support are called *Radon functionals*. We have avoided Bourbaki and Dieudonné's use of the term Radon measure for a Radon functional, which is just too confusing.

We define complex measures, and following Rudin, we present the Radon–Riesz correspondence between bounded Radon functionals and complex (regular) measures (Theorem 7.30). This theorem is absolutely crucial to the construction of the Haar measure and to the definition of the convolution of complex measures and of functions.

Chapter 8 contains a rather complete discussion of the Haar measure on a locally compact group, convolution, and the application of convolution to regularization. After some preliminaries about topological groups (Section 8.1), we describe the method for constructing a left Haar measure from a left Haar functional, following essentially Weil's proof as presented in Folland [33] (see Sections 8.2 and 8.3). We prove almost everything, except for a technical lemma. Then we prove the uniqueness of the left Haar measure up to a positive constant, using Dieudonné's method [24] (Section 8.4). We introduce the modular function and the modulus of an automorphism. We show how to use the Haar measure to construct a hermitian inner product invariant under the representation of a compact group. We discuss G-invariant measures on homogeneous spaces.

One of the main applications of the Haar measure is the definition of the convolution $\mu * \nu$ of (complex) measures and the convolution $f * g$ of functions; see Section 8.11. Under convolution, the set $\mathcal{M}^1(G)$ of complex regular measures is a Banach algebra with an involution, and a multiplicative unit element. This algebra contains the Banach subalgebra $\mathrm{L}^1(G)$, which doesn't have a multiplicative unit in general. In Section 8.14, we show that by convolving a function f with functions g_n from a "well-behaved" family we obtain a sequence $(f * g_n)$ of functions more regular that f that converge to f. This technique is known as *regularization*.

Chapter 8 is the last of the chapters dealing with background material. Similar material is coved in Folland [33], and very extensively in Hewitt and Ross [49] (over 400 pages).

The main chapters presenting some elements of harmonic analysis, in particular the Fourier transform, are:

1. Chapter 6, in which the classical theory of the Fourier transform (and cotransform) on $\mathbb{T}$, $\mathbb{R}$, and then $\mathbb{T}^n$ and $\mathbb{R}^n$, is presented. We also present the sampling theorem due to Shannon, and discuss the Heisenberg uncertainty principle. Our presentation is inspired by Rudin [79], Folland [32, 34], Stein and Shakarchi [94], and Malliavin [68].

2. Chapter 10, which is devoted to harmonic analysis on *locally compact abelian groups*, based on the seminal work of A. Weil, Gelfand, and Pontrjagin. Our presentation is based on Folland [33] and Bourbaki [9].

3. Chapter 13, which gives an exposition of harmonic analysis on a *compact not necessarily abelian* group G. The main result is the beautiful theorem of Peter and Weyl, which among other things, gives the structure of the algebra $\mathrm{L}^2(G)$ as a Hilbert sum of finite-dimensional spaces corresponding to irreducible representations of G. We rely heavily on Dieudonné [24, 21], Folland [33], and Hewitt and Ross [48].

4. Chapter 17, which presents a theory of the Fourier transform that generalizes all previous definitions, based on the concept of a *Gelfand pair* (G, K). We follow Dieudonné's exposition in [22].

Chapters 10, 13, and 17, require more preparatory material.

If G is a commutative locally compact group, then the domain of the Fourier transform on $\mathrm{L}^1(G)$ is the group $\widehat{G}$ of characters of G, the homomorphisms $\chi\colon G \to \mathbb{C}$ such that $|\chi(g)| = 1$ for all $g \in G$. The group $\widehat{G}$ is called the *Pontrjagin dual* of G. It turns out that $\widehat{G}$ is homeomorphic to the space $\mathsf{X}(\mathrm{L}^1(G))$ of characters of the Banach algebra $\mathrm{L}^1(G)$. Thus we need some knowledge about normed algebras. Chapter 9 presents the basic theory of normed algebras and their spectral theory needed for Chapter 10. The study of algebras and normed algebras focuses on three concepts:

(1) The notion of *spectrum* $\sigma(a)$ of an element a of an algebra A.

(2) If A is a commutative algebra, the notion of *character*, and the space $\mathsf{X}(A)$ of characters of A.

(3) If A is a commutative algebra, the notion of *Gelfand transform*, $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A); \mathbb{C})$.

The Gelfand transform from $\mathrm{L}^1(G)$ to $\mathsf{X}(\mathrm{L}^1(G))$ is the Fourier cotransform on $\mathrm{L}^1(G)$. Our presentation is inspired by Dieudonné [24], Bourbaki [9], and Rudin [80].

If G is a locally compact abelian group, then for any function $f \in \mathrm{L}^1(G)$, the Fourier transform $\mathcal{F}(f)$ of f is then a function

$$\mathcal{F}(f)\colon \widehat{G} \to \mathbb{C}.$$

In general, $\widehat{G}$ is completely different from G, and this creates problems. For the familiar cases, $G = \mathbb{T} \cong \mathbf{U}(1)$, $G = \mathbb{Z}$, $G = \mathbb{R}$, and $G = \mathbb{Z}/n\mathbb{Z}$, the characters are well known. The case $G = \mathbb{Z}/n\mathbb{Z}$ corresponds to the *discrete Fourier transform*.

For the groups listed above, we know that under some suitable restriction, we have *Fourier inversion*, which means that there is some transform $\overline{\mathcal{F}}$ (called *Fourier cotransform*) such that

$$f = \overline{\mathcal{F}}(\mathcal{F}(f)). \tag{$*$}$$

We have to be a bit careful because the domain of $\overline{\mathcal{F}}$ is $\mathrm{L}^1(\widehat{G})$, and not $\mathrm{L}^1(G)$, are they are usually very different beause in general G and $\widehat{G}$ are *not* isomorphic. Then (assuming that it makes sense), $\overline{\mathcal{F}}(\mathcal{F}(f))$ is a function with domain $\widehat{\widehat{G}}$, so there seems no hope, except in very special cases such as $G = \mathbb{R}$, that $(*)$ could hold. Fortunately, *Pontrjagin duality* asserts that G and $\widehat{\widehat{G}}$ are isomorphic, so $(*)$ holds (under suitable conditions) in the form

$$f = \overline{\mathcal{F}}(\mathcal{F}(f)) \circ \eta,$$

where $\eta\colon G \to \widehat{\widehat{G}}$ is a canonical isomorphism.

If G is a commutative abelian group, there is a beautiful and well understood theory of the Fourier transform based on results of Gelfand, Pontrjagin, and André Weil presented in Chapter 10. In particular, even though the Fourier transform is not defined on $\mathrm{L}^2(G)$ in general, for any function $f \in \mathrm{L}^1(G) \cap \mathrm{L}^2(G)$, we have $\mathcal{F}(f) \in \mathrm{L}^2(\widehat{G})$, and by Plancherel's theorem, the Fourier transform extends in a unique way to an isometric isomorphism between $\mathrm{L}^2(G)$ and $\mathrm{L}^2(\widehat{G})$. Furthermore, if we identify G and $\widehat{\widehat{G}}$ by Pontrjagin duality, then $\mathcal{F}$ and $\overline{\mathcal{F}}$ are mutual inverses.

If G is *not* commutative, things are a lot tougher. Characters no longer provide a good input domain, and instead one has to turn to *unitary representations*. A unitary representation is a homomorphism $U\colon G \to \mathbf{U}(H)$ satisfying a certain continuity property, where $\mathbf{U}(H)$ is the group of unitary operators on the Hilbert space H. Then $\widehat{G}$ is the set of equivalence classes of irreducible unitary representations of G, but it is no longer a group.

Chapters 11 and 12 provide the background material needed in Chapter 13. Chapter 11 discusses representations of algebras, and gives an introduction to Hilbert algebras. For our purposes, the most important example of a complete Hilbert algebra is $\mathrm{L}^2(G)$, where G is a *compact* (metrizable) group. One of the main theorems of this chapter is a structure theorem for complete separable algebras (Theorem 11.31). This theorem is the key result for proving a major part of the Peter–Weyl theorem in Chapter 13. We follow closely Dieudonné [24].

Chapter 12 gives a brief introduction to the theory of unitary representations of locally compact groups. We prove that there is a bijection between unitary representations of a locally compact group G and nondegenerate representations of the algebra $\mathrm{L}^1(G)$. We define functions and measures of positive type, and prove that there is a bijection between the set of functions of positive type and cyclic unitary representations (Gelfand–Raikov, Godement). We follow Dieudonné [21, 22] and Folland [33].

One more preparatory chapter is needed for Chapter 17. Chapter 15 gives an introduction to induced representations. The goal is to construct a unitary representation of a group G from a representation of a closed subgroup H of G.

A way to deal with noncommutativity, due to Gelfand, is to work with pairs (G, K) where K is a compact subgroup of G. This theory is presented in Chapter 17. Then, instead of working with functions on G, which is "too big," we work with functions on the homogeneous space G/K, the space of left cosets. Under certain assumptions on G and K, which makes (G, K) a *Gelfand pair*, it is possible to consider a commutative algebra of functions on the set of double cosets KsK $(s \in G)$, so that some results from the commutative theory can be used. The domain of the Fourier transform is a set of functions called *spherical functions*, and this set happens to be homeomorphic to the set of characters on the commutative algebra mentioned above. There is a very nice theory of the Fourier transform and its inverse, but how useful it is in practice remains to be seen.

More basic background material dealing with elementary topology, matrix norms, groups and group actions, and Hilbert spaces is found in Appendices A, B, C, and D. These chapters should be considered as appendices and should be consulted by need.

Even though the present document is already quite long, it is by no means complete. If a locally compact group is a *Lie group*, then the whole machinery of Lie algebras and Lie groups developed by Élie Cartan and Hermann Weyl involving weights and roots becomes available. In particular, if G is a connected *semisimple Lie group*, there is a beautiful and extensive theory of harmonic analysis due to Harish–Chandra. We lack the expertise to discuss this difficult theory and refer the ambitious reader to Warner's monographs [103, 104], and Helgason's treatises [47], [46] (especially Chapter IV), and [45] (especially Chapter III, Section 12).

To keep the length of this book under control, we resigned ourselves to omit many proofs. This is unfortunate because some beautiful proofs (such as the proof of the Radon–Riesz theorem for bounded Radon functional) had to be omitted. However, whenever a proof is omitted, we provide precise pointers to sources where such a proof is given.

After Chapter 1 the logical starting point of this book is Chapter 2, followed by the other chapters in consecutive order. However, some readers might find it more illuminating to proceed directly to Chapter 6 which provides a less abstract view of Fourier analysis and harmonic analysis. Readers not familiar with the Lebesgue theory of integration should not be concerned, and they should replace this fancy notion with the notion of integral that they are familiar with. The consequence of such a simplifying assumption is that some of the results may not be quite correct, but this should be a good motivation to return to the chapters dealing with measure theory and integration.

Chapter 2

Function Spaces Often Encountered

Various spaces of functions $f\colon E \to F$ from a topological space E to a metric space or a normed vector space F come up all the the time. The most frequently encountered are the spaces $(F^E)_b$ of bounded functions, the spaces $\mathcal{K}(E;F)$ of continuous functions with compact support, the spaces $\mathcal{C}_0(E;F)$ of continuous functions which tend to zero at infinity, and the spaces $\mathcal{C}_b(E;F)$ of continuous bounded functions. When F is a normed vector space, all these spaces are normed vector spaces with the sup norm. An important issue about function spaces is the convergence of sequences of functions. We review the main three notions, pointwise convergence (also known as simple convergence), uniform convergence, and compact convergence. A sequence of continuous functions may converge pointwise to a function which is not continuous. Uniform convergence has a better behavior. If F is a complete normed vector space, then both spaces $\mathcal{C}_b(E;F)$ and $(F^E)_b$ are also complete under uniform convergence. An interesting family of functions in $(F^{[a,b]})_b$ is the space $\mathrm{Reg}([a,b];F)$ of regulated functions. These functions have at most only countably many simple kinds of discontinuities called discontinuities of the first kind. If F is a complete normed vector space, then the space $\mathrm{Reg}([a,b];F)$ is complete. It contains a subspace $\mathrm{Step}([a,b];F)$ consisting of very simple functions called step functions, which take finitely many different values on consecutive open intervals. The space $\mathrm{Step}([a,b];F)$ is dense in $\mathrm{Reg}([a,b];F)$. If E is a locally compact space, then the space $\mathcal{C}_0(E;\mathbb{C})$ is the closure of $\mathcal{K}(E;\mathbb{C})$ in $\mathcal{C}_b(E;\mathbb{C})$. This chapter relies heavily on the material discussed Appendix A so the reader may want to refer to this appendix whenever the need arises.

2.1 The Function Space F^E and Pointwise Convergence

In this section we study the space of functions $f\colon E \to F$, where E and F are arbitrary topological spaces. We denote the set of all functions from E to F by F^E.

Our first goal is to make F^E into a topological space in its own right. Surprisingly, one

of the easiest ways to describe a topology on F^E is to follow Tychonoff and observe that

$$F^E \cong \prod_{x \in E} F_x, \qquad F_x = F.$$

Since F^E is isomorphic to an E-indexed product space, we may give it a product topology as follows: a subset of functions in F^E is open if it is the union of subsets U_A of functions $f\colon E \to F$ for which there is some *finite* subset A of E such that $f(x) \in U_x$ for all $x \in A$, where U_x is an open subset of F, and $f(x) \in F$ is arbitrary for all $x \in E - A$; see Figure 2.1.

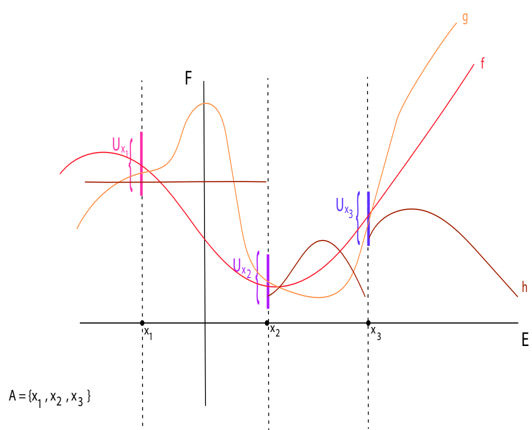


Figure 2.1: A schematic illustration of an open set U_A of F^E, (where the reader may assume $E = F = \mathbb{R}$). The three functions $f, g, h \in U_A$ since they pass "through" the open sets U_{x_i}, for $1 \leq i \leq 3$.

Equivalently, for any $x \in E$ and any open subset U of F, let $S(x, U)$ be the set

$$S(x, U) = \{f \mid f \in F^E,\ f(x) \in U\};$$

see Figure 2.2. Then observe that

$$U_A = \bigcap_{x \in A} S(x, U_x), \quad A \text{ finite},$$

that is, the sets $S(x, U)$ form a subbasis of the product topology on F^E.

For every $x \in E$, if $\pi_x\colon F^E \to F$ is the projection map given by

$$\pi_x(f) = f(x), \qquad f \in F^E,$$

(evaluation at x), then the product topology on F^E is the weakest topology that makes all the π_x continuous. Indeed, the weakest topology on F^E making all the π_x continuous consists

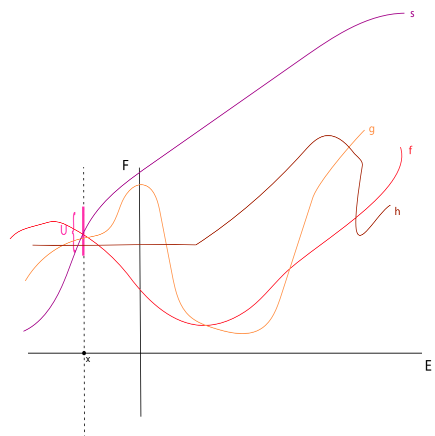


Figure 2.2: A schematic illustration of an open set $S(x, U)$ of F^E, (where the reader may assume $E = F = \mathbb{R}$). The four functions $f, g, h, s \in S(x, U)$ since they pass "through" the open set U.

of all unions of finite intersections of subsets of F^E of the form $\pi_x^{-1}(U_x)$, for any open subset U_x of F, but

$$\pi_x^{-1}(U_x) = S(x, U_x),$$

is one of the sets in the subbasis defined above. For this reason, the product topology on F^E is also called the *weak topology* induced by the family of functions $(\pi_x)_{x \in E}$; see Rudin [80] (Chapter 3, Section 3.8).

Now that we have made F^E into a topological space, we can ask ourselves what it means for a sequence $(f_n)_{n \geq 1}$ of functions $f_n \colon E \to F$ to converge to f. By definition of the product topology, $(f_n)_{n \geq 1}$ converges to f if and only if given any subbasic open set $S(x, U)$ containing f, there exists $n_0 \geq 0$ such that $f_n \in S(x, U)$ whenever $n \geq n_0$. A moment of reflection shows that we may reinterpret the previous statement as saying for a *fixed* point $x \in E$, $f_n(x)$ becomes "arbitrarily" close to $f(x)$. This reinterpretation is rigorously stated in terms of pointwise convergence, namely that for *fixed* $x \in E$, $\lim_{n \to \infty} f_n(x) = f(x)$. The notion of pointwise convergence does not require F to be a metric space, but since this is the situation we most often encounter, we give the definition assuming that (F, d) is a metric space,

Definition 2.1. Let (F, d) be a metric space. A sequence $(f_n)_{n \geq 1}$ of functions $f_n \colon E \to F$ *converges pointwise* (or *converges simply*) to a function $f \colon E \to F$ if for every $x \in E$, for every $\epsilon > 0$, there is some $N > 0$ such that

$$d(f_n(x), f(x)) < \epsilon \quad \text{for all } n \geq N.$$

See Figure 2.3.

To reiterate, Definition 2.1 says that for every $x \in E$, the sequence $(f_n(x))_{n \geq 1}$ converges to $f(x)$. Observe that the above ϵ depends on x.

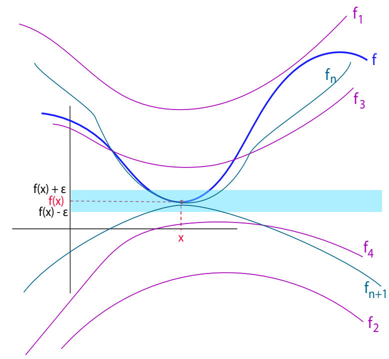


Figure 2.3: A schematic illustration of $f_n(x)$ converging pointwise $f(x)$, where $E = F = \mathbb{R}$. As n increases, the graph of $f_n(x)$ near x must be in the band determined by the graphs of $f(x) - \epsilon$ and $f(x) + \epsilon$.

A sequence $(f_n)_{n\geq 1}$ of elements of F^E converges pointwise to $f \in F^E$ iff the sequence $(f_n)_{n\geq 1}$ converges to f in the product topology; see Munkres [75] (Chapter 7, Section 46, Theorem 46.1), or Folland [34] (Chapter 4, Proposition 4.12). Consequently, the product (weak) topology is also called the topology of pointwise convergence and pointwise convergence is also known as *weak convergence*. We summarize the previous discussion in the following definition.

Definition 2.2. If F is any topological space and E is any set, the topology on F^E having the sets

$$S(x, U) = \{f \mid f \in F^E,\ f(x) \in U\}, \qquad x \in E,\ U \text{ open in } F,$$

as a subbasis is the *topology of pointwise convergence*. An open subset of F^E in this topology is any union (possibly infinite) of finite intersections of subsets of the form $S(x, U)$ as above.

If F is Hausdorff, so is the topology of pointwise convergence. Indeed, if $f, g \in F^E$ and $f \neq g$, then there is some $x \in E$ such that $f(x) \neq g(x)$, and since F is Hausdorff, there exist two disjoint open subsets $U_{f(x)}$ and $U_{g(x)}$ with $f(x) \in U_{f(x)}$ and $g(x) \in U_{g(x)}$. Then $\pi_x^{-1}(U_{f(x)})$ and $\pi_x^{-1}(U_{g(x)})$ are disjoint open subsets with $f \in \pi_x^{-1}(U_{f(x)})$ and $g \in \pi_x^{-1}(U_{g(x)})$; see Figure 2.4.

When F is a metric space there are two important subsets within F^E, the subspace of *continuous* functions $\mathcal{C}(E; F)$, and the subspace of *bounded* functions $(F^E)_b$. As shown in Figure 2.5, both $\mathcal{C}(E; F)$ and $(F^E)_b$ inherit subspace topologies from the product topology of F^E. But if F is either a metric or a normed vector space, we can place "finer" topologies on both $\mathcal{C}(E; F)$ and $(F^E)_b$. In the next section we discuss how such a topology makes $(F^E)_b$ into its own metric space by considering it an independent space in its own right, not necessarily embedded in F^E.

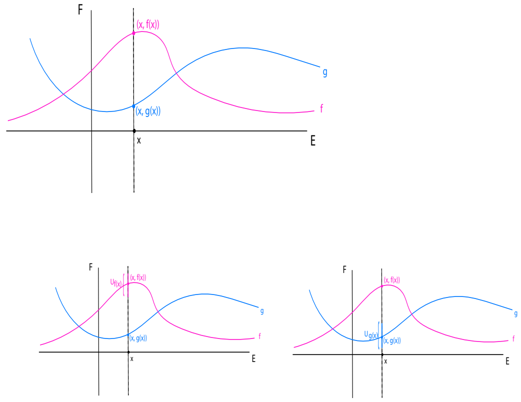


Figure 2.4: If F is Hausdorff, so is the topology of pointwise convergence. For convenience, let $E = F = \mathbb{R}$. The top figure illustrates two distinct elements of F^E. The bottom left figure illustrates the open set $\pi_x^{-1}(U_{f(x)})$, while the bottom right figure illustrates the open set $\pi_x^{-1}(U_{g(x)})$. These two sets separate f and g within F^E.

2.2 Spaces of Bounded Functions

In this section we are dealing with functions $f\colon E \to F$, where F is either a metric space or a normed vector space.

First assume that F is a metric space with metric d. We would like to make F^E into a metric space. It is natural to define a metric on F^E by setting

$$d_\infty(f, g) = \sup_{x \in E} d(f(x), g(x))$$

for any two functions $f, g\colon E \to F$, but if $d(f(x), g(x))$ is unbounded as x ranges over E, the expression $\sup_{x\in E} d(f(x), g(x))$ is undefined. Therefore, we consider the space of bounded functions defined as follows.

Definition 2.3. If (F, d) is a metric space, a function $f\colon E \to F$ is *bounded* if its image $f(E)$ is bounded in F, which means that $f(E) \subseteq B(a, \alpha)$, for some closed ball $B(a, \alpha)$ of center a and radius $\alpha > 0$. See Figure 2.6. The space of bounded functions $f\colon E \to F$ is denoted by $(F^E)_b$.

If $f\colon E \to F$ and $g\colon E \to F$ are bounded functions, then it is easy to see that if $f(E) \subseteq B(a, \alpha)$ and if $g(E) \subseteq B(b, \beta)$, then

$$d(f(x), g(x)) \leq \alpha + \beta + d(a, b) \quad \text{for all } x \in E;$$

see Figure 2.7. Therefore, $\sup_{x\in E} d(f(x), g(x))$ is well defined. It is easy to check that if we define

$$d_\infty(f, g) = \sup_{x \in E} d(f(x), g(x))$$

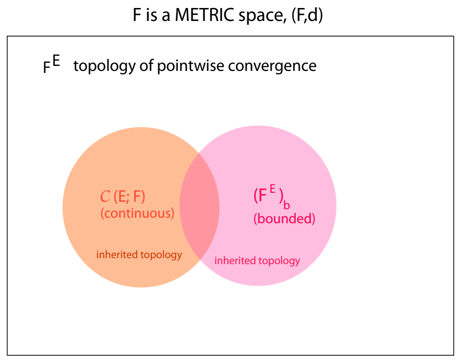


Figure 2.5: A Venn diagram of illustration of F^E and the subsets $\mathcal{C}(E;F)$ and $(F^E)_b$ with the inherited topology of pointwise convergence.

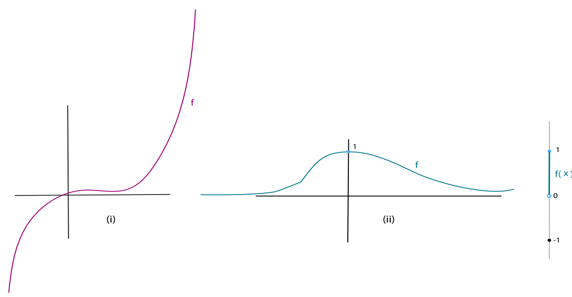


Figure 2.6: Let $E = F = \mathbb{R}$ with the Euclidean metric. In Figure (i), f is unbounded since $f(E) = \mathbb{R}$. In Figure (ii), $f \in (F^E)_b$ since $f(E) = (0,1]$ and $(0,1] \subset B(0,1) = [-1,1]$.

for any two bounded functions f, g, then d is indeed a metric on $(F^E)_b$.

Definition 2.4. If (F, d) is a metric space, then for any two bounded functions $f, g \in (F^E)_b$, the quantity

$$d_\infty(f,g) = \sup_{x \in E} d(f(x), g(x))$$

is a metric on $(F^E)_b$. See Figure 2.8.

If $(F, \| \|)$ is normed metric space, then F^E is a vector space, and it is easy to check that $(F^E)_b$ is also a vector space. For any bounded function $f \colon E \to F$ (which means that $f(E) \subseteq B(0,\alpha)$, for some closed ball $B(0,\alpha)$), then

$$\|f\|_\infty = \sup_{x \in E} \|f(x)\|$$

is a norm on the vector space $(F^E)_b$.

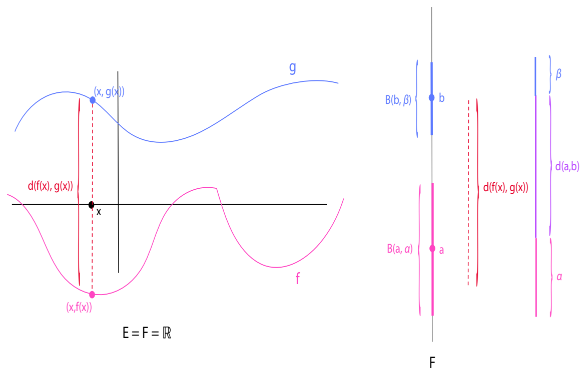


Figure 2.7: An illustration of $d(f(x), g(x)) \leq \alpha + \beta + d(a, b)$, when $E = F = \mathbb{R}$ with the Euclidean metric.

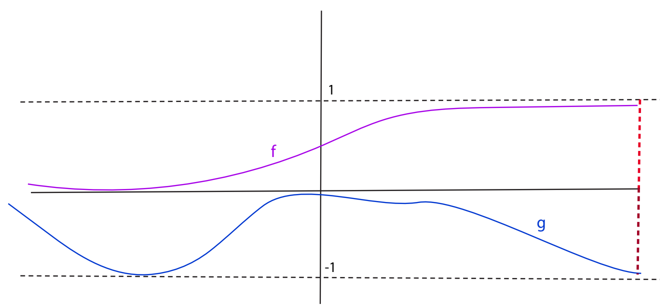


Figure 2.8: Let $E = F = \mathbb{R}$ with the Euclidean metric. Both $f, g \in (F^E)_b$ since $f(E) = (0, 1)$, while $g(E) = [-1, 0)$. The concatenation of the vertical dashed red lines is $d_\infty(f, g) = \sup_{x\in E} d(f(x), g(x)) = 1 - (-1) = 2$.

Definition 2.5. If $(F, \| \, \|)$ is a normed vector space, then for any bounded function $f \in (F^E)_b$, the quantity

$$\|f\|_\infty = \sup_{x\in E} \|f(x)\|$$

is a norm on $(F^E)_b$, often called the *sup norm*; see Figure 2.9.

The following important theorem can be shown; see Schwartz [83] (Chapter XV, Section 1, Theorem 1).

Theorem 2.1. *(1) If (F, d) is a complete metric space, then $((F^E)_b, d_\infty)$ is also a complete metric space.*

(2) If $(F, \| \, \|)$ is a complete normed vector space, then $((F^E)_b, \| \, \|_\infty)$ is also a complete normed vector space.

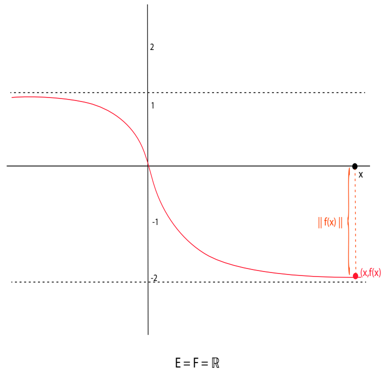


Figure 2.9: Let $f \in (F^E)_b$, where $E = F = \mathbb{R}$ with norm given by the absolute value. Then $\|f\|_\infty = 2$.

2.3 Uniform Convergence of Functions

When dealing with spaces of functions, a crucial issue is to identify notions of limit that preserve certain desirable properties, such as continuity.

Unfortunately the notion of pointwise convergence within F^E does not have such a property. If a sequence $(f_n)_{n\geq 1}$ of continuous functions converges pointwise to a function f, this f is not necessarily continuous. For example, the functions $f_n\colon [0,1] \to \mathbb{R}$ given by $f_n(x) = x^n$ are continuous, and the sequence $(f_n)_{n\geq 1}$ converges pointwise to the discontinuous function $f\colon [0,1] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x = 1, \end{cases}$$

as evidenced by Figure 2.10.

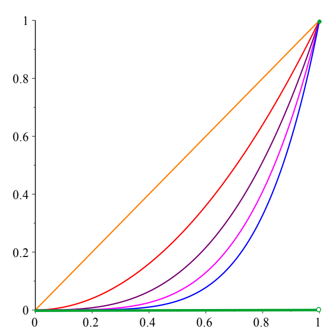


Figure 2.10: The sequence of functions $f_n(x) = x^n$ over $[0,1]$ converges pointwise to the discontinuous green graph.

However, if F is a metric space there is a stronger notion of convergence, uniform convergence, which ensure that continuity *is preserved* in the limit.

Definition 2.6. Let (F, d) be a metric space. A sequence $(f_n)_{n \geq 1}$ of functions $f_n \colon E \to F$ *converges uniformly* to a function $f \colon E \to F$ if for every $\epsilon > 0$, there is some $N > 0$ such that

$$d(f_n(x), f(x)) < \epsilon \quad \text{for all } n \geq N \text{ and for all } x \in E.$$

See Figure 2.11.

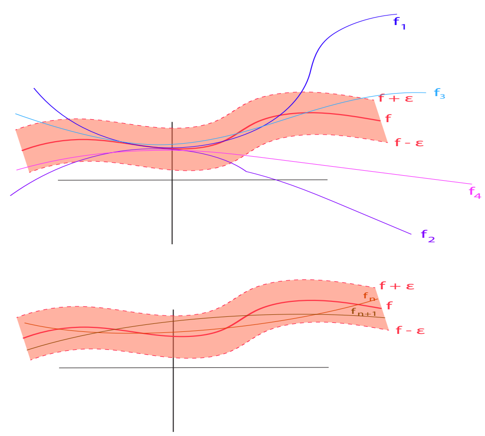


Figure 2.11: A schematic illustration of f_n converging uniformly to f, where $E = F = \mathbb{R}$. As n increases, the graph of f_n must lie entirely in the band determined by the graphs of $f - \epsilon$ and $f + \epsilon$.

Observe that convergence in the metric space of bounded functions $((F^E)_b, d_\infty)$ is the uniform convergence of sequences of functions. Similarly, convergence in the normed vector space of bounded functions $((F^E)_b, \| \, \|_\infty)$ is the uniform convergence of sequences of functions. For this reason, the topology on $(F^E)_b$ induced by the metric d_∞ (or the norm $\| \, \|_\infty$) is sometimes called the *topology of uniform convergence*. Figure 2.12 illustrates how the intrinsic metric based topology of uniform convergence is the finer topology which replaces the inherited topology of pointwise convergence.

The difference between simple (pointwise) and uniform convergence is that in uniform convergence, ϵ is independent of x. For example the functions $f_n \colon [0, 2\pi] \to \mathbb{R}$ defined by $f_n(x) = n \sin\left(\frac{x}{n}\right)$ converges uniformly to $f(x) = x$, as evidenced by Figure 2.13. Consequently, uniform convergence implies simple convergence, but the converse is false, as the following examples illustrate.

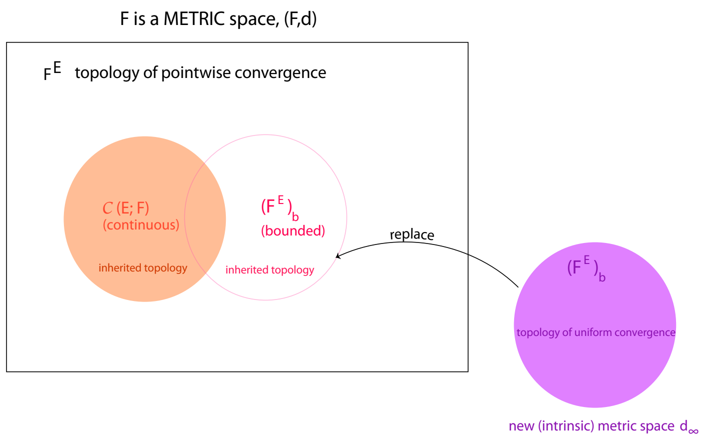


Figure 2.12: A Venn diagram illustration of F^E and two of its subspaces; $\mathcal{C}(E;F)$, which has the inherited topology of pointwise convergence, and $(F^E)_b$, which has the inherited topology of pointwise convergence replaced with the topology of uniform convergence.

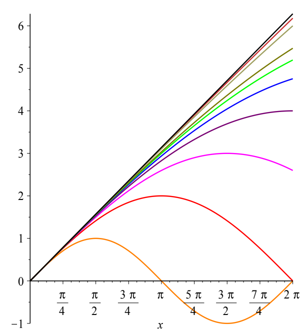


Figure 2.13: The colored functions $f_n(x) = n\sin\left(\frac{x}{n}\right)$, over the domain $[0, 2\pi]$, converge uniformly to the black line $f(x) = x$.

Example 2.1. Let $g\colon \mathbb{R} \to \mathbb{R}$ be the function given by

$$g(x) = \frac{1}{1+x^2},$$

and for every $n \geq 1$, let $f_n\colon \mathbb{R} \to \mathbb{R}$ be the function given by

$$f_n(x) = \frac{1}{1+(x-n)^2}.$$

The function f_n is obtained by translating g to the right using the translation $x \mapsto x + n$; see Figure 2.14.

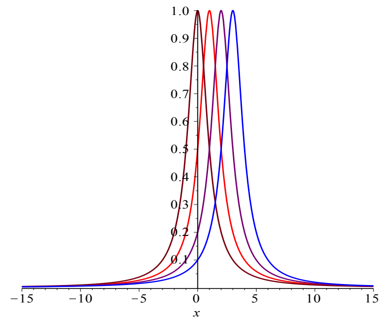


Figure 2.14: The bell curve graphs of Example 2.1; $g(x)$ in brown; $f_1(x)$ in red; $f_2(x)$ in purple; $f_3(x)$ in blue.

Since
$$\lim_{n\mapsto\infty} \frac{1}{1+(x-n)^2} = 0,$$
the sequence $(f_n)_{n\geq 1}$ converges pointwise to the zero function f given by $f(x) = 0$ for all $x \in \mathbb{R}$. However, since the the maximum of each f_n is 1, we have
$$d_\infty(f_n, f) = 1 \quad \text{for all } n \geq 1,$$
so the sequence $(f_n)_{n\geq 1}$ does not converge uniformly to the zero function.

Example 2.2. Pick any positive real $\alpha > 0$. For each $n \geq 1$, let $f_n \colon \mathbb{R} \to \mathbb{R}$ be the piecewise affine function defined as follows:
$$f_n(x) = \begin{cases} 0 & \text{if } x \leq 0 \text{ or } x \geq 1/n \\ (2n)n^\alpha x & \text{if } 0 \leq x \leq 1/(2n) \\ 2n^\alpha(1-nx) & \text{if } 1/(2n) \leq x \leq 1/n. \end{cases}$$
See Figure 2.15.

For every $x > 0$, there is some n such that $1/n < x$, so $\lim_{n\mapsto\infty} f_n(x) = 0$ for $x > 0$, and since $f_n(x) = 0$ for $x \leq 0$, we see that the sequence $(f_n)_{n\geq 1}$ converges pointwise to the zero function f. However, the maximum of f_n is n^α (for $x = 1/(2n)$) so
$$d_\infty(f_n, f) = n^\alpha,$$
and $\lim_{n\mapsto\infty} d_\infty(f_n, f) = \infty$, so the sequence $(f_n)_{n\geq 1}$ does not converge uniformly to the zero function.

If E is a topological space, it is useful to define the following local notion of uniform convergence.

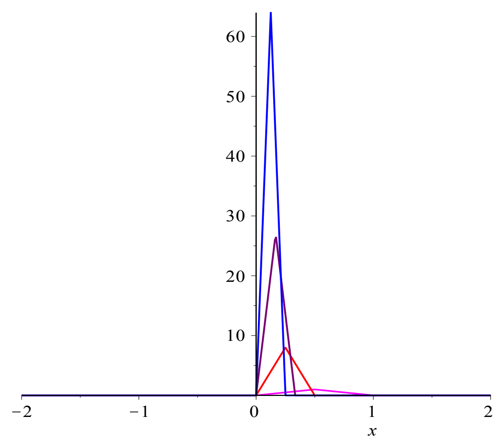


Figure 2.15: The piecewise affine functions of Example 2.2 with $\alpha = 3$; $f_1(x)$ in magenta; $f_2(x)$ in red; $f_3(x)$ in purple; $f_4(x)$ in blue. Each $f_n(x)$ has a symmetrical triangular peak. As n increases, the peak becomes taller and thinner.

Definition 2.7. Let E be a topological space and let (F, d) be a metric space. A sequence $(f_n)_{n\geq 1}$ of functions $f_n\colon E \to F$ *converges locally uniformly* to a function $f\colon E \to F$ if for every $x \in E$, there is some open subset U of E containing x such that for every $\epsilon > 0$, there is some $N > 0$ such that

$$d(f_n(x), f(x)) < \epsilon \quad \text{for all } n \geq N \text{ and for all } x \in U;$$

see Figure 2.16.

If E is locally compact, it is easy to see that a sequence $(f_n)_{n\geq 1}$ converges locally uniformly iff it converges uniformly on every compact subset of E.

As we saw at the beginning of this section, the pointwise limit of a sequence $(f_n)_{n\geq 1}$ of continuous functions needs not be continuous. However, if the convergence is locally uniform, then the limit is continuous. The following theorem gives sufficient conditions for the limit of a sequence of continuous functions to be continuous.

Theorem 2.2. *Let E be a topological space, (F, d) be a metric space, and let $(f_n)_{n\geq 1}$ be a sequence of functions $f_n\colon E \to F$ converging locally uniformly to a function $f\colon E \to F$. Then the following properties hold:*

(1) If the functions f_n are continuous at some point $a \in E$, then the limit f is also continuous at a.

(2) If the functions f_n are continuous (on the whole of E), then the limit f is also continuous (on the whole of E).

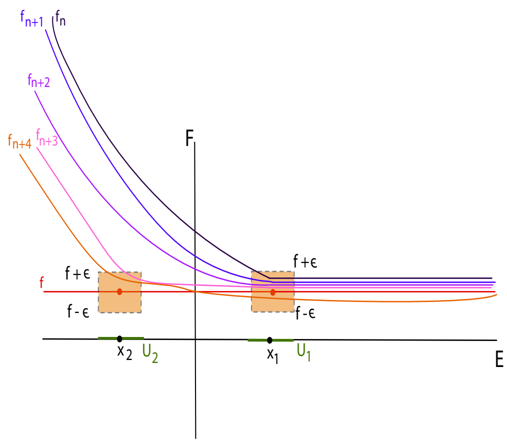


Figure 2.16: Let $E = F = \mathbb{R}$ with the Euclidean metric, and let f be red horizontal line. The sequence $(f_n)_{n\geq 1}$ converges locally uniformly to f. Note that for a given x and a given ϵ, the N will vary. For example, for x_1, $N = n$, while for x_2, $N = n + 4$.

(3) If E is a metric space, the sequence $(f_n)_{n\geq 1}$ converges uniformly to f, and the f_n are uniformly continuous on E, then the limit f is also uniformly continuous on E.

The proof of Theorem 2.2 can be found in Schwartz [83] (Chapter XV, Section 4, Theorem 1).

Here are a few applications of Theorem 2.2.

Definition 2.8. Let E be a topological space, and let (F, d) be a metric space. The metric subspace of $((F^E)_b, d_\infty)$ consisting of all continuous bounded functions $f\colon E \to F$ is denoted $\mathcal{C}_b(E; F)$. If $(E, \| \|)$ is a normed vector space, the normed subspace of $((F^E)_b, \| \|_\infty)$ consisting of all continuous bounded functions $f\colon E \to F$ is also denoted $\mathcal{C}_b(E; F)$.

Proposition 2.3. *Let E be a topological space, and let (F, d) be a metric space. The metric subspace $\mathcal{C}_b(E; F)$ of $((F^E)_b, d_\infty)$ is closed. If (F, d) is a complete metric space, then $(\mathcal{C}_b(E; F), d_\infty)$ is also complete.*

Proposition 2.4. *Let E be a topological space, and let $(F, \| \|)$ be a normed vector space. The normed subspace $\mathcal{C}_b(E; F)$ of $((F^E)_b, \| \|_\infty)$ is closed. If $(F, \| \|)$ is a complete normed vector space, then $(\mathcal{C}_b(E; F), \| \|_\infty)$ is also complete.*

An important special case of Proposition 2.4 is the case where $F = \mathbb{R}$ or $F = \mathbb{C}$, namely, our functions are real-valued continuous and bounded functions $f\colon E \to \mathbb{R}$, or complex-valued continuous and bounded functions $f\colon E \to \mathbb{C}$. The spaces of functions $(\mathcal{C}_b(E; \mathbb{R}), d_\infty)$ and $(\mathcal{C}_b(E; \mathbb{C}), \| \|_\infty)$ are complete.

If E is compact and if $(F, \| \|)$ is a complete normed vector space, then every continuous function $f\colon E \to F$ is bounded. As a consequence, the space $\mathcal{C}(E; F)$ of continuous functions $f\colon E \to F$ is complete.

2.4 Compact Convergence and the Space of Continuous Functions

In the past two sections, for the case of a metric space (F, d), we investigated $(F^E)_b$. The topology we placed on $(F^E)_b$, that of uniform convergence, was intrinsic in nature and *was not* induced by the topology of pointwise convergence on F^E, but in fact finer than the induced topology. Still assuming that F is a metric space, we now want to investigate $\mathcal{C}(E; F)$. Unlike the case of $(F^E)_b$, we *can* use F^E to induce an appropriate topology on $\mathcal{C}(E; F)$, but the key is to create a *new* finer topology on F^E, namely that of compact convergence. This is not an arbitrary choice, but one based on experience, since the topology of compact convergence occurs in the definition of the dual of an abelian locally compact group.

Definition 2.9. Let E be any topological space and let (F, d) be a metric space. For any $\epsilon > 0$, any $f \in F^E$, and any compact subset K of E, define the set $B_K(f, \epsilon)$ by

$$B_K(f, \epsilon) = \left\{ g \in F^E \mid \sup_{x \in K} d(f(x), g(x)) < \epsilon \right\};$$

see Figure 2.17. The family of sets $B_K(f, \epsilon)$ is a subbasis of the *topology of compact convergence*; that is, an open set of F^E in this topology is any union (possibly infinite) of finite intersections of subsets of the form $B_K(f, \epsilon)$. The space of continuous functions from E to F with the topology of compact convergence is denoted by $(F^E)_c$.

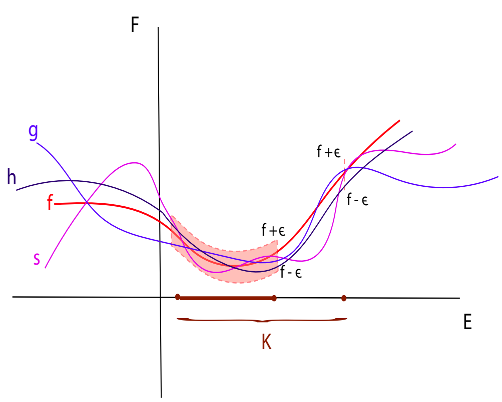


Figure 2.17: Let $E = F = \mathbb{R}$ with the Euclidean metric, and let K be the disjoint union of the brown closed interval and single point. Then $g, h, s \in B_K(f, \epsilon)$

The difference between this topology and the topology of pointwise convergence is that a general basis subset containing a function f contains functions that are close to f not just at finitely many points, but at all points of some compact subset. Thus the topology

of pointwise convergence is weaker than the topology of compact convergence, which itself is weaker than the topology of uniform convergence. It is easy to see the sets $B_K(f, \epsilon)$ actually form a basis of the topology of compact convergence (they are closed under finite intersections).

It is easy to show that a sequence (f_n) of functions in F^E converges to a function f in the topology of compact convergence iff for every compact subset K of E, the sequence (f_n) converges uniformly to f on K.

If the space E is compactly generated, then the topology of compact convergence is even better behaved.

Definition 2.10. A topological space E is *compactly generated* if any subset U of E is open if and only if $U \cap K$ is open in K for every compact subset K.

The following result is shown in Munkres [75] (Chapter 7, Section 46, Lemma 46.3).

Proposition 2.5. *If a topological space E is locally compact or first countable, then it is compactly generated.*

A nice consequence of E being compactly generated is that, as in the case of uniform convergence, the limit of a sequence of continuous functions that converges to a function f in the topology of compact convergence is continuous.

Proposition 2.6. *Let E be a compactly generated topological space and let (F, d) be a metric space. Then the space $\mathcal{C}(E; F)$ of continuous functions from E to F is closed in F^E in the topology of compact convergence.*

Proposition 2.6 is proven in Munkres [75] (Chapter 7, Section 46, Theorem 46.5).

In many applications we are interested in considering the space $\mathcal{C}(E; F)$ of continuous functions from E to F as an independent space in its own right, not as a subspace embedded in F^E. As such there is an *intrinsic* way to define a topology on $\mathcal{C}(E; F)$ which has the advantage of *not* requiring F to be a metric space. Fortunately, as we will discover, and as illustrated in Figure 2.18, if F is a metric space, this intrinsic methodology corresponds to the inherent topology of compact convergence.

Definition 2.11. Let E and F be two topological spaces. For any compact subset K of E and any open subset U of F, let $S(K, U)$ be the set of continuous functions)

$$S(K, U) = \{f \mid f \in \mathcal{C}(E; F),\ f(K) \subseteq U\};$$

see Figure 2.19. The sets $S(K, U)$ form a subbasis for a topology on $\mathcal{C}(E; F)$ called the *compact-open topology*. An open set in the topology is any union (possibly infinite) of finite intersections of subsets of the form $S(K, U)$.

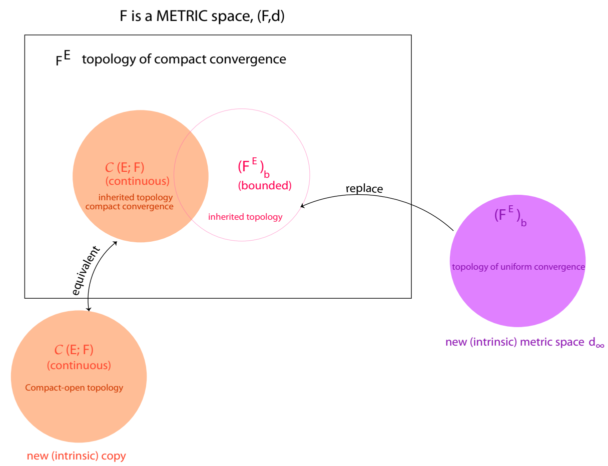


Figure 2.18: A Venn diagram illustration of F^E with the finer topology of compact convergence, along with the subsets $\mathcal{C}(E;F)$ and $(F^E)_b$. There are two equivalent approaches for placing a topology on $\mathcal{C}(E;F)$, an inherited subspace approach and an intrinsic approach. The metric topology on $(F^E)_b$ still requires the intrinsic approach.

It is immediately verified that if F is Hausdorff, then the compact-open topology on $\mathcal{C}(E;F)$ is Hausdorff.

Remark: Observe that the open subsets $S(x,U)$ of the topology of pointwise convergence can be viewed as the result of restricting K to be a single point but relaxing f to belong to F^E.

The compact-open topology is interesting in its own right and coincides with the topology of compact convergence when F is a metric space. The following result is proven in Munkres [75] (Chapter 7, Section 46, Theorem 46.8).

Proposition 2.7. *If E is a topological space and if (F,d) is a metric space, then on the space $\mathcal{C}(E;F)$ of continuous functions from E to F, the compact-open topology and the topology of compact convergence coincide.*

2.5 Equicontinuous Sets of Continuous Functions

Recall that in uniform convergence the limit of a sequence of continuous function is continuous. Another notion that is often useful to show that a sequence of continuous functions converges pointwise to a continuous function is the notion of an equicontinuous set of functions. Intuitively speaking equicontinuity is of sort of uniform continuity for sets of functions.

Definition 2.12. Let E be a topological space and let (F,d_F) be a metric space. A subset $S \subseteq \mathcal{C}(E;F)$ of the set of continuous functions from E to F is *equicontinuous at a point*

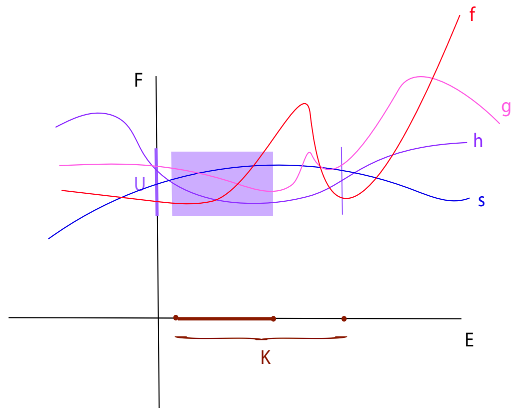


Figure 2.19: Let $E = F = \mathbb{R}$ with the Euclidean metric, and let K be the disjoint union of the brown closed interval and single point. Let U be the purple open interval. Then $f, g, h, s \in S(K, U)$ since each function passes through the light purple region.

$x_0 \in E$ if for every $\epsilon > 0$, there is some open subset $U \subseteq E$ containing x_0 such that $d_F(f(x), f(x_0)) \leq \epsilon$ for all $x \in U$ and for *all* $f \in S$; see Figure 2.20. If E is also a metric space with metric d_E, then the above condition says that for every $\epsilon > 0$ and for *all* $f \in S$, there is some $\eta > 0$ such that $d_F(f(x), f(x_0)) \leq \epsilon$ whenever $d_E(x, x_0) \leq \eta$. The set of functions S is *equicontinuous* if it is equicontinuous at every point $x \in E$.

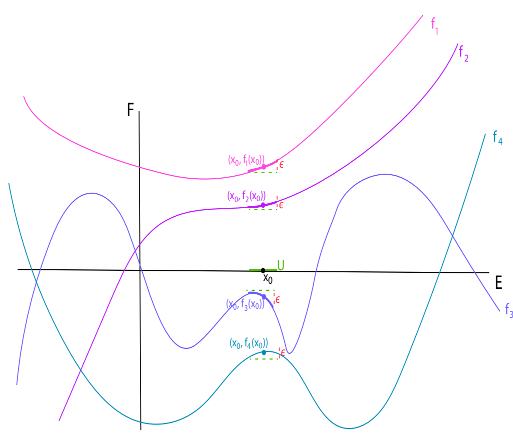


Figure 2.20: Let $E = F = \mathbb{R}$ with the Euclidean metric, and let U be the green open interval containing x_0. The set $S = \{f_1, f_2, f_3, f_4\}$ is equicontinuous at x_0.

For example, if E is a metric space and if there exists two constants $c, \alpha > 0$ such that

we have the Lipschitz condition

$$d_F(f(x), f(y)) \leq c(d_E(x, y))^{\alpha}, \qquad \text{for all } f \in S \text{ and all } x, y \in E,$$

then S is equicontinuous.

Proposition 2.8. *Let (f_n) be a sequence of functions $f_n \in \mathcal{C}(E; F)$, and let (x_n) be a sequence of points $x_n \in E$. If the set $\{f_n\}$ is equicontinuous, the sequence (x_n) converges to $x \in E$, and the sequence (f_n) converges pointwise to some function $f \colon E \to F$, then the sequence $(f_n(x_n))$ converges to $f(x) \in F$.*

Proof. We have, as shown in Figure 2.21, the inequality

$$d_F(f_n(x_n), f(x)) \leq d_F(f_n(x_n), f_n(x)) + d_F(f_n(x), f(x)).$$

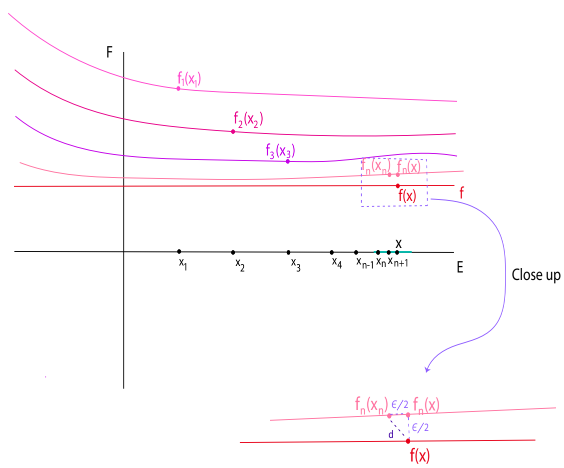


Figure 2.21: An illustration of $d_F(f_n(x_n), f(x)) \leq d_F(f_n(x_n), f_n(x)) + d_F(f_n(x), f(x))$, where $E = F = \mathbb{R}$. For simplicity we suppressed the first coordinate of the ordered pair.

For every $\epsilon > 0$, since the sequence (f_n) converges pointwise to f, there is some $N_2 > 0$ such that $d_F(f_n(x), f(x)) \leq \epsilon/2$ for all $n \geq N_2$. Since $\{f_n\}$ is equicontinuous, there is some open subset $U \subseteq E$ containing x such that

$$d_F(f_n(y), f_n(x)) \leq \epsilon/2 \qquad \text{for all } n \geq 1 \text{ and all } y \in U.$$

Since (x_n) converges to x, there is some $N_1 > 0$ such that $x_n \in U$ for all $n \geq N_1$, so

$$d_F(f_n(x_n), f_n(x)) \leq \epsilon/2 \qquad \text{for all } n \geq N_1,$$

and for all $n \geq \max\{N_1, N_2\}$, we have $d_F(f_n(x_n), f(x)) \leq \epsilon$, which proves that $(f_n(x_n))$ converges to $f(x)$. $\square$

There are various results about equicontinuous sets of functions usually known as variants of *Ascoli's theorem*. Schwartz [83] (Chapter XX) gives one of the most complete expositions we are aware of. We only consider three variants of Ascoli's theorem that we will need.

Theorem 2.9. *(Ascoli I) Let E be a topological space, let (F, d_F) be a metric space, and let $S \subseteq \mathcal{C}(E; F)$ be a set of equicontinuous functions at some $x_0 \in E$. Then the closure $\overline{S}$ of S in F^E with the topology of pointwise convergence is also equicontinuous at x_0. As a corollary, if $S \subseteq \mathcal{C}(E; F)$ is a set of equicontinuous functions, then every function $f \in \overline{S}$ is continuous, and for every sequence (f_n) of functions $f_n \in S$, if (f_n) converges pointwise to a function $f \in F^E$, then f is continuous.*

Proof. Since S is equicontinuous at x_0, for every $\epsilon > 0$, there is some open subset $U \subseteq E$ containing x_0 such that

$$d_F(f(x_0), f(x)) \leq \epsilon, \qquad \text{for all } f \in S \text{ and all } x \in U.$$

But for $x \in U$ fixed, the map $f \mapsto (f(x_0), f(x))$ from F^E to $F^2 = F \times F$ is continuous (this is a projection onto a product), and d_F is continuous on F^2. As a consequence, the set

$$\{f \in F^E \mid d_F(f(x_0), f(x)) \leq \epsilon\}$$

is closed in F^E, and since it contains S, it also contains $\overline{S}$. Thus, for every $\epsilon > 0$, we found an open subset U containing x_0 such that $d_F(f(x_0), f(x)) \leq \epsilon$ for all $x \in U$ and all $f \in \overline{S}$, which means that $\overline{S}$ is equicontinuous.

Since every function in an equicontinuous set of functions is continuous, every function $f \in \overline{S}$ is continuous. By definition of the pointwise topology, if a sequence (f_n) of functions $f_n \in S$ converges pointwise to a function $f \in F^E$, then $f \in \overline{S}$, so f is continuous. $\square$

Dieudonné proves a weaker version of Theorem 2.9, namely that for every subset S of the space of bounded continuous functions $\mathcal{C}_b(E; F)$, if S is equicontinuous, then its closure $\overline{S}$ is also equicontinuous. This is Proposition 7.5.4 in Dieudonné [25] (Chapter 7, Section 5).

The second version of Ascoli's theorem involves a dense subset E_0 of E. We need the following variant of Definition 2.2.

Definition 2.13. The topology of *pointwise convergence in E_0* is the topology on F^E having the sets

$$S(x, U) = \{f \mid f \in F^E,\ f(x) \in U\}, \qquad x \in E_0,\ U \text{ open in } F,$$

as a subbasis.

Theorem 2.10. *(Ascoli II) Let E be a topological space, let (F, d_F) be a metric space, E_0 be a dense subset of E, and $S \subseteq \mathcal{C}(E; F)$ be a set of equicontinuous functions. Then the topology of pointwise convergence in E_0, the topology of pointwise convergence, and the topology of compact convergence (all three topologies being defined in F^E), induce identical topologies on S.*

Theorem 2.10 is proven in Schwartz [83] (Chapter XX, Theorem XX.3.1). The following corollaries of Theorem 2.10 are particularly useful. The first of these two propositions is an immediate consequence of Theorem 2.10.

Proposition 2.11. *Let E be a topological space and let (F, d_F) be a metric space. If a sequence (f_n) of continuous functions $f_n \in \mathcal{C}(E; F)$ converges pointwise to a function $f \in F^E$ and if $\{f_n\}$ is equicontinuous, then f is continuous and the sequence (f_n) converges uniformly to f on every compact subset.*

Proposition 2.12. *Let E be a topological space, E_0 a dense subset of E, and let (F, d_F) be a metric space. If the following properties hold:*

(1) The sequence (f_n) of continuous functions $f_n \in \mathcal{C}(E; F)$ converges pointwise for every $x \in E_0$;

(2) The set $\{f_n\}$ is equicontinuous;

(3) The set $\{f_n(x) \mid n \geq 1\}$ is contained in a complete subset of F for every $x \in E$;

then the sequence (f_n) converges pointwise (for all $x \in E$) to a continuous function f, and (f_n) converges uniformly to f on every compact subset. If F complete, then Condition (3) is automatically satisfied and can be omitted.

Proof. Since by Theorem 2.10, the topology of pointwise convergence on E_0 is identical to the topology of pointwise convergence on E, as the sequence (f_n) converges pointwise for every $x \in E_0$, it also converges pointwise for every $x \in E$. This implies that for every x, the sequence $(f_n(x))$ is a Cauchy sequence in F, but since by (3) the set $\{f_n(x) \mid n \geq 1\}$ is contained in a complete subset of F, the sequence $(f_n(x))$ converges. Thus (f_n) converges pointwise to a function $f \in F^E$, and since $\{f_n\}$ is equicontinuous, by Proposition 2.11, the function f is continuous, and (f_n) converges uniformly to f on every compact subset. $\square$

Dieudonné proves a special case of Proposition 2.12 where E is a metric space, F is a complete normed vector space (a Banach space), the functions f_n are continuous and bounded, and $\{f_n\}$ is equicontinuous; see Proposition 7.5.5 and Proposition 7.5.6 in [25] (Chapter 7, Section 5).

In most applications of Ascoli I and II, E is a metric space and F is a (complete) normed vector space. The following result about sets of continuous linear maps will be needed.

Proposition 2.13. *Let E be a metrizable vector space and F be a normed vector space. A subset of continuous linear maps $S \subseteq \mathcal{L}(E; F)$ is equicontinuous if and only if there is some open subset $V \subseteq E$ containing 0 and some real $c > 0$ such that $\|f(x)\| \leq c$ for all $x \in V$ and all $f \in S$.*

Proof. If S is equicontinuous, then obviously the property of the proposition holds. Conversely, for any $\epsilon > 0$, the condition $\|f(x)\| \leq c$ for all $x \in V$ and all $f \in S$ implies that $\|f(x)\| \leq \epsilon$ for all $x \in (\epsilon/c)V$ and all $f \in S$, so S is equicontinuous at 0. For any $x_0 \in E$, and for all $x \in x_0 + (\epsilon/c)V$, we have

$$\|f(x) - f(x_0)\| = \|f(x - x_0)\| \leq \epsilon$$

for all $f \in S$, that is, S is equicontinuous at x_0. □

A third version of Ascoli's theorem involving relative compactness will be needed in Section 17.1. Recall that a subset A of a Hausforff space X is relatively compact if its closure $\overline{A}$ is compact in X.

Theorem 2.14. *(Ascoli III) Let E be a topological space, let (F, d_F) be a metric space, and let $S \subseteq \mathcal{C}(E; F)$ be a set of continuous functions. Assume the following two conditions hold:*

(1) The set S is equicontinuous.

(2) For every $x \in E$, the set $S(x) = \{f(x) \mid f \in S\}$ is relatively compact in F.

Then the set S is relatively compact in the space $(F^E)_c$ of continuous functions from E to F with the topology of compact convergence. Conversely, if E is locally compact and if the set S is relatively compact in the space $(F^E)_c$, then Conditions (1) and (2) hold.

Proof. A complete proof is given in Schwartz [83] (Chapter XX, Theorem XX.4.1). We only prove the first part of the theorem. The proof uses Tychonoff's powerful product theorem. By hypothesis, for every $x \in E$, the closure $\overline{S(x)}$ of $S(x)$ is compact in F, so by Tychonoff's theorem, the product $\prod_{x \in E} \overline{S(x)}$ is compact in F. By definition of the above product, this means that the set $\widehat{S}$ of functions $f \in F^E$ such that $f(x) \in \overline{S(x)}$ for all $x \in E$ is compact in F^E with the topology of pointwise convergence. Since S is contained in the compact set $\widehat{S}$, we deduce that its closure $\overline{S}$ is compact in F^E (with the topology of pointwise convergence). By Ascoli I (Theorem 2.9), since S is equicontinuous, the set $\overline{S}$ is also equicontinuous. By Ascoli II (Theorem 2.10), since the restriction to $\overline{S}$ of the topology of pointwise convergence on F^E coincides with the restriction to $\overline{S}$ of the topology of compact convergence on F^E, the set $\overline{S}$ is compact in $(F^E)_c$, and thus S is relatively compact in $(F^E)_c$. □

The special case of Theorem 2.14 in which E is compact and F is a Banach space is proven in Dieudonné [25] (Chapter 7, Section 5, Theorem 5.7.5). Because F is complete the proof is simpler and does not use Tychonoff's theorem.

2.6 Continuous Functions of Compact Support

In this section we consider F to be a normed vector space. We know that two important subspaces of F^E are $(F^E)_b$, the space of bounded functions, and $\mathcal{C}(E; F)$, the space of continuous functions. The intersection of these subspaces, with the inherited sup norm of $(F^E)_b$, is the space of continuous bounded functions $\mathcal{C}_b(E; F)$. Within $\mathcal{C}_b(E; F)$ there is another interesting subspace, namely $\mathcal{C}_c(E; F)$, the space of continuous functions of compact support. In this section we investigate $\mathcal{C}_c(E; F)$ and describe its closure within $\mathcal{C}_b(E; F)$. So first we recall what is the support of a function.

Definition 2.14. Given any function $f\colon E \to F$, where E is a topological space and F is a vector space, the support $\mathrm{supp}(f)$ of f is the closure of the subset of E where f is nonzero, that is, $\mathrm{supp}(f) = \overline{\{x \in E \mid f(x) \neq 0\}}$. The function f has *compact support* if its support $\mathrm{supp}(f)$ is compact. If E is Hausdorff, this is equivalent to saying that f vanishes outside some compact subset K of E. See Figure 2.22.

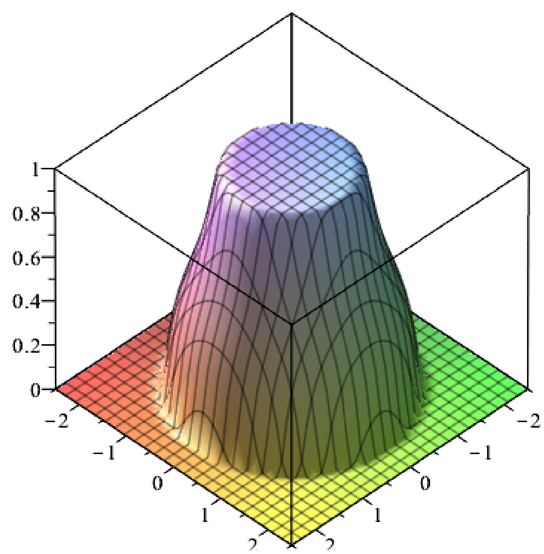


Figure 2.22: The graph of $f\colon \mathbb{R}^2 \to \mathbb{R}$ with compact support $\mathrm{supp} = \overline{B(0,2)} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 2\}$.

It is easy to see that the set of continuous functions $f\colon E \to F$ with compact support is a vector space.

Definition 2.15. The vector space of continuous functions $f\colon E \to F$ with compact support is denoted by $\mathcal{C}_c(E; F)$, or $\mathcal{K}(E; F)$. For every compact subset K of E, we denote by $\mathcal{K}(K; F)$ the space of continuous functions whose support is contained in K. Then

$$\mathcal{K}(E; F) = \bigcup_{K \subseteq E,\ K \text{ compact}} \mathcal{K}(K; F).$$

Observe that every function in $\mathcal{K}(E;F)$ is bounded, that is, $\mathcal{K}(E;F) \subseteq \mathcal{C}_b(E;F)$.

If $F = \mathbb{R}$ or $F = \mathbb{C}$, then we write $\mathcal{K}_{\mathbb{R}}(E)$ or $\mathcal{K}_{\mathbb{C}}(E)$ for $\mathcal{K}(E;F)$. Radon functionals are certain kinds of linear forms on $\mathcal{K}_{\mathbb{C}}(E)$.

The following results will be needed in Chapter 12.

Proposition 2.15. *If E is a compact metric space, then the spaces $\mathcal{C}_{\mathbb{R}}(E)$ and $\mathcal{C}_{\mathbb{C}}(E)$ are separable.*

Proof sketch. The proof is nontrivial and can be found in Dieudonné [25] (Chapter 7, Theorem 7.4.4). The proof makes a crucial use of the Stone–Weierstrass Theorem (Theorem 9.36). The first step is to observe that it suffices to prove that $\mathcal{C}_{\mathbb{R}}(E)$ is separable because $\mathcal{C}_{\mathbb{C}}(E)$ is the direct (topological) sum of $\mathcal{C}_{\mathbb{R}}(E)$ and $i\mathcal{C}_{\mathbb{R}}(E)$. As a second step, we observe that by Proposition A.47, since E is a compact metric space, it is separable, and by Proposition A.46, a separable metric space is second countable. Thus there is a countable base (U_n) for the topology. Then the trick is to define the family of continuous functions $g_n(t) = d(t, E - U_n)$ (see Definition A.5 for the definition of the distance to a subset). The next step is to define the subalgebra B of $\mathcal{C}_{\mathbb{R}}(E)$ generated by the monomials $g_{i_1}^{m_1}(t) \cdots g_{i_k}^{m_k}(t)$ and to check that B satisfies the hypotheses of the Stone–Weierstrass Theorem (Theorem 9.36). The final step is to show that by using rational linear combinations of the monomials $g_{i_1}^{m_1}(t) \cdots g_{i_k}^{m_k}(t)$ we obtain a countable dense subset of $\mathcal{C}_{\mathbb{R}}(E)$ (see Dieudonné [25] (Chapter 5, Theorem 5.10.1)). $\square$

Proposition 2.16. *If E is a locally compact separable metric space, then the spaces $\mathcal{K}_{\mathbb{R}}(E)$ and $\mathcal{K}_{\mathbb{C}}(E)$ are separable.*

Proof sketch. A proof is implicitly given in Dieudonné [24] (Chapter XIII, Theorem 13.11.6)). As in the proof of Proposition 2.15 it suffices to prove our result for $\mathcal{K}_{\mathbb{R}}(E)$. By Proposition A.49(1), since E is locally compact, metric, and separable, there is a countable sequence (K_n) of compact subsets of E such that $K_n \subseteq K_{n+1}$ and $E = \bigcup_{n \geq 1} K_n$. Then

$$\mathcal{K}_{\mathbb{R}}(E) = \bigcup_{n \geq 1} \mathcal{K}_{\mathbb{R}}(K_n).$$

By Proposition 2.15, for each $n \geq 1$, there is a dense sequence $(g_{m,n})_{n \geq 1}$ in $\mathcal{K}_{\mathbb{R}}(K_n)$. Then the countable double sequence $(g_{m,n})$ is dense in $\mathcal{K}_{\mathbb{R}}(E)$. $\square$

If $(F, \| \|)$ is a Banach space and K is a fixed compact subset of E, then so is $\mathcal{K}(K;F)$ (for the sup norm $\| \|_\infty$), because it is closed in $\mathcal{C}_b(E;F)$. However, the normed vector space $(\mathcal{K}(E;F), \| \|_\infty)$ is *not* complete!

Example 2.3. For every $n \geq 1$, consider the function $u_n \colon \mathbb{R} \to \mathbb{R}$ defined as follows:

$$u_n(x) = \begin{cases} 1 & \text{if } -n \leq x \leq n \\ x + n + 1 & \text{if } -(n+1) \leq x \leq -n \\ -x + n + 1 & \text{if } n \leq x \leq n + 1 \\ 0 & \text{if } |x| \geq n + 1. \end{cases}$$

Now consider the sequence of functions (f_n) given by

$$f_n(x) = u_n e^{-|x|}.$$

Each function f_n is continuous and has compact support $[-(n+1), n+1]$, and it is easy to show that the sequence (f_n) converges uniformly to the function f given by $f(x) = e^{-|x|}$, but f does not have compact support. The problem is that the domains of the functions f_n, although compact, keep growing as n goes to infinity. See Figure 2.23.

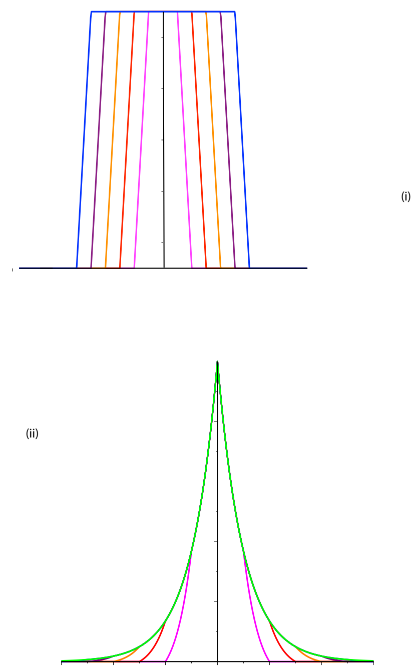


Figure 2.23: The functions of Example 2.3. Figure (i) illustrates the $u_1(x)$ in magenta; $u_2(x)$ in red, $u_3(x)$ in orange, $u_4(x)$ in purple, and $u_5(x)$ in blue. Figure (ii) uses the same color scheme to illustrate the corresponding $f_n(x)$. Note these $f_n(x)$ converge uniformly to green $f(x) = e^{-|x|}$.

Example 2.3 shows that the normed vector space $(\mathcal{K}(E;F), \| \|_\infty)$ is *not closed* in the complete normed vector space $(\mathcal{C}_b(E;F), \| \|_\infty)$. It would be useful to identify the closure $\overline{\mathcal{K}(E;F)}$ of $\mathcal{K}(E;F)$ in $\mathcal{C}_b(E;F)$, and this can indeed be done when E is locally compact.

Assume that f belongs to the closure $\overline{\mathcal{K}(E;F)}$ of $\mathcal{K}(E;F)$. This means that there is a sequence (f_n) of functions $f_n \in \mathcal{K}(E;F)$ such that $\lim_{n\mapsto\infty} \|f - f_n\|_\infty = 0$, so for every $\epsilon > 0$, there is some $n \geq 1$ such that $\|f(x) - f_n(x)\| \leq \epsilon$ for all $x \in E$, and since f_n has compact support, there is some compact subset K of E such that $\|f(x)\| \leq \epsilon$ for all $x \in E - K$. This suggests the following definition.

Definition 2.16. The subspace of $\mathcal{C}_b(E;F)$, denoted $\mathcal{C}_0(E;F)$, consisting of the continuous functions f such that for every $\epsilon > 0$, there is some compact subset K of E such that $\|f(x)\| \leq \epsilon$ for all $x \in E - K$, is called the space of *continuous functions which tend to* 0 *at infinity*; see Figure 2.24.

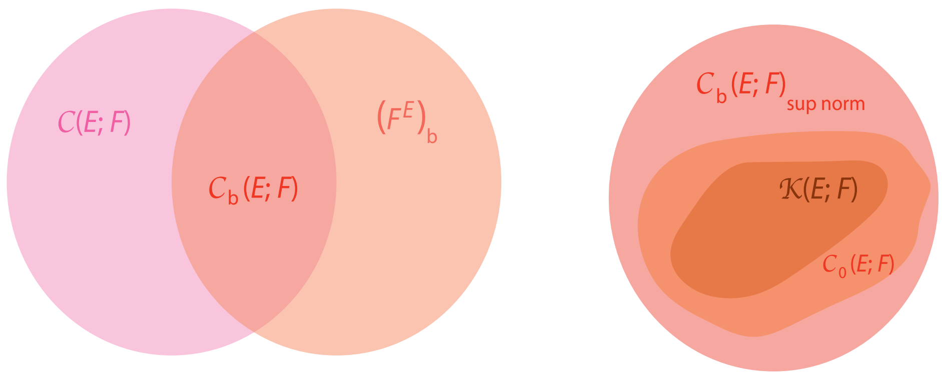


Figure 2.24: The Venn diagram relationships between $\mathcal{C}_b(E;F) = (F^E)_b \cap \mathcal{C}(E;F)$, and the subspaces $\mathcal{K}(E;F)$ and $\mathcal{C}_0(E;F)$, where $\mathcal{K}(E;F) \subseteq \mathcal{C}_0(E;F) \subseteq \mathcal{C}_b(E;F)$.

If E is compact, we can pick $K = E$, in which case $E - K = \emptyset$. This shows that Definition 2.16 has been designed so that if E is compact, then $\mathcal{C}_0(E;F) = \mathcal{C}(E;F) = \mathcal{K}(E;F)$.

Observe that if $E = \mathbb{R}$, then a function $f \in \mathcal{C}_0(\mathbb{R};F)$ does indeed have the property that $\lim_{x \mapsto -\infty} f(x) = \lim_{x \mapsto +\infty} f(x) = 0$; see Figure 2.25.

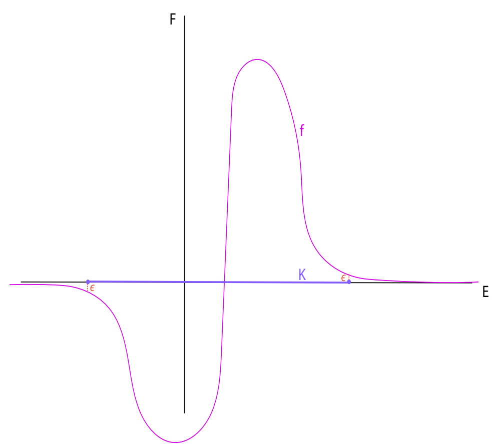


Figure 2.25: A schematic illustration of $f \in \mathcal{C}_0(\mathbb{R};F)$, where the reader may consider $F = \mathbb{R}$.

We showed that $\mathcal{K}(E;F) \subseteq \mathcal{C}_0(E;F)$. If E is locally compact, then we have the following result from Dieudonné [24] (Chapter XIII, Section 20) and Rudin [79] (Chapter 3, Theorem 3.17).

Proposition 2.17. *If E is locally compact, then $\mathcal{C}_0(E;\mathbb{C})$ is the closure of $\mathcal{K}(E;\mathbb{C})$ in $\mathcal{C}_b(E;\mathbb{C})$. Consequently, $\mathcal{C}_0(E;\mathbb{C})$ is complete.*

Proof. We already showed just before Definition 2.16 that if a function f belongs to the closure of $\mathcal{K}(E;\mathbb{C})$, then it tends to zero at infinity. Conversely, pick any f in $\mathcal{C}_0(E;\mathbb{C})$. For every $\epsilon > 0$, there is a compact subset K of E such that $|f(x)| < \epsilon$ outside of K. By Proposition A.39, there is continuous function $g\colon E \to [0,1]$ with compact support such that $g(x) = 1$ for all $x \in K$. Clearly $fg \in \mathcal{K}_{\mathbb{C}}(E)$, and $\|fg - f\|_\infty < \epsilon$. This shows that $\mathcal{K}(E;\mathbb{C})$ is dense in $\mathcal{C}_0(E;\mathbb{C})$. $\square$

In summary, if E is locally compact, then we have the inclusions

$$\mathcal{K}(E;\mathbb{C}) \subseteq \mathcal{C}_0(E;\mathbb{C}) \subseteq \mathcal{C}_b(E;\mathbb{C}),$$

with $\mathcal{C}_0(E;\mathbb{C})$ and $\mathcal{C}_b(E;\mathbb{C})$ complete, and $\mathcal{K}(E;\mathbb{C})$ dense in $\mathcal{C}_0(E;\mathbb{C})$. If E is not compact, these inclusions are strict in general. It turns out that the space of continuous linear forms on $\mathcal{C}_0(E;\mathbb{C})$ is isomorphic to the space of bounded Radon functionals.

2.7 Topologies Defined by Semi-Norms; Fréchet Spaces

Certain function spaces, such as the space $\mathcal{C}(X;\mathbb{C})$ of continuous functions on a topological space X, do not come with “natural” topologies defined by a norm or a metric for which they are complete. However, the weaker notion of semi-norm can be used to define a topology, and under certain conditions, although such topologies are not defined by any norm, they are metrizable and complete. In this section we briefly discuss the use of semi-norms to define topologies. It turns out that the corresponding spaces are locally convex.

Recall from Definition B.1 that a semi-norm satisfies Properties (N2) and (N3) of a norm, but in general does not satisfy Condition (N1), so $\|x\| = 0$ does not necessarily imply that $x = 0$. Here is a method for defining a topology on a vector space using a family of semi-norms.

Definition 2.17. Let X be a vector space and let $(p_\alpha)_{\alpha\in I}$ be a family of semi-norms on X. For every $x \in X$, every $\epsilon > 0$, and every $\alpha \in I$, let

$$U_{x,\alpha,\epsilon} = \{y \in X \mid p_\alpha(y - x) < \epsilon\}.$$

The *topology induced by the family of semi-norms* $(p_\alpha)_{\alpha\in I}$ is the weakest (coarsest) topology whose open sets are arbitrary unions of finite intersections of subsets of the form $U_{x,\alpha,\epsilon}$.

We can think of the subset $U_{x,\alpha,\epsilon}$ as an open ball of center x and radius ϵ in X, determined by the semi-norm p_α.

Two good examples of topologies induced by families of semi-norms are the topology of pointwise convergence and the topology of compact convergence on a normed vector space F.

Example 2.4. Let E be any set and let F be a normed vector space. If we define the family of semi-norms $(p_x)_{x\in E}$ by

$$p_x(f) = \|f(x)\|, \quad f \in F^E,\ x \in E,$$

then it is easy to see that the topology defined by the family $(p_x)_{x\in E}$ is the topology of pointwise convergence on F^E, which has the subsets

$$S(x, U) = \{f \mid f \in F^E,\ f(x) \in U\}, \qquad x \in E,\ U \text{ open in } F,$$

as a subbasis.

Example 2.5. Let E be a topological space and let F be a normed vector space. If we define the family of semi-norms $\{p_K \mid K \text{ compact in } E\}$, by

$$p_K(f) = \sup_{x\in K} \|f(x)\|, \quad f \in F^E,\ K \text{ compact in } E,$$

then it is easy to see that the topology defined by the family (p_K) is the topology of compact convergence on F^E, which has the subsets

$$B_K(f, \epsilon) = \left\{ g \in F^E \mid \sup_{x\in K} d(f(x), g(x)) < \epsilon \right\}$$

as a subbasis.

We have made our vector space X into a topological space but it is not clear that the operations (addition and scalar multiplication) are continuous. Also, in general, this topology is not Hausdorff. The following proposition addresses these issues.

Proposition 2.18. *Let X be a vector space and let $(p_\alpha)_{\alpha\in I}$ be a family of semi-norms on X.*

(1) With the topology induced by the family of semi-norms $(p_\alpha)_{\alpha\in I}$, addition and scalar multiplication are continuous, so X is a topological vector space.

(2) For every $x \in X$, the finite intersections of subsets of the form $U_{x,\alpha,\epsilon}$ is a neighborhood base of x.

(3) Every open set $U_{x,\alpha,\epsilon}$ is convex.

(4) Every p_α is continuous.

(5) The topology induced by the family of semi-norms is Hausdorff if and only if, for every $x \neq 0$, there is some $\alpha \in I$ such that $p_\alpha(x) \neq 0$.

Proposition 2.18 is proven in Folland [34] (Chapter 5, Section 5.4, Theorem 5.14), or Rudin [80] (Chapter 1, Theorem 1.37). In view of Property (3), the topological space X is said to be *locally convex*.

In a vector space X whose topology defined by a family of semi-norms $(p_\alpha)_{\alpha\in I}$ is Hausdorff, it is easy to see that the convergence of a sequence (x_n) to a limit x is expressed conveniently as follows.

Proposition 2.19. *Let X be a space whose topology is defined by a family of semi-norms $(p_\alpha)_{\alpha\in I}$. If X is Hausdorff, then a sequence (x_n) converges to a limit x iff for every $\alpha \in I$, for every $\epsilon > 0$, there is some $N_\alpha > 0$ such that $p_\alpha(x - x_n) \leq \epsilon$ for all $n \geq N_\alpha$; equivalently, $\lim_{n\mapsto\infty} p_\alpha(x - x_n) = 0$, for every $\alpha \in I$.*

When the index family I is countable and the topology induced by a family of semi-norm is Hausdorff, then X is actually metrizable.

Proposition 2.20. *Let X be a vector space and let $(p_\alpha)_{\alpha\in I}$ be a family of semi-norms on X. If the topology induced by $(p_\alpha)_{\alpha\in I}$ is Hausdorff and if I is countable, then X is metrizable with a translation-invariant metric d (this means that $d(a, b) = d(a+u, b+u)$ for all $a, b, u \in X$). In fact, we can use the metric d given by*

$$d(x, y) = \sum_{m=0}^{\infty} \frac{1}{2^m} \frac{p_m(y - x)}{1 + p_m(y - x)}.$$

Proposition 2.18 is proven in Dieudonné [24] (Chapter 12, Section 4, Theorem 12.4.6), and in Rudin [80] (Chapter 1, Page 29, with an equivalent metric).

Definition 2.18. A vector space X whose topology is defined by a countable family of semi-norms, and which is Hausdorff and complete for some translation-invariant metric defining the topology of X is called a *Fréchet space*.

A prime example of a Fréchet space is the space $\mathcal{C}(X; \mathbb{C})$ of continuous functions on a separable, locally compact, metrizable space X. This will be proven shortly.

The following technical result is needed to prove Proposition 2.22.

Proposition 2.21. *Let X be a metrizable Hausdorff topological vector space. For any translation-invariant metric d defining the topology of X, a sequence (x_n) is a Cauchy sequence if and only if for every neighborhood V of 0, there is some $N > 0$ such that $x_m - x_n \in V$ for all m, n such that $m \geq N$ and $n \geq N$.*

Proof. A slightly more general result is proven for topological groups in Dieudonné [24] (Chapter 12, Section 9, Theorem 12.9.2) and Rudin [80] (Chapter 1, Page 21). If a metric d defining the topology of X is translation-invariant, then

$$d(x_n, x_m) = d(0, x_m - x_n),$$

and the sequence (x_n) is a Cauchy sequence iff for every $\epsilon > 0$, there is some $N > 0$ such that $d(0, x_m - x_n) < \epsilon$ for all $m \geq N$ and $n \geq N$, which is equivalent to saying that $x_m - x_n \in V$, where V is the open ball of center 0 and radius ϵ, which is an open subset of X, by definition of the metric topology. Conversely, since the topology of X is defined by the metric d, every open ball of center 0 is an open set, so the condition of the proposition implies that (x_n) is a Cauchy sequence for every translation-invariant metric defining the topology of X. $\square$

Proposition 2.22. *If a metrizable topological vector space X is Hausdorff and complete for some translation-invariant metric d defining the topology of X, then it is also complete for every translation-invariant metric d' defining the topology of X,*

We now prove that the space $\mathcal{C}(X; \mathbb{C})$ of continuous functions on a separable, locally compact, metrizable space X is a Fréchet space.

Recall from Proposition A.49 that since X is metrizable, there is a sequence $(U_n)_{n \geq 0}$ of open subsets such that for all $n \in \mathbb{N}$, $U_n \subseteq U_{n+1}$, $\overline{U_n}$ is compact, $\overline{U_n} \subseteq U_{n+1}$, and $X = \bigcup_{n \geq 0} U_n = \bigcup_{n \geq 0} \overline{U_n}$. For every $n \in \mathbb{N}$, define the function $p_n \colon \mathcal{C}(X; \mathbb{C}) \to \mathbb{R}$ by

$$p_n(f) = \sup_{x \in U_n} |f(x)|, \qquad f \in \mathcal{C}(X; \mathbb{C}).$$

It is immediately verified that the p_n are semi-norms (but none of the p_n are norms if X is not compact). For each $f \in \mathcal{C}(X; \mathbb{C})$, if $f \neq 0$, then there is some n such that $x \in U_n$, hence $p_n(f) \neq 0$. Thus, by Proposition 2.18(5), the space $\mathcal{C}(X; \mathbb{C})$ with the topology induced by the family of semi-norms (p_n) is Hausdorff. By Proposition 2.20, this topology is metrizable. Note that the restriction of p_{n+1} to the compact subset $\overline{U_n}$ is actually a norm, and by definition of the metric d given by Proposition 2.20, the restriction of d to $\overline{U_n}$ is equivalent to p_{n+1}.

Proposition 2.23. *Let X be a separable, locally compact, metrizable space. The space $\mathcal{C}(X; \mathbb{C})$ with the topology induced by the family of semi-norms (p_n) is complete. Therefore, it is a Fréchet space.*

Proof. Since the restriction of the metric d to $\overline{U_n}$ is equivalent to p_{n+1}, by Proposition 2.22, a sequence (f_k) of functions in $\mathcal{C}(X; \mathbb{C})$ is a Cauchy sequence if for every n, the sequence of restrictions $f_k | \overline{U_n}$ is a Cauchy sequence in the Banach space $\mathcal{C}(\overline{U_n}; \mathbb{C})$, hence converges uniformly in $\overline{U_n}$ to a continuous function $g_n \in \mathcal{C}(\overline{U_n}; \mathbb{C})$. Since $g_{n+1} | \overline{U_n} = g_n$, there exists a continuous function $f \in \mathcal{C}(X; \mathbb{C})$ whose restriction to each U_n agrees with the restriction of g_n to U_n; see Figure 2.26. It is clear that $\lim_{m \mapsto \infty} p_n(f - f_m) = 0$ for all $n \geq 0$, hence by Proposition 2.19, f is the limit of the Cauchy sequence (f_k), and $\mathcal{C}(X; \mathbb{C})$ is complete. $\square$

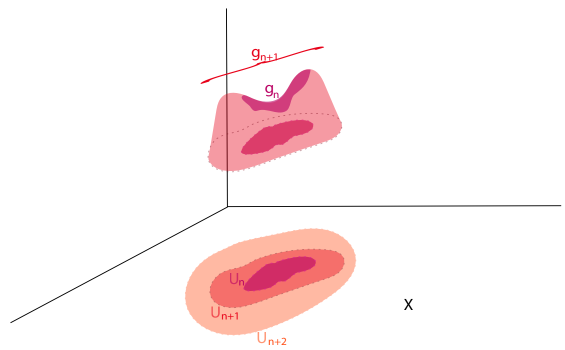


Figure 2.26: A schematic illustration of the function g_n and its continuous extension g_{n+1}. In this figure X is represented by the horizontal plane and $\mathbb{C}$ is the vertical axis. The graph of g_{n+1} is the dusty rose surface while the graph of g_n is the plum surface patch inside of that surface.

It is shown in Rudin [80] (Chapter 1, Example 1.44) that the Fréchet space $\mathcal{C}(X;\mathbb{C})$ is not normable.

The following result is shown in Dieudonné [24] (Chapter 12, Section 14, Theorem 12.14.6.2).

Proposition 2.24. *Let X be a separable, locally compact, metrizable space. The Fréchet space $\mathcal{C}(X;\mathbb{C})$ is separable. In fact, there is a countable dense set consisting of continuous functions with compact support.*

Another good example of a Fréchet space is the Schwartz space; see Section 6.8.

2.8 Regulated Functions

In the last two sections we focused on $\mathcal{C}(E;F)$ where F is a normed vector space. We return to $(F^E)_b$ and in preparation for the next chapter on the Riemann integral investigate two important subspaces of $(F^{\mathbb{R}})_b$, the space of regulated functions and then the space of step functions, both of which inherit the sup norm from $(F^E)_b$. Since the space of regulated functions contains the space of step functions we begin with the definition of the larger subspace. Recall that there are four kinds of intervals of $\mathbb{R}$: (a,b), $[a,b)$, $(a,b]$, and $[a,b]$, with $a<b$. By convention, $(a,b)=[a,b)$ if $a=-\infty$, and $(a,b)=(a,b]$ if $b=\infty$.

Definition 2.19. Let I be an interval of $\mathbb{R}$, and let F be a metric space (or a normed vector space). Given a function $f\colon I\to F$, for any $x\in I$ with $x\neq b$, we say that f *has a limit to the right in x* if $\lim_{y\in I,\, y>x} f(y)$ exists as $y\in I$ tends to x from above. This limit is denoted by $f(x+)$. For any $x\in I$ with $x\neq a$, we say that f *has a limit to the left in x* if $\lim_{y\in I,\, y<x} f(y)$

exists as $y \in I$ tends to x from below. This limit is denoted by $f(x-)$. Given any interval I, a function $f\colon I \to F$ is a *regulated function* (or *ruled function*) if it has a left limit and a right limit for every $x \in I$. If F is a metric space (or a normed vector space), a function $f\colon \mathbb{R} \to F$ is a *regulated function* (or *ruled function*) if there is some interval I such that f vanishes outside I, and the restriction $f\colon I \to F$ of f to I is regulated. See Figure 2.27.

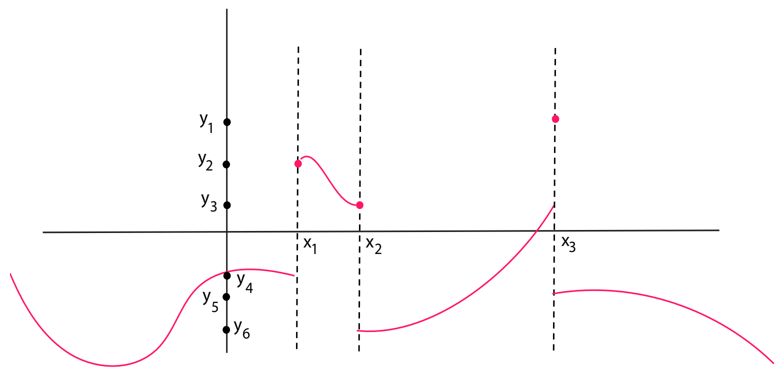


Figure 2.27: An illustration of a regulated function $f\colon \mathbb{R} \to \mathbb{R}$. This function has three discontinuities x_1, x_2, and x_3, each of the first kind. Note that $f(x_1-) = y_4$, $f(x_1+) = f(x_1) = y_2$, $f(x_2-) = f(x_2) = y_3$, $f(x_2+) = y_6$, $f(x_3-) = y_3$, $f(x_3+) = y_5$, yet $f(x_3) = y_1$.

The notion of a regulated function can also be defined in terms of certain kinds of discontinuities.

Definition 2.20. Let I be an interval of $\mathbb{R}$, and let F be a metric space (or a normed vector space). Given a function $f\colon I \to F$, we say that a point $x \in I$ is a *discontinuity of the first kind* if the left limit $f(x-)$ and the right limit $f(x+)$ both exist, but $f(x-) \neq f(x)$ or $f(x+) \neq f(x)$.

Is clear that a function $f\colon I \to F$ is regulated iff for every $x \in I$, either f is continuous or x is a discontinuity of the first kind. Thus every continuous function is a regulated function. It is also easy to see that a monotonic function $f\colon I \to \mathbb{R}$ is a regulated function.

The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is discontinuous at $x = 0$, but this is not a discontinuity of the first kind. See Figure 2.28.

The following result is shown in Schwartz [85] (Chapter III, Section 2, Theorem 3.2.3).

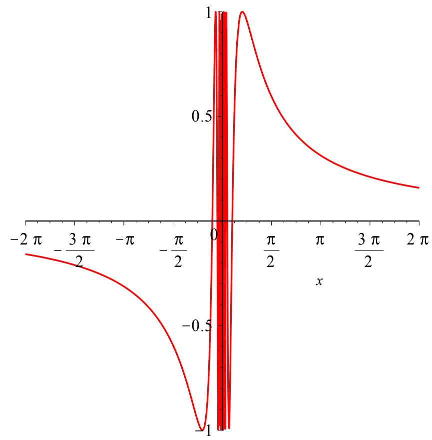


Figure 2.28: The graph of $f(x) = \sin\left(\frac{1}{x}\right)$, $x \neq 0$.

Proposition 2.25. *If $f \colon I \to F$ is a regulated function (where F is a metric space), then f has at most countably many discontinuities of the first kind.*

Regulated functions on a closed and bounded interval $[a, b]$ must be bounded. As a consequence, they arise as limits of uniformly convergent sequences of step functions.

Definition 2.21. A function $f \colon \mathbb{R} \to F$ (where F is any set) is a *step function* if there is a finite sequence $(a_0, a_1, \ldots, a_n)$ of reals such that $a_k < a_{k+1}$ for $k = 0, \ldots, n-1$, and f is constant on each of the open intervals $(-\infty, a_0)$, (a_k, a_{k+1}) for $k = 0, \ldots, n-1$, and $(a_n, +\infty)$. The sequence $(a_0, a_1, \ldots, a_n)$ is called an *admissible subdivision* for f. See Figure 2.29. If a step function f has compact support, then we assume that f vanishes on $(-\infty, a_0)$ and on $(a_n, +\infty)$ for any admissible subdivision $(a_0, a_1, \ldots, a_n)$ for f. By a step function $f \colon [a, b] \to F$, we mean a step function such that $f(x) = 0$ for all $x \leq a$ and for all $x \geq b$.

Observe that Definition 2.21 does not make any restriction on the values $f(a_k)$, but a step function is regulated. Also, by refining a given subdivision, a given step function admits infinitely many admissible subdivisions.

The following result is easy to prove.

Proposition 2.26. *If F is a vector space, then the set of step functions $f \colon \mathbb{R} \to F$ is a vector space denoted by* $\mathrm{Step}(\mathbb{R}; F)$. *The set of step functions $f \colon [a, b] \to F$ is also vector space denoted by* $\mathrm{Step}([a, b]; F)$.

The following proposition is much more interesting.

Proposition 2.27. *Let F be a metric space and let $[a, b]$ be a closed and bounded interval. Then every regulated function $f \colon [a, b] \to F$ is the limit of a uniformly convergent sequence*

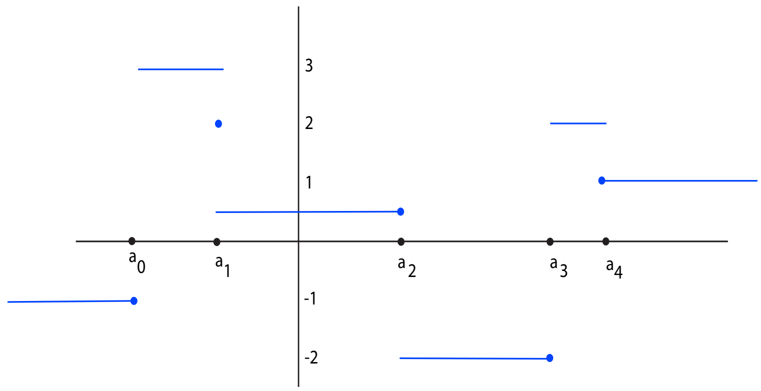


Figure 2.29: An illustration of a step $f\colon \mathbb{R} \to \mathbb{R}$ with admissible subdivision $(a_0, a_1, a_2, a_3, a_4)$.

$(f_n)_{n\geq 1}$ of step functions $f_n\colon [a,b] \to F$. Furthermore, if F is a complete metric space, then the limit of any uniformly convergent sequence $(f_n)_{n\geq 1}$ of step functions is a regulated function.

The proof of Proposition 2.27 is given in Schwartz [85] (Chapter III, Section 2, Theorem 3.2.9).

As a corollary of Proposition 2.27 we have the following result.

Proposition 2.28. *If F is a complete metric space, then the space of regulated functions on $[a,b]$ is closed in $(F^{[a,b]})_b$, and the space of step functions on $[a,b]$ is dense in the space of regulated functions on $[a,b]$. Thus if F is complete, since $(F^{[a,b]})_b$ is complete, the space of regulated function on $[a,b]$ is also complete.*

Another corollary of Proposition 2.27 is that every continuous function $f\colon [a,b] \to F$ to a metric space F is the limit of a uniformly convergent sequence $(f_n)_{n\geq 1}$ of step functions $f_n\colon [a,b] \to F$.

If F is a vector space, the set of regulated functions defined on the closed and bounded interval $[a,b]$ is a vector space denoted by $\mathrm{Reg}([a,b];F)$. Then Proposition 2.27 implies the following result.

Proposition 2.29. *Let F be a complete normed vector space. The space $\mathrm{Reg}([a,b];F)$ of regulated functions on $[a,b]$ is complete, and the space $\mathrm{Step}([a,b];F)$ is dense in $\mathrm{Reg}([a,b];F)$.*

Step functions can be used to define the Riemann integral. To do so it is convenient to consider functions of finite support. Furthermore, a modified version of step functions involving a measure will be used to define the integral on a measure space.

Chapter 3

The Riemann Integral

Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function. Intuitively, the Riemann integral $\int_a^b f(t)dt$ is the area of the surface "under the curve" $t \mapsto f(t)$ from $x = a$ to $x = b$. It can be approximated by the sum $s_T(f)$ (called *Cauchy-Riemann sum*) of the areas $(t_{k+1} - t_k)f(t_k)$ of $n \geq 1$ narrow rectangles, where $T = (t_0, t_1, \ldots, t_n)$ is any sequence of reals such that $t_0 = a$, $t_n = b$ and $t_k < t_{k+1}$, for $k = 0, \ldots, n-1$; see Figure 3.1. The fact that the function f is continuous on the compact interval $[a,b]$ implies that the sums $s_T(f)$ have a limit when the diameter of the subdivision tends to zero (see Definition 3.1), which means the maximum of the distances $t_{k+1} - t_k$ tends to zero (as n goes to infinity), and this limit is independent of the subdivision. Thus we can define the Riemann integral $\int_a^b f(t)dt$ as this common limit. The mapping $f \mapsto \int_a^b f(t)dt$ is a positive linear form on the space of continuous functions on $[a,b]$. This procedure applies unchanged to continuous functions $f\colon [a,b] \to F$, where F is a complete normed vector space.

The method for constructing the integral of a continuous function can be adapted to define the integral of regulated functions (see Definition 2.19). We proceed in two steps:

(1) The method of Cauchy-Riemann sums is easily adapted to define the notion of integral for a step function (see Definition 2.21). This yields a mapping $\int\colon \mathrm{Step}([a,b];F) \to F$ which is easily seen to be linear and continuous.

(2) By Proposition 2.29, the vector space $\mathrm{Step}([a,b];F)$ of step functions over $[a,b]$ is dense in $\mathrm{Reg}([a,b];F)$, the space of regulated functions over $[a,b]$, and $\mathrm{Reg}([a,b];F)$ is complete. By Theorem A.73, the continuous linear map $\int\colon \mathrm{Step}([a,b];F) \to F$ has a unique extension $\int\colon \mathrm{Reg}([a,b];F) \to F$ to $\mathrm{Reg}([a,b];F)$, which is also continuous and linear. This is how the integral of a regulated function is defined.

In summary, we define an "obvious" notion of integral on the simple set $\mathrm{Step}([a,b];F)$. It is a linear and continuous mapping, so we extend it by continuity to the bigger space $\mathrm{Reg}([a,b];F)$ in which $\mathrm{Step}([a,b];F)$ is dense.

3.1 Riemann Integral of a Continuous Function

In this section we define the Riemann integral of a real-valued continuous function.

Definition 3.1. Let $a < b$ be any two reals. A set $T = \{t_0, t_1, \ldots, t_n\}$ of reals such that $t_0 = a$, $t_n = b$ and $t_k < t_{k+1}$, for $k = 0, \ldots, n-1$, is called a *subdivision* of $[a, b]$. The *diameter* $\delta(T)$ of T is defined by

$$\delta(T) = \max_{0 \le k \le n-1} (t_{k+1} - t_k).$$

Given a continuous function $f\colon [a, b] \to \mathbb{R}$, define the *Cauchy–Riemann sum* $s_T(f)$ by

$$s_T(f) = \sum_{k=0}^{n-1} (t_{k+1} - t_k) f(t_k).$$

See Figure 3.1.

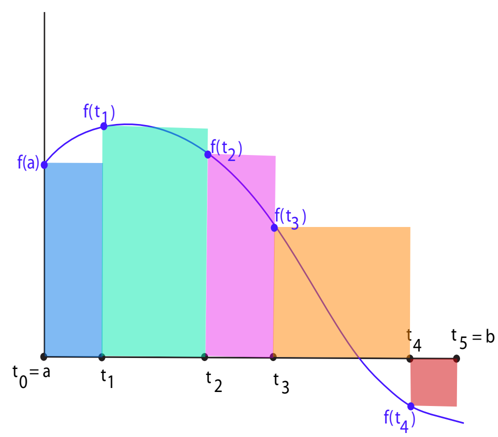


Figure 3.1: The Cauchy–Riemann sum $s_T(f) = \sum_{k=0}^{4} (t_{k+1} - t_k) f(t_k)$ is the "signed" area represented by the pastel shaded boxes.

Observe that

$$\sum_{k=0}^{n-1} (t_{k+1} - t_k) = b - a.$$

We immediately check that s_T is a linear form on the set of continuous functions on $[a, b]$. Furthermore, if $f \ge 0$, which means that $f(t) \ge 0$ for all $t \in [a, b]$, then $s_T(f) \ge 0$.

The question is, as the subdivision T becomes finer and finer, in the sense that $\delta(T)$ becomes smaller and smaller (which means that n gets bigger and bigger), do the sums $s_T(f)$ have a limit?

The answer is *yes*.

The reason is that a continuous function on a compact interval $[a, b]$ is uniformly continuous, and this implies that for any sequence (T_m) of subdivisions such that $\delta(T_m) \to 0$ as m goes to infinity, the sums $s_{T_m}(f)$ form a Cauchy sequence, as we now explain.

Proposition 3.1. *Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function defined on a closed and bounded (compact) interval $[a, b]$. For every $\epsilon > 0$, there is some $\eta > 0$ such that for any two subdivisions T and T' of $[a, b]$ such that $\delta(T) < \eta$ and $\delta(T') < \eta$, we have*

$$|s_T(f) - s_{T'}(f)| < \epsilon.$$

Proof. Since a continuous function on $[a, b]$ is actually uniformly continuous, for any $\epsilon > 0$, we can find some $\eta > 0$ such that

$$|f(x) - f(x')| < \epsilon/2(b-a) \quad \text{for all } x, x' \in [a, b] \text{ such that } |x - x'| < \eta.$$

If $T = \{t_0, t_1, \ldots, t_n\}$ and $T' = \{t'_0, t'_1, \ldots, t'_{n'}\}$, let $T'' = T \cup T'$ and let T''_k be the subdivision $T''_k = T' \cap [t_k, t_{k+1}]$, more precisely, $T''_k = \{s_0, s_1, \ldots, s_r\}$, with $s_0 = t_k$, $s_r = t_{k+1}$, and

$$\{s_1, \ldots, s_{r-1}\} = \{t'_j \mid t_k < t'_j < t_{k+1}\},$$

with $r = 0$ if the above set on the right-hand side is empty, for $k = 0, \ldots, n-1$.

Then we immediately check that

$$T'' = \bigcup_{k=0}^{n-1} T''_k, \quad \text{and} \quad s_{T''}(f) = \sum_{k=0}^{n-1} s_{T''_k}(f).$$

See Figure 3.2.

Since $s_{T''_k}(f)$ is of the form

$$s_{T''_k}(f) = \sum_{i=0}^{r-1} (s_{i+1} - s_i) f(s_i),$$

where $t_k \leq s_i \leq t_{k+1}$ for $i = 0, \ldots, r$, and since $\sum_{i=0}^{r-1}(s_{i+1} - s_i) = s_r - s_0 = t_{k+1} - t_k$, we have

$$\begin{aligned}
|s_{T''_k}(f) - (t_{k+1} - t_k) f(t_k)| &= \left| \sum_{i=0}^{r-1} (s_{i+1} - s_i) f(s_i) - (t_{k+1} - t_k) f(t_k) \right| \\
&= \left| \sum_{i=0}^{r-1} (s_{i+1} - s_i)(f(s_i) - f(t_k)) \right| \\
&\leq \sum_{i=0}^{r-1} |(s_{i+1} - s_i)| |(f(s_i) - f(t_k))| \\
&< \sum_{i=0}^{r-1} |(s_{i+1} - s_i)| \frac{\epsilon}{2(b-a)} \\
&= (t_{k+1} - t_k) \frac{\epsilon}{2(b-a)}.
\end{aligned}$$

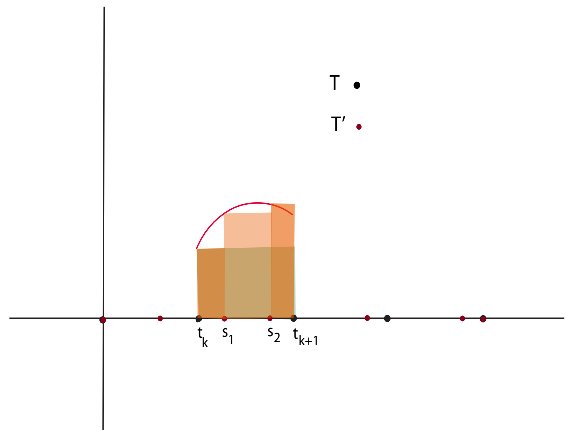


Figure 3.2: An illustration of the refinement $s_{T_k''}(f)$ utilized in the proof of Proposition 3.1. Note that T is given by the black dots while T' is given by the brown dots.

As a consequence, we obtain

$$\begin{aligned}
|s_T(f) - s_{T''}(f)| &= \left| \sum_{k=0}^{n-1} (t_{k+1} - t_k) f(t_k) - \sum_{k=0}^{n-1} s_{T_k''}(f) \right| \\
&\le \sum_{k=0}^{n-1} |s_{T_k''}(f) - (t_{k+1} - t_k) f(t_k)| \\
&< \sum_{k=0}^{n-1} (t_{k+1} - t_k) \frac{\epsilon}{2(b-a)} \\
&\le \frac{\epsilon}{2},
\end{aligned}$$

that is,

$$|s_T(f) - s_{T''}(f)| < \frac{\epsilon}{2}.$$

By a similar argument applied to T', we obtain

$$|s_{T'}(f) - s_{T''}(f)| < \frac{\epsilon}{2}.$$

But then we obtain

$$\begin{aligned}
|s_T(f) - s_{T'}(f)| &= |s_T(f) - s_{T''}(f) + s_{T''}(f) - s_{T'}(f)| \\
&\le |s_T(f) - s_{T''}(f)| + |s_{T''}(f) - s_{T'}(f)| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,
\end{aligned}$$

as claimed. $\square$

Remark: It is easy to check that the proof of Proposition 3.1 is still valid if we use more general Cauchy-Riemann sums. Namely, given a subdivision $T = \{t_0, t_1, \ldots, t_n\}$ of $[a, b]$, and any choice of reals $\theta_1, \ldots, \theta_n$ such that $t_k \leq \theta_{k+1} \leq t_{k+1}$ for $k = 0, \ldots, n-1$, define $s_T(f)$ as

$$s_T(f) = \sum_{k=0}^{n-1} (t_{k+1} - t_k) f(\theta_{k+1});$$

see Figure 3.3.

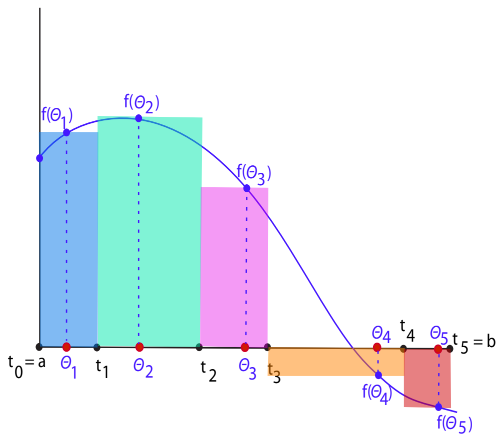


Figure 3.3: The general Cauchy–Riemann sum $s_T(f) = \sum_{k=0}^{4}(t_{k+1} - t_k)f(\theta_{k+1})$ is the "signed" area represented by the pastel shaded boxes.

Proposition 3.1 implies the following result, which establishes the existence of the Riemann integral of a continuous function defined on a closed and bounded (compact) interval $[a, b]$.

Theorem 3.2. *Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function defined on a closed and bounded (compact) interval $[a, b]$. For every sequence $\mathcal{T} = (T_m)_{m\geq 1}$ of subdivisions of $[a, b]$ such that $\lim_{m\mapsto\infty} \delta(T_m) = 0$, the sequence $(s_{T_m}(f))_{m\geq 1}$ is a Cauchy sequence, and thus has a limit $S_{\mathcal{T}}(f)$. For any two sequences $\mathcal{T} = (T_m)_{m\geq 1}$ and $\mathcal{T}' = (T'_m)_{m\geq 1}$ of subdivisions of $[a, b]$, if $\lim_{m\mapsto\infty} \delta(T_m) = 0$ and $\lim_{m\mapsto\infty} \delta(T'_m) = 0$, then $S_{\mathcal{T}}(f) = S_{\mathcal{T}'}(f)$, that is, the limit of the sequence $(s_{T_m}(f))$ is independent of the sequence $\mathcal{T} = (T_m)_{m\geq 1}$ such that $\lim_{m\mapsto\infty} \delta(T_m) = 0$.*

Proof. Pick any $\epsilon > 0$, and let $\eta > 0$ be some number given by Proposition 3.1, such that for any two subdivisions T and T' of $[a, b]$ such that $\delta(T) < \eta$ and $\delta(T') < \eta$, we have

$$|s_T(f) - s_{T'}(f)| < \epsilon.$$

Since $\lim_{m\mapsto\infty} \delta(T_m) = 0$, there is some $N > 0$ such that for all $m, n \geq N$, we have $\delta(T_m) < \eta$ and $\delta(T_n) < \eta$, which by the definition of η, implies that

$$|s_{T_m}(f) - s_{T_n}(f)| < \epsilon \quad \text{for all } m, n \geq N.$$

Therefore, $(s_{T_m}(f))$ is a Cauchy sequence. Since $\mathbb{R}$ is a complete metric space, this sequence has a limit $S_{\mathcal{T}}(f)$. The same argument shows that $(s_{T'_m}(f))$ is a Cauchy sequence which has a limit $S_{\mathcal{T}'}(f)$.

Since by hypothesis $\lim_{m\mapsto\infty}\delta(T_m) = 0$ and $\lim_{m\mapsto\infty}\delta(T'_m) = 0$, there is some $N > 0$ such that for all $m \geq N$, we have $\delta(T_m) < \eta$ and $\delta(T'_m) < \eta$, so by Proposition 3.1,

$$|s_{T_m}(f) - s_{T'_m}(f)| < \epsilon \quad \text{for all } m \geq N. \tag{eq1}$$

By the triangle inequality

$$|S_{\mathcal{T}'}(f) - S_{\mathcal{T}}(f)| \leq |S_{\mathcal{T}'}(f) - s_{T'_m}(f)| + |s_{T'_m}(f) - s_{T_m}(f)| + |s_{T_m}(f) - S_{\mathcal{T}}(f)|,$$

since the Cauchy sequences $(s_{T_m}(f))$ and $(s_{T'_m}(f))$ converge and (eq1) holds, we deduce that $S_{\mathcal{T}'}(f) = S_{\mathcal{T}}(f)$, that is, the sequences $(s_{T_m}(f))$ and $(s_{T'_m}(f))$ have the same limit. $\square$

Theorem 3.2 also holds for the more general Cauchy-Riemann sums defined in the Remark after Proposition 3.1.

Theorem 3.2 justifies the following definition.

Definition 3.2. Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function defined on a closed and bounded (compact) interval $[a,b]$. The common limit $S_{\mathcal{T}}(f)$ of the Cauchy sequences $(s_{T_m}(f))_{m\geq 1}$, for all sequences $\mathcal{T} = (T_m)_{m\geq 1}$ of subdivisions of $[a,b]$ such that $\lim_{m\mapsto\infty}\delta(T_m) = 0$, is called the *Riemann integral* of f, and is denoted by $\int_a^b f(t)dt$.

The following are basic properties of the Riemann integral, which are easy to prove (using suitable subdivisions of $[a,b]$):

1. The mapping $f \mapsto \int_a^b f(t)dt$ is a *linear form* on the space of continuous functions on $[a,b]$. This means that for any two continuous functions $f,g\colon [a,b] \to \mathbb{R}$ and any scalar $\lambda \in \mathbb{R}$,

$$\int_a^b (f+g)(t)dt = \int_a^b f(t)dt + \int_a^b g(t)dt$$
$$\int_a^b (\lambda f)(t)dt = \lambda \int_a^b f(t)dt,$$

where, as usual, $f+g$ is the function given by $(f+g)(t) = f(t) + g(t)$, and λf is the function given by $(\lambda f)(t) = \lambda f(t)$, for all $t \in [a,b]$. Furthermore, it is a positive linear form, which means that if $f \geq 0$, then $\int_a^b f(t)dt \geq 0$. These seemingly innocuous properties turn out to be very important. Indeed, we will see later how the notion of integral on a locally compact space can be defined in terms of such linear forms (Radon functionals).

2.
$$\left|\int_a^b f(t)dt\right| \leq \int_a^b |f(t)|dt \leq (b-a)\max_{t\in[a,b]} |f(t)|;$$
see Figure 3.4.

3. If $f \geq 0$ and $f(t) > 0$ for some $t \in [a,b]$, then $\int_a^b f(t)dt > 0$.

4. If $a < b < c$, then
$$\int_a^b f(t)dt + \int_b^c f(t)dt = \int_a^c f(t)dt;$$
see Figure 3.5.

5. If $H\colon [a,b] \to \mathbb{R}$ is the function given by
$$H(x) = \int_a^x f(t)dt,$$
then H is differentiable on $[a,b]$ and $H'(x) = f(x)$ (the so-called *first fundamental theorem of calculus*).

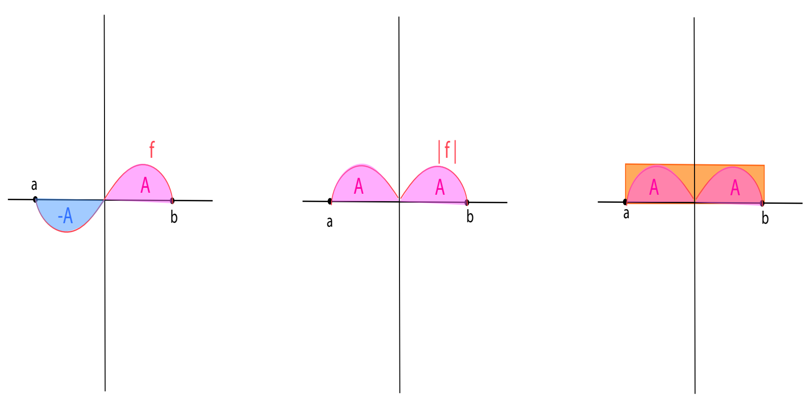


Figure 3.4: The left figure illustrates $\int_a^b f(t)dt = A+(-A) = 0$, while the middle figure illustrates $\int_a^b |f(t)|dt = 2A$, so $\left|\int_a^b f(t)dt\right| \leq \int_a^b |f(t)|dt$. The right figure shows that $\int_a^b |f(t)|dt$ is contained within the orange rectangle of area $(b-a)\max_{t\in[a,b]} |f(t)|$.

The process that we just described only requires that the codomain be complete and that a continuous function $f\colon [a,b] \to F$ be uniformly continuous. We also need the linear combinations $\sum_{k=0}^{n-1}(t_{k+1} - t_k)f(t_k)$ to make sense, so F should be a vector space. If we assume that F is a complete normed vector space (a Banach space), then the Riemann integral of a continuous vector-valued function $f\colon [a,b] \to F$ can be defined by using the method that we just presented.

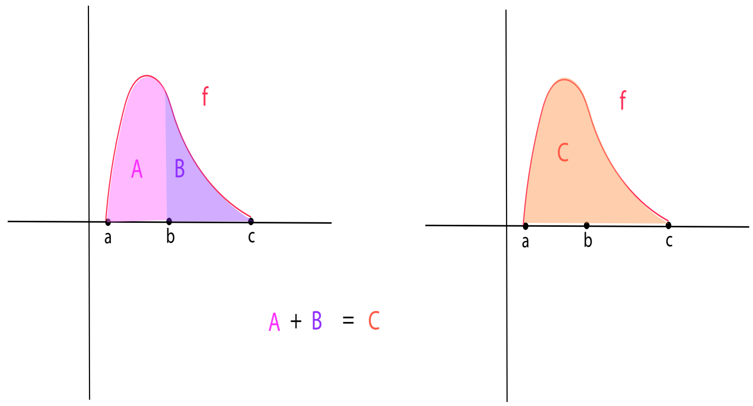


Figure 3.5: In the left figure, $\int_a^b f(t)dt = A$, while $\int_b^c f(t)dt = B$, and this is equal to the peach area under the curve from a to c.

In the next section we show how to define the integral of functions with discontinuities, provided that these discontinuities are "reasonable." For this, a new crucial idea is needed: to define the integral on a class of simple functions with a finite number of reasonable discontinuities, and then to extend the integral to a bigger class of functions by taking limits of simple functions. For this process to work, the bigger space of functions should be complete.

3.2 The Riemann Integral of Regulated Functions

In this section we show how to define the integral of regulated functions $f: [a, b] \to F$, where F is any complete normed vector space, in particular $\mathbb{R}$ or any finite-dimensional vector space (real or complex).

The first key ingredient is that the method of Cauchy-Riemann sums can be immediately adapted to define the notion of integral for a step function. The mapping $\int : \text{Step}([a, b]; F) \to F$ is seen to be linear and continuous.

The second key ingredient is that, by Proposition 2.29, the vector space $\text{Step}([a, b]; F)$ of step functions over $[a, b]$ is dense in $\text{Reg}([a, b]; F)$, the space of regulated functions over $[a, b]$, and $\text{Reg}([a, b]; F)$ is complete, where $[a, b]$ is a closed and bounded interval.

Then, because $\text{Step}([a, b]; F)$ is dense in $\text{Reg}([a, b]; F)$, and $\text{Reg}([a, b]; F)$ is complete, by Theorem A.73, the continuous linear map $\int : \text{Step}([a, b]; F) \to F$ has a unique extension $\int : \text{Reg}([a, b]; F) \to F$ to $\text{Reg}([a, b]; F)$, which is also continuous and linear. This is how the integral of a regulated function is defined.

Thus it remains to define the integral of a step function.

Definition 3.3. Let $f\colon [a,b] \to F$ be a step function. For any admissible subdivision $T = (a_0, a_1, \ldots, a_n)$ for f, for any sequence $\xi = (\xi_1, \ldots, \xi_n)$ of reals such that $\xi_{k+1} \in (a_k, a_{k+1})$ for $k = 0, \ldots, n-1$, define $s_{T,\xi}(f)$ by

$$s_{T,\xi}(f) = \sum_{k=0}^{n-1} (a_{k+1} - a_k) f(\xi_{k+1}).$$

See Figure 3.6.

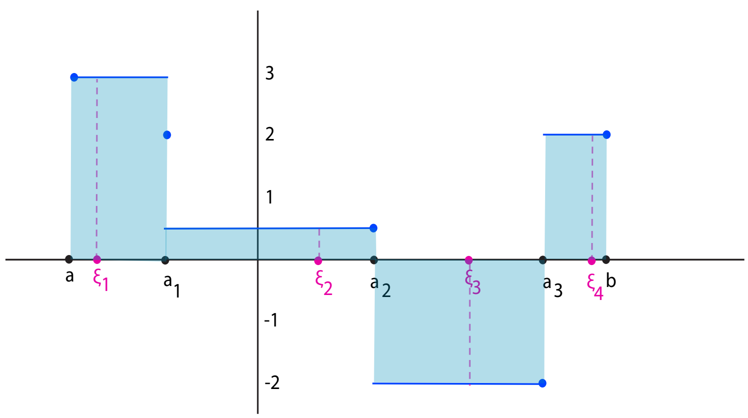


Figure 3.6: An illustration of $s_{T,\xi}(f) = \sum_{k=0}^{3} (a_{k+1} - a_k) f(\xi_{k+1})$ for step function $f\colon [a,b] \to \mathbb{R}$.

The above is a linear combination of vectors in F, and since F is a vector space, it is well defined. Note that because $\xi_{k+1} \in (a_k, a_{k+1})$, $s_{T,\xi}(f)$ does not depend on the value of f at the a_k. For simplicity of language, we refer to a pair (T, ξ) as in Definition 3.3 as an *admissible pair* for f.

The problem with the above definition of $s_{T,\xi}(f)$ is that it depends on the admissible subdivision T, and on ξ, but because f is a step function, it is constant on each interval (a_k, a_{k+1}), so in fact $s_{T,\xi}(f)$ is independent of the admissible pair (T, ξ).

Proposition 3.3. *Given a step function $f\colon [a,b] \to F$, for any two admissible pairs (T, ξ) and (T', ξ') for f, we have $s_{T,\xi}(f) = s_{T',\xi'}(f)$.*

Proposition 3.3 is proved by using an admissible pair which is finer than both (T, ξ) and (T', ξ'). The details are left to the reader, or see Schwartz [86] (Chapter V, Section §1).

Proposition 3.3 justifies the following definition.

Definition 3.4. Let $f\colon [a,b] \to F$ be a step function. The *integral* of f, denoted $\int_{[a,b]} f$, is the common value of the sum $s_{T,\xi}(f)$, for any any admissible pair (T, ξ) for f.

The following proposition follows almost immediately from the definitions.

Proposition 3.4. *The map* $\int\colon \mathrm{Step}([a,b];F) \to F$, *where* $\int f = \int_{[a,b]} f$ *is the integral defined in Definition 3.4, is linear. Furthermore, we have*

$$\left\| \int_{[a,b]} f \right\| \leq \int_{[a,b]} \|f\| \quad \textit{and} \quad \left\| \int_{[a,b]} f \right\| \leq (b-a)\, \|f\|_{\infty},$$

where $\|f\|$ *means the real-valued function* $x \mapsto \|f(x)\|$. *If* $f = \mathbb{R}$ *and if* $f \geq 0$, *then* $\int_{[a,b]} f \geq 0$.

Proposition 3.4 shows that the map $\int\colon \mathrm{Step}([a,b];F) \to F$ is linear and continuous. As we explained earlier, by Theorem A.73, the map $\int\colon \mathrm{Step}([a,b];F) \to F$ has a unique extension $\int\colon \mathrm{Reg}([a,b];F) \to F$ to $\mathrm{Reg}([a,b];F)$, which is also linear and continuous.

Definition 3.5. The *integral* $\int_{[a.b]} f$ of any regulated function $f \in \mathrm{Reg}([a,b];F)$ is equal to $\int f$, where $\int\colon \mathrm{Reg}([a,b];F) \to F$ is the unique linear and continuous extension of the linear and continuous map $\int\colon \mathrm{Step}([a,b];F) \to F$. This integral is called the *Riemann integral* of the regulated function f.

Definition 3.5 is not very constructive. It turns out that the the Riemann integral of a regulated function can be defined more directly in terms of generalized Riemann sums. This approach is presented in Schwartz [86] (Chapter V, Section §1).

Note that we actually haven't defined the notion of Riemann-integrable function. What we did is to exhibit a family of functions, the regulated functions, which are Riemann-integrable function. The notion of Riemann-integrable function is defined in various books, including Schwartz [86]. This can be done using the notion of upper integral $\int^* f$, which is defined for a positive function $f \in \mathcal{K}(\mathbb{R}, F)$ as the infimum of the integrals of the step functions that bound f from above.

The space of Riemann-integrable functions contains other functions besides the regulated functions. For example, functions with compact support which are continuous except at finitely many points, are Riemann-integrable. The function $x \mapsto \sin(1/x)$ is such a function (with value 0 at $x = 0$). It is Riemann-integrable on $[0, 1]$, even though 0 is not a discontinuity of the first kind.

The method of this section, which consists in defining the notion of integral for a "big" set of functions, such as $\mathrm{Reg}([a,b];F)$, by first defining a notion of integral on a very simple set of functions for which the definition is obvious, such as $\mathrm{Step}([a,b];F)$, and then to extend the integral on $\mathrm{Step}([a,b];F)$ to a notion of integral on $\mathrm{Reg}([a,b];F)$ using a completion process, is a key idea. In this situation we are lucky that $\mathrm{Reg}([a,b];F)$ is complete.

In order to define a notion of integral for functions defined on a domain X which is more general than a compact interval $[a,b]$ of $\mathbb{R}$, we can proceed as above, but some additional

structure on X is needed to define step functions and the notion of integral of step functions. This new ingredient is the notion of *measure*. The other technical difficulty is that the completion of the space of generalized step functions is not a space identifiable with a space of familiar functions. By Theorem A.72, the completion always exists, but its elements are equivalence classes of functions, so it will take some work to exhibit this space as a set of functions.

Chapter 4

Measure Theory; Basic Notions

Let X be a nonempty set. Intuitively, a measure on X is a function μ that assigns a nonnegative real number $\mu(A)$ to every subset A in some specified nonempty collection $\mathcal{A}$ of subsets of X, where $\mu(A)$ is a generalization of the notion of length, area, or volume. For example, any natural measure μ on $\mathbb{R}$ should have the property that $\mu((a,b)) = \mu([a,b]) = b-a$ for all $a \leq b$. It is natural to require that if a subset A is sliced into countably many pairwise disjoint small pieces A_i, then $\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} A_i\right) = \sum_{n=1}^{\infty} \mu(A_i)$. This property is called *σ-additivity*. Then the family $\mathcal{A}$ of subsets on which μ is defined should be closed under countable unions. It is also natural to require $\mathcal{A}$ to be closed under complementation. This leads to the important notion of a *σ-algebra*, which is closed under complementation and countable unions. The weaker notion which only requires closure under complementation and closure under finite unions is that of an *algebra*. In general it is not easy to construct nontrivial σ-algebras, so it is useful to have tools to do so. A pair $(X, \mathcal{A})$ consisting of a nonempty set and a σ-algebra $\mathcal{A}$ is called a *measurable space*.

Given any nonempty family $\mathcal{S}$ of subsets of X, there is a smallest σ-algebra $\mathcal{A}(\mathcal{S})$ containing $\mathcal{S}$. If X is a topological space, then the σ-algebra $\mathcal{B}(X)$ containing the open subsets of X is an important σ-algebra called the *Borel σ-algebra*.

The notion of *monotone class* is also useful to construct σ-algebras. Given any nonempty family $\mathcal{S}$ of subsets of X, there is a smallest monotone class $\mathfrak{M}(\mathcal{S})$ containing $\mathcal{S}$. Given an algebra $\mathcal{B}$, the smallest σ-algebra $\mathcal{A}(\mathcal{B})$ containing $\mathcal{B}$ and the smallest monotone class $\mathfrak{M}(\mathcal{B})$ containing $\mathcal{B}$ are identical: $\mathcal{A}(\mathcal{B}) = \mathfrak{M}(\mathcal{B})$.

Next we define (positive) measures on a σ-algebra. A triple $(X, \mathcal{A}, \mu)$ consisting of a nonempty set, a σ-algebra $\mathcal{A}$, and a measure μ on $\mathcal{A}$ is called a *measure space*. We investigate a few properties of measures. In particular, we show that every measure can be extended to a *complete measure*, which means that all $A \in \mathcal{A}$, if $\mu(A) = 0$, then $B \in \mathcal{A}$ for all $B \subseteq A$.

As we said earlier, it is not easy to construct nontrivial measures. A very useful concept to achieve this is the notion of *outer measure*, introduced in Section 4.4. Outer measures are defined for all subsets of X, which makes them much easier to construct. In particular, we construct the *Lebesgue outer measure* μ_L^*.

A fundamental theorem due to Carathéodory shows that every outer measure induces a measure space; see Theorem 4.11.

By applying Theorem 4.11 to the outer measure μ_L^* we obtain the σ-algebra $\mathcal{L}(\mathbb{R})$ of *Lebesgue measurable sets* and the *Lebesgue measure* μ_L; see Section 4.5. The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is properly contained in the σ-algebra $\mathcal{L}(\mathbb{R})$ of Lebesgue measurable sets, and there are subsets of $\mathbb{R}$ that are not Lebesgue measurable sets (assuming the axiom of choice). We also discuss various regularity properties of the Lebesgue measure.

4.1 σ-Algebras

Let X be a nonempty set. We would like to define the notion of "measure" for the subsets of X in such a way that familiar properties of the notion of length, area, or volume of polyhedral objects in $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$ hold. The measure $m(A)$ of a subset of X should be nonnegative, but we have to allow "big" objects to have infinite measure so it is desirable to extend the nonnnegative real numbers by adding a new element corresponding to infinity.

Technically, we define $\overline{R}_+$ as the union

$$\overline{R}_+ = \{\alpha \in \mathbb{R} \mid \alpha \geq 0\} \cup \{+\infty\} = \mathbb{R}_+ \cup \{+\infty\},$$

where $+\infty$ is *not* in $\mathbb{R}_+$, and we assume that the following properties hold:

(a) $\alpha < +\infty$, for all $\alpha \in \mathbb{R}_+$,

(b) $\alpha + (+\infty) = (+\infty) + \alpha = +\infty$, for all $\alpha \in \overline{\mathbb{R}}_+$,

(c) $\alpha \cdot (+\infty) = (+\infty) \cdot \alpha = +\infty$, for all $\alpha \in \overline{\mathbb{R}}_+ - \{0\}$,

(d) $0 \cdot (+\infty) = (+\infty) \cdot 0 = 0$,

(e) If $(\alpha_i)_{i\geq 1}$ is a sequence with $\alpha_i \in \overline{\mathbb{R}}_+$, and if $\alpha_i = +\infty$ for some i, then $\sum_{i=1}^{\infty} \alpha_i = +\infty$.

The set $\overline{R}_+$ is also denoted by $[0, +\infty]$.

In this chapter we closely follow Halmos [44] and some course notes given by Philippe G. Ciarlet in 1970-1971 at ENPC (Paris, France). Other nice (but concise) presentations can be found in Rudin [79], Folland [34], and Lang [62]. A very detailed presentation is given in Schwartz [86].

An "ideal measure" should be a function m satisfying the following properties:

(1) $m\colon 2^X \to [0, +\infty]$, that is, m is defined for *all* subsets of X.

(2) $m(\emptyset) = 0$.

(3) For any countable sequence $(A_i)_{i \geq 1}$ of subsets A_i of X such that $A_i \cap A_j = \emptyset$ for all $i \neq j$,
$$m\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(A_i).$$
This property is called *σ-additivity*.

The intuition behind σ-additivity is that if we slice an object A into countably many pairwise disjoint small pieces A_i, then the measure $m(A)$ of A should be the sum of the measures $m(A_i)$ of the pieces A_i.

Observe that by choosing a sequence $(A_i)_{i \geq 1}$ such that $A_j = \emptyset$ for all $j > n$, and $A_i \cap A_j = \emptyset$ if $i \neq j$, we obtain the property
$$m\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} m(A_i),$$
known as *additivity*; see Figure 4.1.

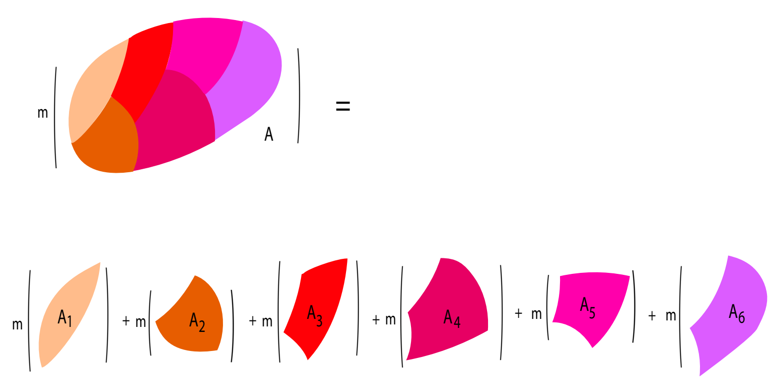


Figure 4.1: A pictorial representation of the identity $m\left(\bigcup_{i=1}^{6} A_i\right) = \sum_{i=1}^{6} m(A_i)$.

For any two subsets A and B, if $A \subseteq B$, we can write $B = A \cup (B - A)$, with $A \cap (B - A) = \emptyset$, so by additivity,
$$m(B) = m(A) + m(B - A),$$
and since $m(B - A) \geq 0$, we obtain
$$m(A) \leq m(B);$$
see Figure 4.2.

We claim that the following property holds.

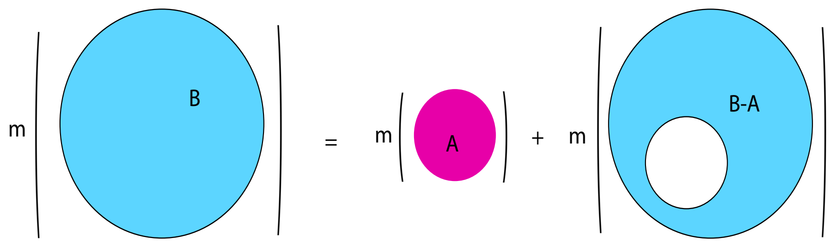


Figure 4.2: A pictorial representation of the identities $m(B) = m(A) + m(B - A)$ and $m(A) \leq m(B)$.

Proposition 4.1. *If a function m satisfies Properties (1–3) above, then for any countable sequence $(A_i)_{i\geq 1}$ of (not necessarily pairwise disjoint) subsets A_i of X,*

$$m\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} m(A_i).$$

Proof. Define the sequence (B_i) of subsets of X as follows: $B_1 = A_1$, $B_2 = A_2 - A_1, \ldots$, $B_i = A_i - \left(\bigcup_{j=1}^{i-1} A_j\right)$, for all $i \geq 2$. See Figure 4.3.

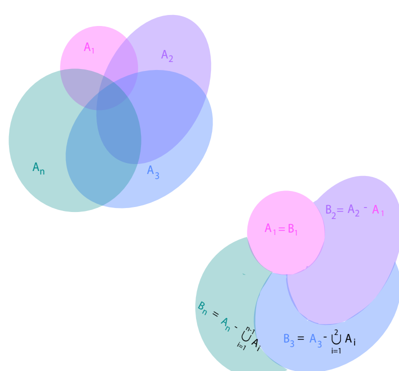


Figure 4.3: A schematic illustration of the set construction (B_i).

It is easy to check that $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$, $B_i \cap B_j = \emptyset$ for all $i \neq j$, and $m(B_i) \leq m(A_i)$ for all $i \geq 1$, so by σ-additivity,

$$m\left(\bigcup_{i=1}^{\infty} A_i\right) = m\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} m(B_i) \leq \sum_{i=1}^{\infty} m(A_i),$$

as claimed. $\square$

In general, for an arbitrary set X, there may be no function m satisfying Properties (1–3) for *all* subsets of X, as well as certain desirable properties. For example, there is no such translation invariant function on $2^{\mathbb{R}}$ such that $m([0,1)) \neq 0$ and $m([0,1)) \neq +\infty$, and no such translation invariant function on $2^{\mathbb{R}}$ such that $m([a,b]) = b - a$ for every interval $[a,b]$; see Section 4.5.

Thus we are led to relax some of these conditions. There are two options:

(1) The first option is to relax (3) by replacing it by the result of Proposition 4.1, namely

(3') For any countable sequence $(A_i)_{i\geq 1}$ of subsets A_i of X,

$$m\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} m(A_i).$$

This approach leads to *outer measures*, and is discussed in Section 4.4.

(2) Condition (3) is highly desirable, so the second option is to restrict the domain of m to be a proper family of subsets of X; the right notion is that of a *σ-algebra*.

The notion of a σ-algebra is more important that the notion of outer measure, which is needed for technical reasons. Thus we now consider Option 2, and define σ-algebras. Once the notion of σ-algebra is defined, we will be able to define the abstract notion of a measure (see Definition 4.9), which is the crucial ingredient in defining a general notion of integral.

Definition 4.1. Let X be any nonempty set. A family $\mathcal{A}$ of subsets of X is a *σ-algebra* if it satisfies the following conditions:

(A1) $X \in \mathcal{A}$.

(A2) For every subset A of X, if $A \in \mathcal{A}$, then $X - A \in \mathcal{A}$ (closure under complementation).

(σ-A3) For every countable family $(A_i)_{i\geq 1}$ of subsets of X, if $A_i \in \mathcal{A}$ for all $i \geq 1$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ (closure under countable unions).

From (A1) and (A2), we see that $\emptyset \in \mathcal{A}$. From (A2) and (σ-A3) and the fact that $A = X - (X - A)$ and $\bigcap_{i=1}^{\infty} A_i = X - (\bigcup_{i=1}^{\infty}(X - A_i))$, if $A_i \in \mathcal{A}$ for all $i \geq 1$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$ (closure under countable intersections). In particular, if we let $A_i = \emptyset$ for all $i \geq 3$, we see that if $A_1 \in \mathcal{A}$ and $A_2 \in \mathcal{A}$, then $A_1 \cup A_2 \in \mathcal{A}$ and $A_1 \cap A_2 \in \mathcal{A}$. Since $A_1 - A_2 = A_1 \cap (X - A_2)$, we also have $A_1 - A_2 \in \mathcal{A}$.

Axiom (σ-A3) is a strong condition, and this is the reason why it is not easy to construct nontrivial σ-algebras. There are two extreme σ-algebras:

1. $\mathcal{A} = \{\emptyset, X\}$.
2. $\mathcal{A} = 2^X$, the family of all subsets of X.

Interesting σ-algebra lie in-between.

Remarks:

1. Some authors use the term *σ-field* instead of σ-algebra. This is a rather unfortunate terminology, because in algebra, a field is a set with two operations that have identity elements. Here the operations are union and intersection, but there is no identity element for intersection.

2. If we weaken Condition σ-A3 to *finite* unions, then we obtain a structure called an *algebra* (or *boolean algebra* of sets).

Definition 4.2. Let X be any nonempty set. A family $\mathcal{B}$ of subsets of X is an *algebra* (or *boolean algebra* of sets) if it satisfies the following conditions:

(A1) $X \in \mathcal{B}$.

(A2) For every subset A of X, if $A \in \mathcal{B}$, then $X - A \in \mathcal{B}$ (closure under complementation).

(A3) For every finite family $(A_i)_{i=1}^n$ of subsets of X, if $A_i \in \mathcal{B}$ for all $i = 1, \ldots, n$, then $\bigcup_{i=1}^n A_i \in \mathcal{B}$ (closure under finite unions).

As in the case of σ-algebras, an algebra contains $\emptyset$ and is closed under (finite) unions and intersections. In the construction of a product of measurable spaces, another notion of algebra will come up. These are the semi-algebras.

Definition 4.3. Let X be any nonempty set. A family $\mathcal{S}$ of subsets of X is a *semi-algebra* if it satisfies the following conditions:

(S1) $X, \emptyset \in \mathcal{S}$.

(S2) For all $A, B \in \mathcal{S}$, we have $A \cap B \in \mathcal{S}$.

(S3) For all $A \in \mathcal{S}$, we have $X - A = X_1 \cup \cdots \cup X_n$, for finitely many pairwise disjoint subsets $X_i \in \mathcal{S}$.

Example 4.1. First consider the family of intervals of $\mathbb{R}$ of the form $[a, b)$, with $a \leq b$, where $a = -\infty$ or $b = \infty$ is allowed. By convention, let $[a, b) = \emptyset$ if $a > b$. This is a semi-algebra, because

$$[a_1, b_1) \cap [a_2, b_2) = [\max(a_1, a_2), \min(b_1, b_2)),$$

and

$$X - [a, b) = [-\infty, a) \cup [b, \infty);$$

see Figure 4.4.

Figure 4.4: The left figure illustrates $[a_1, b_1) \cap [a_2, b_2) = [\max(a_1, a_2), \min(b_1, b_2)) = [a_2, b_1)$, while the right figure illustrates $\mathbb{R} - [a, b) = [-\infty, a) \cup [b, \infty)$.

Example 4.2. Next, let X and Y be two nonempty sets, and let $\mathcal{A}$ be an algebra on X and let $\mathcal{B}$ be an algebra on Y. Define the set $\mathcal{R}$ of *rectangles* in $X \times Y$ as follows:

$$\mathcal{R} = \{A \times B \in X \times Y \mid A \in \mathcal{A},\ B \in \mathcal{B}\}.$$

It is easy to check that $\mathcal{R}$ is a semi-algebra. For example,

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2),$$

and

$$(X \times Y) - (A \times B) = ((X - A) \times (Y - B)) \cup ((X - A) \times B) \cup (A \times (Y - B));$$

see Figure 4.5.

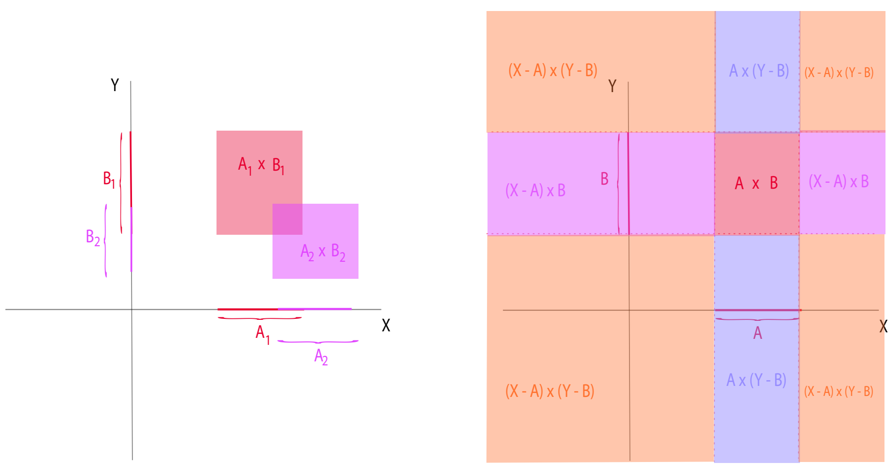


Figure 4.5: Let X and Y be arbitrary topological spaces (for example $\mathbb{R}$). The left figure illustrates $(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$ as the overlap of the red and lilac rectangles while the right figure illustrates $(X \times Y) - (A \times B) = ((X - A) \times (Y - B)) \cup ((X - A) \times B) \cup (A \times (Y - B))$.

Then it can be shown that the set $\mathcal{B}(\mathcal{R})$ of finite unions of pairwise disjoint sets in $\mathcal{R}$ is the smallest algebra containing the semi-algebra $\mathcal{R}$. This algebra will be used to construct the product of measurable spaces.

The following result can be shown.

Proposition 4.2. *Given a semi-algebra $\mathcal{S}$, the smallest algebra $\mathcal{B}(\mathcal{S})$ containing $\mathcal{S}$ is the family of finite unions of pairwise disjoint subsets in $\mathcal{S}$.*

Definition 4.4. Let X be any nonempty set. A pair $(X, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra of subsets of X is called a *measurable space*. The subsets of X that belong to $\mathcal{A}$ are called the *measurable subsets* of X.

Proposition 4.3. *Let X be any nonempty set, and let $\mathcal{S}$ be any nonempty family of subsets of X. Then there is a σ-algebra $\mathcal{A}(\mathcal{S})$ with the following properties:*

(a) $\mathcal{S} \subseteq \mathcal{A}(\mathcal{S})$.

(b) If $\mathcal{A}'$ is any σ-algebra such that $\mathcal{S} \subseteq \mathcal{A}'$, then $\mathcal{A}(\mathcal{S}) \subseteq \mathcal{A}'$.

This means that $\mathcal{A}(\mathcal{S})$ is the smallest σ-algebra containing $\mathcal{S}$.

Definition 4.5. Let X be any nonempty set, and let $\mathcal{S}$ be any nonempty family of subsets of X. The smallest σ-algebra $\mathcal{A}(\mathcal{S})$ containing $\mathcal{S}$ is called the *σ-algebra generated by $\mathcal{S}$*.

The σ-algebra $\mathcal{A}(\mathcal{S})$ is the intersection of the family of all σ-algebras containing $\mathcal{S}$. This family is nonempty since 2^X belongs to it. This way of defining $\mathcal{A}(\mathcal{S})$ is highly nonconstructive. A bottom-up construction of $\mathcal{A}(\mathcal{S})$ can be performed, but to guarantee closure under countable infinite unions, transfinite induction is required; see Schwartz [86] (Chapter V, Section §2) or Folland [34] (Proposition 1.23).

Remark: Readers not familiar with ordinals should skip this remark. For a quick review of the notion of ordinal and their basic properties, see Chapter E. Recall that an ordinal $\alpha > 0$ is either a successor ordinal, which means that $\alpha = \beta + 1$ for some ordinal $\beta < \alpha$, or a limit ordinal, which means that $\alpha = \bigcup_{\beta<\alpha} \beta$. Given $\mathcal{S}$ we define the sequence $\mathcal{S}_\alpha$ by transfinite induction. In fact, it suffices to construct this sequence for countable ordinals. We set

$$\begin{aligned}
\mathcal{S}_0 &= \mathcal{S} \\
\mathcal{S}_{\beta+1} &= \mathcal{S}_\beta \cup \left\{ \bigcup_{i=1}^{\infty} A_i \mid A_i \in \mathcal{S}_\beta \right\} \cup \{X - A \mid A \in \mathcal{S}_\beta\} \\
\mathcal{S}_\alpha &= \bigcup_{\beta<\alpha} \mathcal{S}_\beta
\end{aligned}$$

where α is a limit ordinal. If Ω is the set of all countable ordinals, then we let

$$\mathcal{S}^\dagger = \bigcup_{\alpha\in\Omega} \mathcal{S}_\alpha.$$

Because every increasing sequence in Ω has a supremum, it can shown that $\mathcal{A}(\mathcal{S}) = \mathcal{S}^\dagger$; see Folland [34] (Proposition 1.23). The cardinal of the set $\mathbb{R}$ of real numbers is denoted by $\mathfrak{c}$ or

$2^{\aleph_0}$. The proof also implies that if $\mathcal{S}$ is of cardinality $\aleph_0 \leq |\mathcal{S}| \leq \mathfrak{c}$, then $\mathcal{A}(\mathcal{S})$ has cardinality $\mathfrak{c}$.

An important example arises when X is a topological space $(X, \mathcal{O})$.

Definition 4.6. Let $(X, \mathcal{O})$ be a topological space. The σ-algebra $\mathcal{B}(X)$ generated by the family $\mathcal{O}$ of open sets is called the *Borel σ-algebra* of X. The subsets in $\mathcal{B}(X)$ are called *Borel sets*.

All open subsets and all closed sets are Borel sets. Countably infinite unions of closed sets and countable infinite intersections of open sets are Borel sets. But there are many more Borel sets.

Another way to construct σ-algebras is to use algebras and monotone classes. Although we are not going to use monotone classes in this book, there are a useful tool in constructing σ-algebras. They are used in the proof of Theorem 5.55 on the existence of measures on products of measure spaces. They are also useful in proving that certain functions defined on semi-algebras or algebras $\mathcal{B}$ having some of the properties of measures can be extended to measures on certain σ-algebras induced by $\mathcal{B}$.

Definition 4.7. Let X be any nonempty set. A nonempty family $\mathfrak{M}$ of subsets of X is a *monotone class* if for every countable family $(A_i)_{i \geq 1}$ of subsets of X, if $A_i \in \mathfrak{M}$ for all $i \geq 1$ then:

1. If $A_i \subseteq A_{i+1}$ for all $i \geq 1$, then $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{M}$. See Figure 4.6, Diagram (i).
2. If $A_{i+1} \subseteq A_i$ for all $i \geq 1$, then $\bigcap_{i=1}^{\infty} A_i \in \mathfrak{M}$. See Figure 4.6, Diagram (ii).

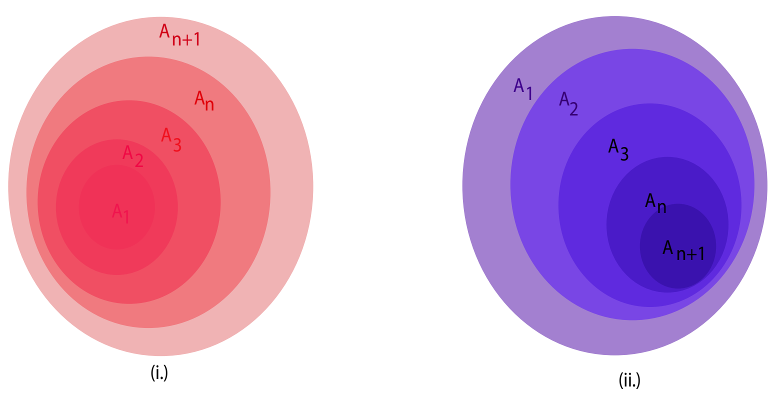


Figure 4.6: The rose colored sets of Figure (i) satisfy the increasing nesting condition of $A_i \subseteq A_{i+1}$, while the periwinkle sets of Figure (ii) satisfy the decreasing nesting condition $A_{i+1} \subseteq A_i$.

Proposition 4.4. *Let X be any nonempty set. For any algebra $\mathcal{B}$, if $\mathcal{B}$ is a monotone class, then $\mathcal{B}$ is a σ-algebra.*

Proof. Let $(A_i)_{i\geq 1}$ be a countable family of subsets of X, such that $A_i \in \mathcal{B}$ for all $i \geq 1$. Since $\mathcal{B}$ is an algebra, it is closed under finite unions, so $B_n = \bigcup_{i=1}^n A_i \in \mathcal{B}$, and obviously $B_n \subseteq B_{n+1}$ for all $n \geq 1$, and $\bigcup_{i=1}^\infty A_i = \bigcup_{n=1}^\infty B_n$. Since $\mathcal{B}$ is a monotone class, $\bigcup_{i=1}^\infty A_i = \bigcup_{n=1}^\infty B_n \in \mathcal{B}$. ∎

Here is a version of Proposition 4.3 for monotone classes.

Proposition 4.5. *Let X be any nonempty set, and let $\mathcal{S}$ be any nonempty family of subsets of X. Then there is a monotone class $\mathfrak{M}(\mathcal{S})$ with the following properties:*

(a) $\mathcal{S} \subseteq \mathfrak{M}(\mathcal{S})$.

(b) If $\mathfrak{M}'$ is any monotone class such that $\mathcal{S} \subseteq \mathfrak{M}'$, then $\mathfrak{M}(\mathcal{S}) \subseteq \mathfrak{M}'$.

This means that $\mathfrak{M}(\mathcal{S})$ is the smallest monotone class containing $\mathcal{S}$.

Definition 4.8. Let X be any nonempty set, and let $\mathcal{S}$ be any nonempty family of subsets of X. The smallest monotone class $\mathfrak{M}(\mathcal{S})$ containing $\mathcal{S}$ is called the *monotone class generated by $\mathcal{S}$*.

The following theorem yields another way of generating a σ-algebra from an algebra.

Theorem 4.6. *Let X be any nonempty set. For any algebra $\mathcal{B}$, the σ-algebra $\mathcal{A}(\mathcal{B})$ generated by $\mathcal{B}$ and the monotone class $\mathfrak{M}(\mathcal{B})$ generated by $\mathcal{B}$ are identical; that is,*

$$\mathcal{A}(\mathcal{B}) = \mathfrak{M}(\mathcal{B}).$$

Theorem 4.6 is proven in Folland [34] (Lemma 2.35).

We now come to the very important notion of measure.

4.2 Measures

Definition 4.9. Let X be any nonempty set. A *measure* on X is a map μ satisfying the following properties:

(μ1) $\mu\colon \mathcal{A} \to [0, +\infty]$, where $\mathcal{A}$ is a σ-algebra of subsets of X.

(μ2) $\mu(\emptyset) = 0$.

(μ3) For any countable sequence $(A_i)_{i\geq 1}$ of subsets A_i of $\mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$,

$$\mu\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mu(A_i).$$

This property is called *σ-additivity*.

A *measure space* is a triple $(X, \mathcal{A}, \mu)$, where $(X, \mathcal{A})$ is a measurable space and μ is a measure on X. A measure μ is also called a *positive measure*, to stress that its range is nonnegative.

Remarks:

1. The degenerate situation where $\mu(A) = +\infty$ for all nonempty subsets in $\mathcal{A}$ is allowed. If μ is *nontrivial*, which means that $\mathcal{A}$ possesses some nonempty subset A such that $\mu(A)$ is finite, then by letting $A_1 = A$ and $A_i = \emptyset$ for all $i \geq 2$, by ($\mu 3$) we get $\mu(A) = \mu(A) + \sum_{i=2}^{\infty} \mu(\emptyset)$, which implies $\mu(\emptyset) = 0$. In this situation Axiom ($\mu 2$) is unnecessary. Rudin makes the assumption that a measure is nontrivial; see [79].

2. Axiom ($\mu 3$) raises a subtle point. If $(A_i)_{i\geq 1}$ is a countable family of pairwise disjoint subsets $A_i \in \mathcal{A}$, the subset $A = \bigcup_{i=1}^{\infty} A_i$ does not depend on the order of the A_i, so for any permutation σ of the positive integers we should have
$$\mu(A) = \sum_{i=1}^{\infty} \mu(A_{\sigma(i)}) = \sum_{i=1}^{\infty} \mu(A_i).$$
How do we know that this is the case?

   But the numbers $\mu(A_i)$ are nonnegative, so the series $\sum_{i=1}^{\infty} \mu(A_i)$ converges absolutely, and thus *commutatively*. For example, see Schwartz [84] (Chapter II, Theorem 2.12.7 and Theorem 2.12.12, which says that in a normed vector space of finite dimension, a series is commutatively convergent iff its is absolutely convergent). Thus there is actually no problem with Axiom ($\mu 3$).

   There are more general measures taking their values in $\mathbb{R}$ or $\mathbb{C}$, or even in a Banach space. For such measures, Condition ($\mu 3$) needs to be slightly strengthened.

3. Some authors use the term *measured space* instead of *measure space*.

Definition 4.10. Let $(X, \mathcal{A}, \mu)$ be a measure space. The measure μ is *finite* if $\mu(X)$ is finite. If $\mu\colon \mathcal{A} \to [0, 1]$ and if $\mu(X) = 1$, then $(X, \mathcal{A}, \mu)$ is called a *probability space*. The measure μ is a *σ-finite* if there exist a countable family $(A_i)_{i\geq 1}$ of subsets $A_i \in \mathcal{A}$ such that $X = \bigcup_{i=1}^{\infty} A_i$, and $\mu(A_i)$ is finite for all $i \geq 1$; see Figure 4.7. The measure μ is *complete* if for all $A \in \mathcal{A}$, if $\mu(A) = 0$, then $B \in \mathcal{A}$ for all $B \subseteq A$. A subset $A \in \mathcal{A}$ such that $\mu(A) = 0$ is called a *set of measure zero*.

Example 4.3. Let X be any nonempty set, and consider the σ-algebra $\mathcal{A} = 2^X$. The map $\mu\colon 2^X \to [0, +\infty]$ given by
$$\mu(A) = \begin{cases} |A| & \text{if } A \text{ is finite} \\ +\infty & \text{if } A \text{ is infinite} \end{cases}$$
is a measure called the *counting measure* on X.

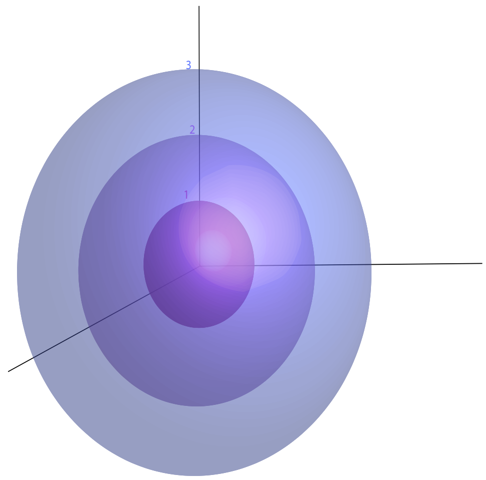


Figure 4.7: Let $X = \mathbb{R}^3$ and let μ be the Lebesgue measure on $\mathbb{R}^3$. Then X is σ-finite since $X = \bigcup_{i=1}^{\infty} A_i$, where $A_i = \{x \in \mathbb{R} \mid ||x|| \leq i\}$. The illustration shows the solid spheres A_1, A_2, and A_3.

Here is a summary of useful properties of measures.

Proposition 4.7. *Let $(X, \mathcal{A}, \mu)$ be a measure space. The following properties hold:*

(1) For any finite sequence $(A_1, \ldots, A_n)$ of subsets $A_i \in \mathcal{A}$ such that $A_i \cap A_j = \emptyset$ whenever $i \neq j$, we have

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i).$$

(2) For any two subsets A, B of X, if $A, B \in \mathcal{A}$ and if $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

(3) For any countable sequence $(A_i)_{i\geq 1}$ of subsets $A_i \in \mathcal{A}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

(4) For any countable sequence $(A_i)_{i\geq 1}$ of subsets $A_i \in \mathcal{A}$, if $A_{i+1} \subseteq A_i$ for all $i \geq 1$ and if $\mu(A_1)$ is finite, then

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \mapsto \infty} \mu(A_n).$$

(5) For any countable sequence $(A_i)_{i\geq 1}$ of subsets $A_i \in \mathcal{A}$, if $A_i \subseteq A_{i+1}$ for all $i \geq 1$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \mapsto \infty} \mu(A_n).$$

Proof. The proof of (1) and (2) is identical to the proof given just before Proposition 4.1, and (3) is Proposition 4.1. We prove (4), leaving the proof of (5) as an exercise.

We can write

$$A_n = \left(\bigcap_{i=1}^{\infty} A_i\right) \cup \left(\bigcup_{i=n}^{\infty} (A_i - A_{i+1})\right),$$

a union of pairwise disjoint subsets since $A_{i+1} \subseteq A_i$ for all $i \geq 1$. By ($\mu 3$), we have

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) + \sum_{i=n}^{\infty} \mu(A_i - A_{i+1}) = \mu(A_n) \leq \mu(A_1) < +\infty,$$

since $A_{i+1} \subseteq A_i$ for all $i \geq 1$ and since $\mu(A_1)$ is assumed to be finite. See Figure 4.8.

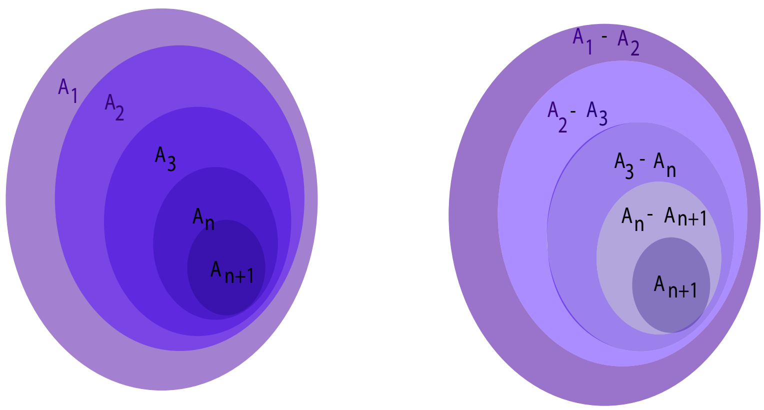


Figure 4.8: Decomposing the decreasing nested sequences of periwinkle sets into disjoint rings. Note $A_1 = (A_1 - A_2) \cup (A_2 - A_3) \cup \cdots \cup (A_n - A_{n+1}) \cup A_{n+1}$.

Consequently, for $n = 1$ we deduce that the series $\sum_{i=1}^{\infty} \mu(A_i - A_{i+1})$ converges, which implies that

$$\lim_{n \mapsto \infty} \sum_{i=n}^{\infty} \mu(A_i - A_{i+1}) = 0.$$

Since

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) + \sum_{i=n}^{\infty} \mu(A_i - A_{i+1}) = \mu(A_n),$$

we conclude that $\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \mapsto \infty} \mu(A_n)$. ☐

The following result shows that every measure can be completed; this is technically useful.

Proposition 4.8. *Let $(X, \mathcal{A}, \mu)$ be a measure space. A measure space $(X, \overline{\mathcal{A}}, \overline{\mu})$ with the following properties can be constructed:*

(a) $\mathcal{A} \subseteq \overline{\mathcal{A}}$.

(b) $\overline{\mu}$ *extends* μ*; that is,* $\overline{\mu}(A) = \mu(A)$ *for all* $A \in \mathcal{A}$.

(c) The measure $\overline{\mu}$ *is complete.*

We force the completeness property by defining $\overline{\mathcal{A}}$ as follows:

$$\overline{\mathcal{A}} = \{\overline{A} \subseteq X \mid (\exists A, A' \in \mathcal{A})(\exists B \subseteq A')(\overline{A} = A \cup B,\ \mu(A') = 0)\}.$$

The measure $\overline{\mu}$ is defined such that

$$\overline{\mu}(\overline{A}) = \overline{\mu}(A \cup B) = \mu(A).$$

See Figure 4.9.

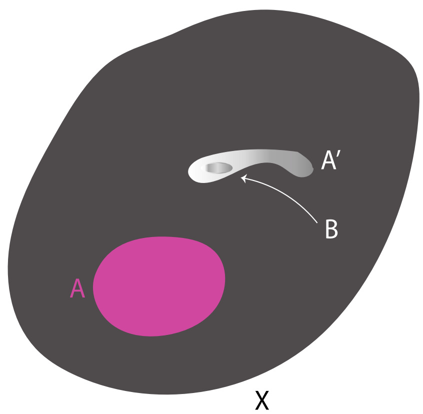


Figure 4.9: A schematic illustration of a set in $\overline{\mathcal{A}}$. The magenta set A has positive measure, while the grayish set A', and all of its subsets, including B, have zero measure. Then $\overline{A} = A \cup B$.

Proposition 4.8 is proven in Rudin [79] (Theorem 1.36). The verification that $\overline{\mathcal{A}}$ is a σ-algebra with the required properties and that $\overline{\mu}$ is a measure with the required properties is tedious (among other things, one needs to check that $\overline{\mu}(\overline{A})$ does not depend on the representation of $\overline{A}$).

Definition 4.11. The measure $\overline{\mu}$ given by Proposition 4.8 is called the *completed measure* of μ.

4.3 Null Subsets and Properties Holding Almost Everywhere

One of the secrets of measure theory is that subsets of measure zero should be ignored. Since a measure is not necessarily complete the correct technical definition is as follows.

Definition 4.12. Let $(X, \mathcal{A}, \mu)$ be a measure space. A subset $E \subseteq X$ is *null*[1] (or *negligeable*) if there is some $A \in \mathcal{A}$ such that $E \subseteq A$ and $\mu(A) = 0$. A property P of the elements of X *holds almost everywhere*, abbreviated *holds a.e.*, if the subset were it fails is null; that is, the set $\{x \in X \mid P(x) = \mathbf{false}\}$ is null.

To be very precise, we should say *μ-null* and *holds μ-a.e.*, since these notions depend on the measure μ. In most cases there is no risk of confusion, and we drop μ.

Observe that if the measure μ is complete, then a subset $E \subseteq X$ is *null* iff $\mu(E) = 0$, and a property P *holds a.e.* iff $\mu(\{x \in X \mid P(x) = \mathbf{false}\}) = 0$. In general, a null set may either be measurable or nonmeasurable, and a nonmeasurable set has no reason to be null, but may be null.

Here are a few properties of null sets.

Proposition 4.9. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Every subset of a null set is null. Every countable union of null sets is a null set.*

Proof. The first property follows immediately from the definition. Let $(A_i)_{i \geq 1}$ be a countable family of null sets. There are subsets $B_i \in \mathcal{A}$ such that $A_i \subseteq B_i$ and $\mu(B_i) = 0$ for all $i \geq 1$. We have

$$\bigcup_{i=1}^{\infty} A_i \subseteq \bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$$

because $\mathcal{A}$ is a σ-algebra, so it remains to show that $\bigcup_{i=1}^{\infty} B_i$ has measure zero. For this, observe that

$$0 \leq \mu\left(\bigcup_{i=1}^{\infty} B_i\right) \leq \sum_{i=1}^{\infty} \mu(B_i) = 0,$$

so $\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = 0$, as desired. $\square$

Let P and P' be two properties of X. If P implies P' and if P holds a.e., then P' holds a.e.

Definition 4.13. Consider the set of functions $f \colon X \to \mathbb{R}$, where $(X, \mathcal{A}, \mu)$ is a measure space. We say that *f and g are equal a.e.* if the set $\{x \in X \mid f(x) \neq g(x)\}$ is null. Write $f = g$ (a.e.).

It is an easy exercise to show that equality a.e. is an equivalence relation.

It should be observed that the notion of equality a.e. is more subtle than one might think.

[1]Beware that in measure theory, the notion of null set has more than one meaning. Some authors mean something different from what we define here.

Example 4.4. For example, consider the function $\chi_{\mathbb{Q}}\colon \mathbb{R} \to \mathbb{R}$, given by

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}; \end{cases}$$

see Figure 4.10.

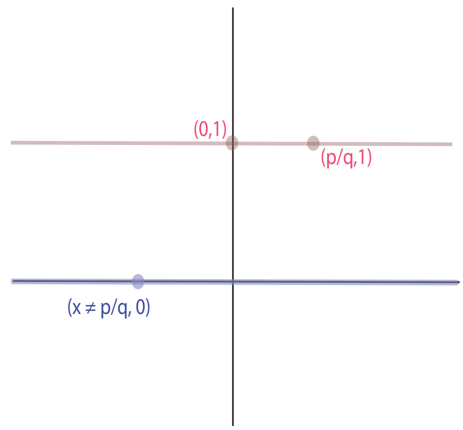


Figure 4.10: The graph of $\chi_{\mathbb{Q}}$. The points on the light brown line have rational x-coordinates while the points on the light gray line have irrational x-coordinates.

In other words, $\chi_{\mathbb{Q}}$ is the characteristic function of the rationals. It is easy to see that $\chi_{\mathbb{Q}}$ is discontinuous at every point $x \in \mathbb{R}$ (if x is irrational, then every small interval containing x contains some rational number; similarly, if x is rational, then every small interval containing x contains some irrational number, say of the form $x + \frac{\sqrt{2}}{2^n}$ for n large enough). Now, the Lebesgue measure μ_L discussed in Section 4.5 has the property that every countable set has measure zero, so in particular $\mathbb{Q}$ has Lebesgue measure zero. It follows that $\chi_{\mathbb{Q}}$ is equal to the zero function (on $\mathbb{Q}$) a.e., and the zero function is a "very nice" function; it is infinitely differentiable.

This is the beauty of equality a.e. Given a "very bad" function, we can ignore its bad behavior on a set of measure zero, as least from the point of view of integration.

An interesting variation of $\chi_{\mathbb{Q}}$ is the following function $D_{\mathbb{Q}}\colon \mathbb{R} \to \mathbb{R}$, given by

$$D_{\mathbb{Q}}(x) = \begin{cases} \frac{1}{q} & \text{if } x = p/q \in \mathbb{Q},\ q > 0,\ p \neq 0,\ \gcd(p,q) = 1, \\ 0 & \text{if } x \notin \mathbb{Q}, \\ 1 & \text{if } x = 0. \end{cases}$$

It is easy to show that $D_{\mathbb{Q}}$ is discontinuous at every rational point x, but is continuous at every irrational point x. In fact, $D_{\mathbb{Q}}$ is a regulated function. Again $D_{\mathbb{Q}}$ is equal to the zero function a.e. (with respect to the Lebesgue measure).

A property that will play an important role is *pointwise convergence a.e.*

Definition 4.14. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let F be any topological space (in most cases a normed vector space). A sequence $(f_n)_{n\geq 1}$ of functions $f_n \colon X \to F$ *converges pointwise a.e.* to a function $f \colon X \to F$ if there is a null set $Z \subseteq X$ such that the sequence $(f_n(x))_{n\geq 1}$ converges to $f(x)$ for all $x \in X - Z$. See Figure 4.11.

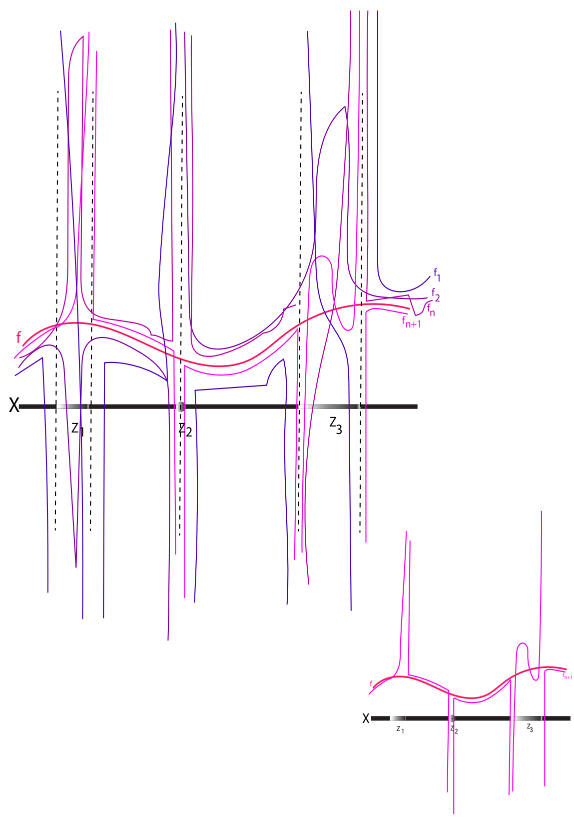


Figure 4.11: A schematic illustration of pointwise convergence a.e. Let X be the solid black line, $F = \mathbb{R}$, and $Z = Z_1 \cup Z_2 \cup Z_3$, where each Z_i has measure zero. The sequence (f_n) converges pointwise to the graph f (in red) for all $x \in X - Z$. As shown in the bottom right corner, the magenta graph f_{n+1} is "close" to f outside of Z.

4.4 Construction of a Measure from an Outer Measure

It turns out that defining explicitly a function m satisfying Conditions (2) and (3) from the beginning of Section 4.1 on a σ-algebra is not easy, but defining a function μ^* on 2^X

satisfying (1), (2), and (3'), is quite easy. Furthermore, given such a function μ^*, called an outer measure, there is a way of generating a σ-algebra and a measure on it.

If X is a locally compact topological space, then there is a way to construct a σ-algebra and a function m satisfying (2) and (3) on this σ-algebra using *Radon functionals*. This method will be explored in Chapter 7.

We now consider Option 1 from Section 4.1 and define outer measures.

Definition 4.15. Given a nonempty set X, an *outer measure* μ^* on X is a function satisfying the following properties:

(μ^*1) $\mu^*\colon 2^X \to [0, +\infty]$, that is, μ^* is defined for *all* subsets of X.

(μ^*2) $\mu^*(\emptyset) = 0$.

(μ^*3) For any countable sequence $(A_i)_{i\geq 1}$ of subsets A_i of X,

$$\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i).$$

This property is called *σ-subadditivity*.

(μ^*4) If $A \subseteq B$, then $\mu^*(A) \leq \mu^*(B)$.

Example 4.5. (Outer measure of Dirac) Let X be any nonempty set, and let a be any point chosen in X. The map $\mu_a^*\colon 2^X \to [0, +\infty]$ given by

$$\mu_a^*(A) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{if } a \notin A \end{cases}$$

is an outer measure called the *outer measure of Dirac*.

Here is a simple way to construct outer measures.

Proposition 4.10. *Let X be a nonempty set, and $\mathfrak{I} \subseteq 2^X$ be a family of subsets with the following properties:*

(a) $\emptyset \in \mathfrak{I}$.

(b) For every subset A of X, there is a countably infinite sequence $(I_n)_{n\geq 1}$ of subsets $I_n \in \mathfrak{I}$ such that $A \subseteq \bigcup_{n=1}^{\infty} I_n$.

Moreover, let $\lambda\colon \mathfrak{I} \to [0, +\infty]$ be any function such that

(c) $\lambda(\emptyset) = 0$.

Then the map μ^* given by

$$\mu^*(A) = \inf\left\{\sum_{n=1}^{\infty} \lambda(I_n) \mid A \subseteq \bigcup_{n=1}^{\infty} I_n,\ I_n \in \mathfrak{I}\right\}$$

is an outer measure on X.

Proof. The verification of (μ^*1), (μ^*2), and (μ^*4) is immediate and left to the reader.

Let $(A_i)_{i\geq 1}$ be an arbitrary family of subsets A_i of X. We may assume that $\sum_{i=1}^{\infty} \mu^*(A_i) < +\infty$, since otherwise ($\mu^*3$) holds trivially. Then we have $\mu^*(A_i) < +\infty$ for all $i \geq 1$. By definition of $\mu^*(A_i)$ as an infimum, for every $\epsilon > 0$, for every fixed $i \geq 1$, there is a countable family $(I_{i_n})_{n\geq 1}$ of subsets $I_{i_n} \in \mathfrak{I}$ such that

$$A_i \subseteq \bigcup_{n=1}^{\infty} I_{i_n} \text{ and } \mu^*(A_i) \leq \sum_{n=1}^{\infty} \lambda(I_{i_n}) \leq \mu^*(A_i) + \frac{\epsilon}{2^i}.$$

Since, as shown in Figure 4.12,

$$\bigcup_{i=1}^{\infty} A_i \subseteq \bigcup_{i=1}^{\infty}\bigcup_{n=1}^{\infty} I_{i_n},$$

by definition of $\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right)$ as an infimum and since

$$\sum_{n=1}^{\infty} \lambda(I_{i_n}) \leq \mu^*(A_i) + \frac{\epsilon}{2^i},$$

we have

$$\begin{aligned}\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) &\leq \sum_{i=1}^{\infty}\sum_{n=1}^{\infty} \lambda(I_{i_n}),\\ &\leq \sum_{i=1}^{\infty} \mu^*(A_i) + \sum_{i=1}^{\infty} \frac{\epsilon}{2^i}\\ &= \sum_{i=1}^{\infty} \mu^*(A_i) + \epsilon,\end{aligned}$$

since

$$\sum_{i=1}^{\infty} \frac{1}{2^i} = \frac{1}{2}\left(\sum_{i=0}^{\infty} \frac{1}{2^i}\right) = \frac{1}{2} \times 2 = 1.$$

Since $\epsilon > 0$ is arbitrary, we conclude that

$$\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i),$$

which is (μ^*3). ∎

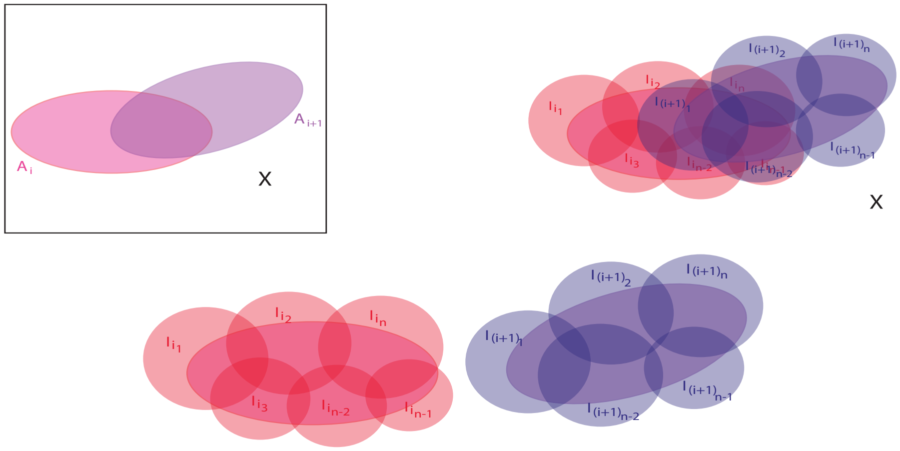


Figure 4.12: A Venn diagram illustration of $\bigcup_{i=1}^{\infty} A_i \subseteq \bigcup_{i=1}^{\infty} \bigcup_{n=1}^{\infty} I_{i_n}$.

As an application of Proposition 4.10, we obtain the outer Lebesgue measure.

Example 4.6. Let $\mathfrak{I}$ consist of the set of all open intervals (a, b), where $a = -\infty$ or $b = +\infty$ is allowed. It is easy to see that Properties (a) and (b) of Proposition 4.10 are satisfied. Let $\lambda\colon \mathfrak{I} \to [0, +\infty]$ be given by $\lambda((a, b)) = b - a$. Obviously, Property (c) holds. The outer measure given by Proposition 4.10 is called the *outer Lebesgue measure* μ_L^* on $\mathbb{R}$.

A similar construction could be performed on $\mathbb{R}^n$ by using products of open intervals $(a_1, b_1) \times \cdots \times (a_n, b_n)$ and $\lambda((a_1, b_1) \times \cdots \times (a_n, b_n)) = \prod_{i=1}^{n}(b_i - a_i)$.

We now state a fundamental theorem due to C. Carathéodory which gives a method for constructing a measure space from an outer measure.

Theorem 4.11. *(Carathéodory) Let $\mu^*\colon 2^X \to [0, +\infty]$ be an outer measure. Define the family $\mathcal{A}$ of subsets of X as follows:*

$$\mathcal{A} = \{A \in 2^X \mid \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap (X - A)), \text{ for all } E \subseteq X\}. \tag{C}$$

See Figure 4.13. Then the following properties hold:

(a) $\mathcal{A}$ is a σ-algebra which contains all subsets $A \subseteq X$ such that $\mu^*(A) = 0$.

(b) The restriction μ of μ^* to $\mathcal{A}$ is a measure. Furthermore, μ is a complete measure.

Proof. (a) To prove that $\mathcal{A}$ is a σ-algebra, we show that in order to prove the defining equation in (C) it suffices to prove the inequality (C') shown below. For this we prove

Claim 1. A subset A of X belongs to $\mathcal{A}$ if and only if, *for all $E \subseteq X$ such that $\mu^*(E) < +\infty$,*

$$\mu^*(E \cap A) + \mu^*(E \cap (X - A)) \leq \mu^*(E). \tag{C'}$$

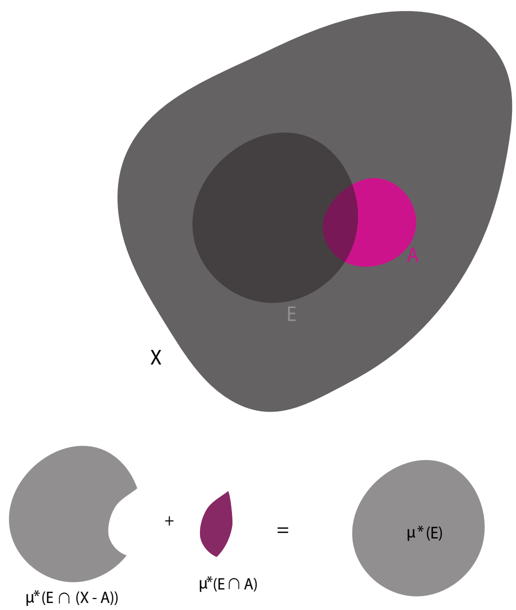


Figure 4.13: A schematic illustration of the Carathéodory construction of $\mathcal{A}$. The σ-algebra $\mathcal{A}$ consists of those magenta sets A which "cut" (with respect to μ^*) arbitrary subsets E in a "nice" manner.

Proof of Claim 1. We have $E = (E \cap A) \cup (E \cap (X - A))$ and by Condition (μ^*3),

$$\text{if } E = (E \cap A) \cup (E \cap (X - A)), \text{ then } \mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap (X - A)),$$

so it suffices to prove the reverse inequality when $\mu^*(E) < +\infty$, because if $\mu^*(E) = +\infty$, then

$$\mu^*(E \cap A) + \mu^*(E \cap (X - A)) = +\infty = \mu^*(E),$$

since both sides are equal to $+\infty$. $\square$

Next the proof consists of several steps.

Step 1. Verification of (A1). By (μ^*2), for every $E \subseteq X$, we have

$$\mu^*(E \cap X) + \mu^*(E \cap (X - X)) = \mu^*(E) + \mu^*(\emptyset) = \mu^*(E) + 0 = \mu^*(E),$$

which shows that X satisfies Equation (C), and thus $X \in \mathcal{A}$.

Step 2. Verification of (A2). This follows from the fact that Equation (C) implies that $A \in \mathcal{A}$ iff $X - A \in \mathcal{A}$.

Step 3. Verification of (σ-A3). We begin by verifying (σ-A3) for finite unions. Since by *Step 2*, $\mathcal{A}$ is closed under complementation, this shows that $\mathcal{A}$ is an algebra.

Step 3a. The case of any finite union $\bigcup_{i=1}^n A_i$ reduces by induction to the case where $n = 2$, so it suffices to prove that for all $A_1, A_2 \subseteq X$, if $A_1, A_2 \in \mathcal{A}$, then $A_1 \cup A_2 \in \mathcal{A}$. In view of *Claim 1*, this is equivalent to checking that for all $E \subseteq X$ (with $\mu^*(E) < +\infty$),

$$\mu^*(E \cap (A_1 \cup A_2)) + \mu^*(E \cap (X - (A_1 \cup A_2))) \leq \mu^*(E). \tag{$*_1$}$$

We begin by rewriting the terms $\mu^*(E \cap (A_1 \cup A_2))$ and $\mu^*(E \cap (X - (A_1 \cup A_2)))$. Since, (see Figure 4.14),

$$E \cap (A_1 \cup A_2) = (E \cap A_1) \cup (E \cap A_2) = (E \cap A_1) \cup (E \cap (X - A_1) \cap A_2),$$

by (μ^*3), we have

$$\mu^*(E \cap (A_1 \cup A_2)) \leq \mu^*(E \cap A_1) + \mu^*(E \cap (X - A_1) \cap A_2). \tag{$*_2$}$$

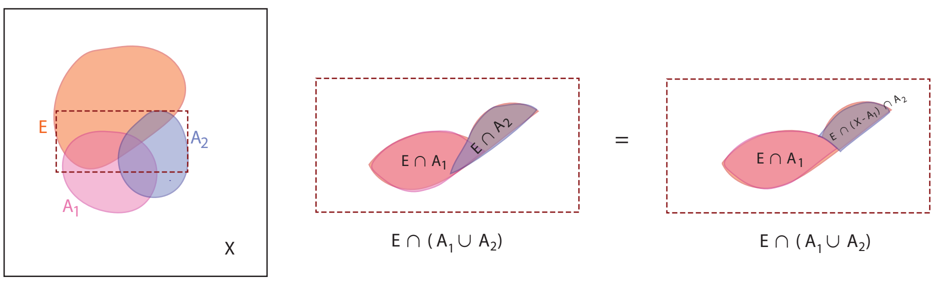


Figure 4.14: A Venn diagram illustration of $E \cap (A_1 \cup A_2) = (E \cap A_1) \cup (E \cap A_2) = (E \cap A_1) \cup (E \cap (X - A_1) \cap A_2)$.

Since, (see Figure 4.15), we also have

$$E \cap (X - (A_1 \cup A_2)) = E \cap (X - A_1) \cap (X - A_2),$$

by ($*_2$), we obtain

$$\begin{aligned}\mu^*(E \cap (A_1 \cup A_2)) + \mu^*(E \cap (X - (A_1 \cup A_2))) \leq \mu^*(E \cap A_1) + \mu^*(E \cap (X - A_1) \cap A_2)\\ + \mu^*(E \cap (X - A_1) \cap (X - A_2)).\end{aligned} \tag{$*_3$}$$

Since $A_1 \in \mathcal{A}$ and $A_2 \in \mathcal{A}$, for any $E \subseteq X$, by applying (C) to A_1 with E we have

$$\mu^*(E) = \mu^*(E \cap A_1) + \mu^*(E \cap (X - A_1)),$$

and by applying (C) to A_2 with $E \cap (X - A_1)$ we have

$$\begin{aligned}\mu^*(E \cap (X - A_1)) &= \mu^*((E \cap (X - A_1)) \cap A_2) + \mu^*((E \cap (X - A_1)) \cap (X - A_2))\\ &= \mu^*(E \cap (X - A_1) \cap A_2) + \mu^*(E \cap (X - A_1) \cap (X - A_2)),\end{aligned}$$

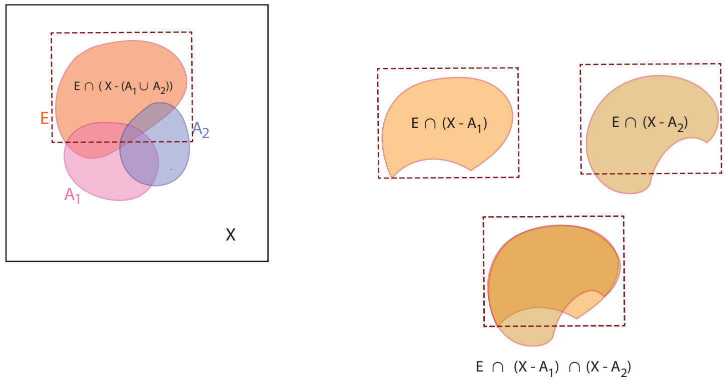


Figure 4.15: A Venn diagram illustration of $E \cap (X - (A_1 \cup A_2)) = E \cap (X - A_1) \cap (X - A_2)$.

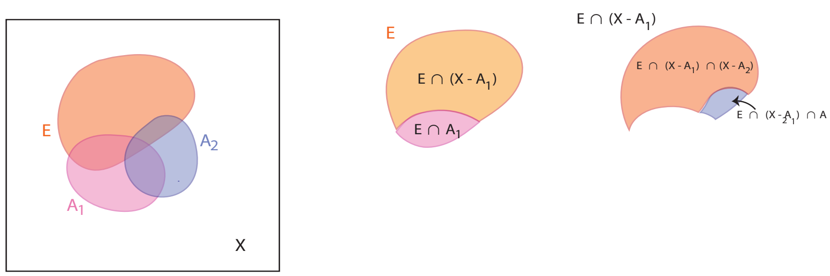


Figure 4.16: Venn diagram illustrations associated with the identities $\mu^*(E) = \mu^*(E \cap A_1) + \mu^*(E \cap (X - A_1))$ and $\mu^*(E \cap (X - A_1)) = \mu^*((E \cap (X - A_1)) \cap A_2) + \mu^*((E \cap (X - A_1)) \cap (X - A_2))$.

so we obtain

$$\mu^*(E) = \mu^*(E \cap A_1) + \mu^*(E \cap (X - A_1) \cap A_2) + \mu^*(E \cap (X - A_1) \cap (X - A_2)); \qquad (*_4)$$

see Figure 4.16.

Since the the right-hand sides of $(*_3)$ and $(*_4)$ are identical, we obtain

$$\mu^*(E \cap (A_1 \cup A_2)) + \mu^*(E \cap (X - (A_1 \cup A_2))) \leq \mu^*(E),$$

as desired.

Step 3b. We prove that (σ-A3) holds for countably infinite unions $B = \bigcup_{i \geq 1} B_i$, with $B_i \in \mathcal{A}$ and $B_i \cap B_j = \emptyset$ for all $i \neq j$, which means that we have to show that for all $E \subseteq X$ such that $\mu^*(E) < +\infty$, we have

$$\mu^*(E \cap B) + \mu^*(E \cap (X - B)) \leq \mu^*(E).$$

We begin by analyzing the term $\mu^*(E \cap B)$. Since $E \cap B = \bigcup_{i=1}^{\infty}(E \cap B_i)$, by ($\mu^*3$) we have

$$\mu^*(E \cap B) = \mu^* \left(\bigcup_{i=1}^{\infty} (E \cap B_i) \right) \leq \sum_{i=1}^{\infty} \mu^*(E \cap B_i),$$

which implies the inequality

$$\mu^*(E \cap B) + \mu^*(E \cap (X - B)) \leq \sum_{i=1}^{\infty} \mu^*(E \cap B_i) + \mu^*(E \cap (X - B)). \qquad (*_5)$$

Thus if we prove that

$$\sum_{i=1}^{\infty} \mu^*(E \cap B_i) + \mu^*(E \cap (X - B)) \leq \mu^*(E),$$

we are done. To deal with the infinite sum on the left-hand side we use *Step 3a*. By the result of *Step 3a*, we have $C_n = \bigcup_{i=1}^{n} B_i \in \mathcal{A}$ for all $n \geq 1$. Since, as shown in Figure 4.17, $E \cap (X - B) \subseteq E \cap (X - C_n)$, by (μ^*4), we have

$$\mu^*(E \cap C_n) + \mu^*(E \cap (X - B)) \leq \mu^*(E \cap C_n) + \mu^*(E \cap (X - C_n)) = \mu^*(E). \qquad (*_6)$$

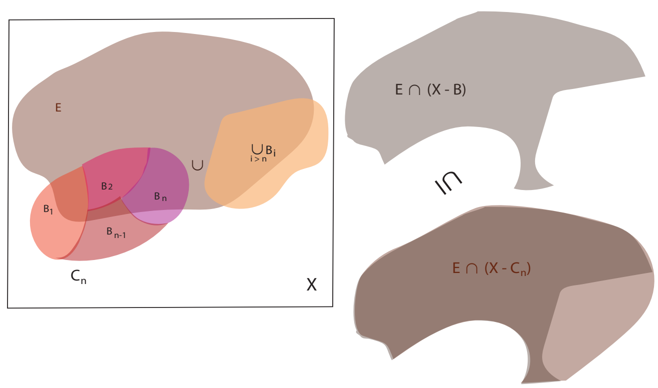


Figure 4.17: Venn diagram illustration of $E \cap (X - B) \subseteq E \cap (X - C_n)$.

On the other hand, since $B_i \in \mathcal{A}$, we can show by induction using the fact that $C_n = \bigcup_{i=1}^{n} B_i$ and the B_i are pairwise disjoint that

$$\begin{aligned}\mu^*(E \cap C_n) &= \mu^*(E \cap C_n \cap B_n) + \mu^*(E \cap C_n \cap (X - B_n)) \\ &= \mu^*(E \cap B_n) + \mu^*(E \cap C_{n-1}) \\ &= \sum_{i=1}^{n} \mu^*(E \cap B_i);\end{aligned}$$

see Figure 4.18.

Consequently, by $(*_6)$ and the above equation, we obtain

$$\mu^*(E \cap C_n) + \mu^*(E \cap (X - B)) = \sum_{i=1}^{n} \mu^*(E \cap B_i) + \mu^*(E \cap (X - B)) \leq \mu^*(E). \qquad (*_7)$$

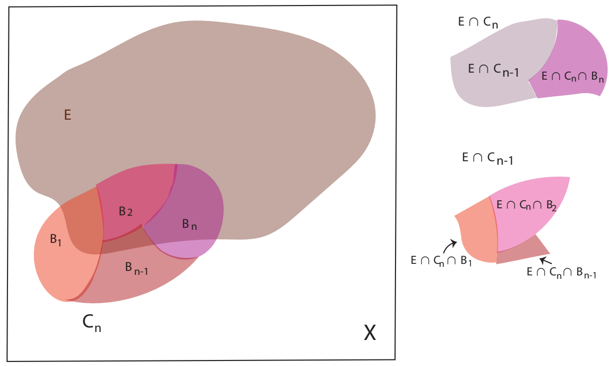


Figure 4.18: Venn diagram illustration associated with $\mu^*(E \cap C_n) = \sum_{i=1}^n \mu^*(E \cap B_i)$.

Since by hypothesis $\mu^*(E) < +\infty$ and by (μ^*4),

$$\mu^*(E \cap (X - B)) \leq \mu^*(E) < +\infty,$$

passing to the limit the inequality $(*_7)$ implies that

$$\sum_{i=1}^{\infty} \mu^*(E \cap B_i) < +\infty,$$

and also that

$$\sum_{i=1}^{\infty} \mu^*(E \cap B_i) + \mu^*(E \cap (X - B)) \leq \mu^*(E), \tag{$\dagger$}$$

as desired.

Step 3c. We prove that (σ-A3) holds for arbitrary countably infinite unions $A = \bigcup_{i \geq 1} A_i$, with $A_i \in \mathcal{A}$.

The trick (already used in the proof of Proposition 4.1) is to define the family $(B_i)_{i \geq 1}$ as follows:

$$\begin{aligned}
B_1 &= A_1 \\
B_i &= A_i - \left(\bigcup_{j=1}^{i-1} A_j\right) = X - \left(\left(\bigcup_{j=1}^{i-1} A_j\right) \cup (X - A_i)\right);
\end{aligned}$$

see Figure 4.3. Since $\mathcal{A}$ is an algebra, it is closed under finite unions and complementation, so $B_i \in \mathcal{A}$. Furthermore, by definition, $B_i \cap B_j = \emptyset$ for all $i \neq j$, and

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i,$$

so by *Step 3b*, we get $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$.

Therefore, we proved that $\mathcal{A}$ is a σ-algebra.

Step 4. Proving (a). If $\mu^*(A) = 0$, since by (μ^*4) we have

$$\mu^*(E \cap A) \leq \mu^*(A) = 0$$

and

$$\mu^*(E \cap (X - A)) \leq \mu^*(E),$$

we obtain

$$\mu^*(E \cap A) + \mu^*(E \cap (X - A)) \leq \mu^*(E)$$

for all $E \subseteq X$ such that $\mu^*(E) < +\infty$, which by *Claim 1* means that $A \in \mathcal{A}$.

(b) We prove that the restriction μ of μ^* to $\mathcal{A}$ is a measure, which means that we need to check Condition $(\mu 1)$, $(\mu 2)$ and $(\mu 3)$, which is achieved in three steps.

Step 5. Property $(\mu 1)$ is obvious.

Step 6. Since $\mathcal{A}$ is a σ-algebra, $\emptyset \in \mathcal{A}$, so by (μ^*2),

$$\mu(\emptyset) = \mu^*(\emptyset) = 0,$$

which is $(\mu 2)$.

Step 7. Let $B = \bigcup_{i=1}^{\infty} B_i$ be a countably infinite union of subsets $B_i \in \mathcal{A}$ which are pairwise disjoint. For all $E \subseteq X$ such that $\mu^*(E) < +\infty$, we proved in $(\dagger)$ that

$$\sum_{i=1}^{\infty} \mu^*(E \cap B_i) + \mu^*(E \cap (X - B)) \leq \mu^*(E).$$

If $\mu^*(B) < +\infty$, then we can let $E = B$ in the above inequality, and we get

$$\sum_{i=1}^{\infty} \mu^*(B_i) \leq \mu^*(B). \tag{$*_7$}$$

By (μ^*3), since $B = \bigcup_{i=1}^{\infty} B_i$, we also have

$$\mu^*(B) \leq \sum_{i=1}^{\infty} \mu^*(B_i). \tag{$*_8$}$$

Then $(*_7)$ and $(*_8)$ yield

$$\sum_{i=1}^{\infty} \mu^*(B_i) = \mu^*(B).$$

Since $B \in \mathcal{A}$ and $B_i \in \mathcal{A}$, $\mu(B) = \mu^*(B)$ and $\mu(B_i) = \mu^*(B_i)$, we get

$$\sum_{i=1}^{\infty} \mu(B_i) = \mu(B),$$

which is (μ3). If $\mu^*(B) = +\infty$, then $(*_8)$ implies that trivially

$$\sum_{i=1}^{\infty} \mu(B_i) = \sum_{i=1}^{\infty} \mu^*(B_i) = \mu^*(B) = \mu(B) = +\infty.$$

Step 8. Finally it remains to show that μ is a complete measure. Let $A \in \mathcal{A}$ such that $\mu(A) = 0$ and consider any subset $B \subseteq A$. Since $A \in \mathcal{A}$, by definition of μ,

$$\mu^*(A) = \mu(A) = 0,$$

and by (μ^*4), $B \subseteq A$ implies that

$$\mu^*(B) \leq \mu^*(A) = 0,$$

so $\mu^*(B) = 0$, and we proved in *Step 4* that $B \in \mathcal{A}$. Therefore, μ is a complete measure. $\square$

Example 4.7. If we apply Theorem 4.11 to the Dirac outer measure μ_a^* of Example 4.5, we find easily that $\mathcal{A} = 2^X$ and that $\mu = \mu_a^*$. The *Dirac measure* μ_a^* is usually denoted by δ_a.

If we apply Theorem 4.11 to the Lebesgue outer measure of Example 4.6, we obtain the *Lebesgue measure* on $\mathbb{R}$. It can be shown that the σ-algebra of *Lebesgue-measurable sets* obtained from the construction contains the σ-algebra of Borel sets of $\mathbb{R}$. This example is considered in slightly more details in the next section.

4.5 The Lebesgue Measure on $\mathbb{R}$

Recall that in Example 4.6 we defined the outer Lebesgue measure μ_L^* on $\mathbb{R}$. For this we considered the set $\mathfrak{I}$ consisting of all open intervals (a, b), where $a = -\infty$ or $b = +\infty$ is allowed. By Proposition 4.10 applied to the function $\lambda\colon \mathfrak{I} \to [0, +\infty]$ given by $\lambda((a, b)) = b - a$, we obtained the outer Lebesgue measure μ_L^* given by

$$\mu_L^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \lambda(I_n) \mid A \subseteq \bigcup_{n=1}^{\infty} I_n,\ I_n \in \mathfrak{I} \right\}.$$

By applying Theorem 4.11 to the outer measure μ_L^*, we obtain the σ-algebra $\mathcal{L}(\mathbb{R})$ of *Lebesgue-measurable sets*, and the *Lebesgue measure* μ_L.

The construction used by Theorem 4.11 yields very little explicit information regarding what the Lebesgue-measurable sets look like, but it is possible to describe some of them. In

particular, if $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra generated by the open sets of $\mathbb{R}$, it turns out that $\mathcal{B}(\mathbb{R}) \subset \mathcal{L}(\mathbb{R})$, a proper inclusion. Actually, every open subset of $\mathbb{R}$ can be expressed as a countable disjoint union of finite open intervals, so the Borel σ-algebra is generated by the open intervals (a, b). The following proposition gives convenient characterizations of the Borel sets.

Proposition 4.12. *The σ-algebra $\mathcal{B}(\mathbb{R})$ of Borel sets of $\mathbb{R}$ is generated by the following intervals:*

1. $[a, b]$, *with $a \leq b$ finite.*
2. $[a, b)$, *with $a \leq b$ finite.*
3. (a, ∞) *and* $(-\infty, a)$, *with a finite.*

Proof. (1) We know that the open intervals (a, b) generate $\mathcal{B}(\mathbb{R})$. We have

$$(a, \infty) = \bigcup_{n=1}^{\infty} (a, a+n), \quad (-\infty, a) = \bigcup_{n=1}^{\infty} (a-n, a),$$

so (a, ∞) and $(-\infty, a)$ are elements of $\mathcal{B}(\mathbb{R})$; see Figure 4.19.

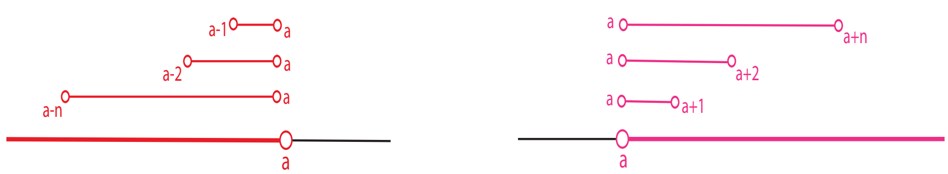


Figure 4.19: The left figure illustrates $(-\infty, a) = \bigcup_{n=1}^{\infty} (a-n, a)$, while the right figure illustrates $(a, \infty) = \bigcup_{n=1}^{\infty} (a, a+n)$.

Observe that

$$[a, b] = \overline{(-\infty, a)} \cap \overline{(b, \infty)} = [a, \infty) \cap (-\infty, b],$$

so $[a, b] \in \mathcal{B}(\mathbb{R})$. We also have

$$(a, b) = \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b - \frac{1}{n}\right],$$

so the closed intervals $[a, b]$ generate $\mathcal{B}(\mathbb{R})$; see Figure 4.20.

(2) We have

$$[a, b) = \bigcup_{n=1}^{\infty} \left[a, b - \frac{1}{n}\right],$$

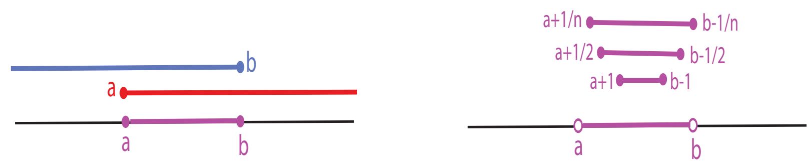


Figure 4.20: The left figure illustrates $[a, b] = [a, \infty) \cap (-\infty, b]$, while the right figure illustrates $(a, b) = \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b - \frac{1}{n}\right]$.

so $[a, b) \in \mathcal{B}(\mathbb{R})$. Then

$$(a, b) = \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b\right),$$

so the intervals $[a, b)$ generate $\mathcal{B}(\mathbb{R})$. See Figure 4.21.

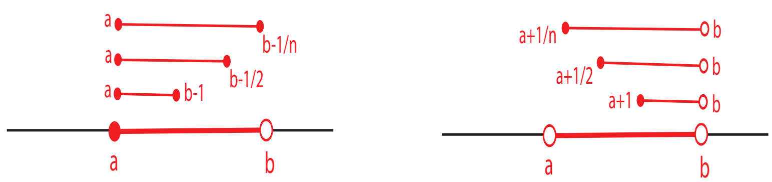


Figure 4.21: The left figure illustrates $[a, b) = \bigcup_{n=1}^{\infty} \left[a, b - \frac{1}{n}\right]$, while the right figure illustrates $(a, b) = \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}, b\right)$.

(3) We already know from (1) that $(a, \infty) \in \mathcal{B}(\mathbb{R})$. This implies that

$$(-\infty, a] = \overline{(a, \infty)} \in \mathcal{B}(\mathbb{R}),$$

so

$$(-\infty, a) = \bigcup_{n=1}^{\infty} \left(-\infty, a - \frac{1}{n}\right] \in \mathcal{B}(\mathbb{R}),$$

and thus

$$(a, b) = (-\infty, b) \cap (a, \infty),$$

so the intervals (a, ∞) generate $\mathcal{B}(\mathbb{R})$. See Figure 4.22. $\square$

Let's use the notation $\wr a, b \wr$ to denote any of the four types of intervals $(a, b), [a, b), (a, b]$, and $[a, b]$ (with $a = -\infty$ or $b = +\infty$ allowed, and $a = b$ allowed). The following result can be shown.

Theorem 4.13. *Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra of open sets, $\mathcal{L}(\mathbb{R})$ be the σ-algebra of Lebesgue-measurable sets, and μ_L be the Lebesgue measure for $\mathbb{R}$.*

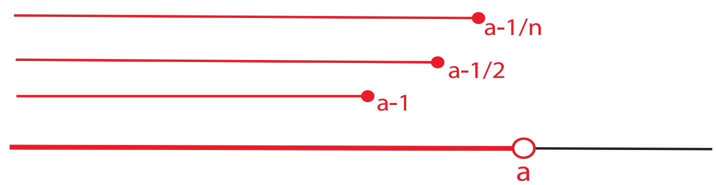


Figure 4.22: An illustration of the identity $(-\infty, a) = \bigcup_{n=1}^{\infty} \left(-\infty, a - \frac{1}{n}\right]$.

(1) $\mathcal{L}(\mathbb{R}) \neq 2^{\mathbb{R}}$; *that is, there exist non-measurable sets. The proof requires the axiom of choice.*

(2) $\mathcal{B}(\mathbb{R}) \subset \mathcal{L}(\mathbb{R})$, *the inclusion being strict. This is because* $|\mathcal{B}(\mathbb{R})| = 2^{\aleph_0} = \mathfrak{c}$, *but* $|\mathcal{L}(\mathbb{R})| = 2^{\mathfrak{c}}$.

(3) The Borel σ-algebra contains all four types of intervals, and

$$\mu_L(\wr a, b \wr) = \begin{cases} b - a & \textit{if } a \neq -\infty \textit{ and } b \neq +\infty \\ +\infty & \textit{if } a = -\infty \textit{ or } b = +\infty. \end{cases}$$

(4) The restriction of the Lebesgue measure μ_L to the Borel σ-algebra is a measure μ_B. The completion of the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_B)$ given by Proposition 4.8 gives back the measure space $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \mu_L)$ of Lebesgue-measurable sets.

The proofs for most parts of Theorem 4.13 are given in Halmos [44] (some of them as exercises). The fact that $|\mathcal{B}(\mathbb{R})| = 2^{\aleph_0}$ follows from the fact that $\mathcal{B}(\mathbb{R})$ is generated by the open intervals (a, b) and the remark just before Definition 4.6. It is surprising how much work it takes to prove Part (3) of Theorem 4.13. See also Folland [34] and Rudin [79].

As a corollary, every one-point set $\{a\}$ has Lebesgue measure 0, and thus every countable subset has Lebesgue measure 0. There are also uncountable subsets of Lebesgue measure 0. The *Cantor set* is such an example; see Folland [34], Section 1.5.

The Lebesgue measure also has the following regularity properties which show that every Lebesgue-measurable set can be approximated either by an open set or by a closed set; see Folland [34] (Section 1.5).

Proposition 4.14. *For any subset A of $\mathbb{R}$, we have*

$$\mu_L^*(A) = \inf\{\mu_L(O) \mid A \subseteq O,\ O \textit{ is open}\}.$$

For every Lebesgue-measurable set $A \in \mathcal{L}(\mathbb{R})$, the following facts hold:

(a) For every $\epsilon > 0$, there is some open subset O such that $A \subseteq O$ and $\mu_L(O - A) < \epsilon$.

(b) For every $\epsilon > 0$, there is some closed subset F such that $F \subseteq A$ and $\mu_L(A - F) < \epsilon$.

As a corollary of Proposition 4.14 we have the following facts.

Proposition 4.15. *For every Lebesgue-measurable set $A \in \mathcal{L}(\mathbb{R})$:*

(a') $\mu_L(A) = \inf\{\mu_L(O) \mid A \subseteq O,\ O \text{ is open}\}$.

(b') $\mu_L(A) = \sup\{\mu_L(F) \mid F \subseteq A,\ F \text{ is closed}\}$;

see Figure 4.23.

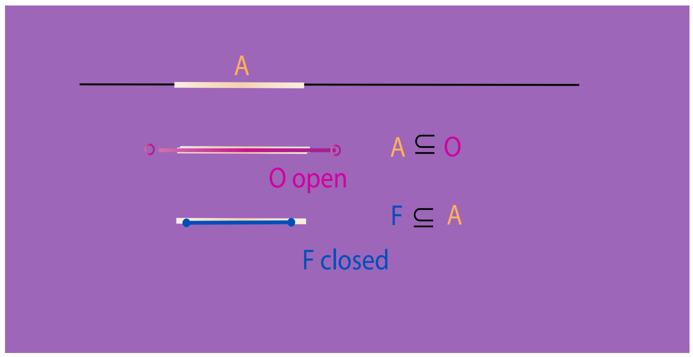


Figure 4.23: A Lebesgue-measurable set A of $\mathbb{R}$ is approximated from the "outside" by an open set O; it is also approximated from the "inside" by a closed set F.
.

It should be noted that Properties (a') and (b') are weaker than Properties (a) and (b), because they imply Properties (a) and (b) only when $\mu(A)$ is finite.

It can also be shown that for every Lebesgue-measurable set $A \in \mathcal{L}(\mathbb{R})$, we have

$$\mu_L(A) = \sup\{\mu_L(K) \mid K \subseteq A,\ K \text{ is compact}\}.$$

Proposition 4.14 also holds for the Lebesgue-measurable subsets of $\mathbb{R}^n$.

Another important property of the Lebesgue measure is that it is translation-invariant.

Proposition 4.16. *For any Lebesgue measurable set $A \in \mathcal{L}(\mathbb{R})$, we have $\mu_L(x+A) = \mu_L(A)$ for all $x \in \mathbb{R}$, where $x + A = \{x + a \mid a \in \mathbb{R}\}$. This property is called translation-invariance.*

For a proof, see Section 8.5, Example 8.1.

Proposition 4.17. *There is no translation-invariant measure μ defined on all subsets of $\mathbb{R}$ such that $\mu([0,1)) \neq 0$ and $\mu([0,1)) \neq +\infty$. As a consequence, there is no translation-invariant measure defined on all subsets of $\mathbb{R}$ such that $\mu([a,b)] = b - a$.*

Proof. To prove the proposition, we consider the quotient set $\mathbb{R}/\mathbb{Q}$ of the reals modulo the equivalence relation $x \sim y$ iff $x - y \in \mathbb{Q}$. Using the axiom of choice, we can form a subset $E \subseteq [0, 1)$ which contains exactly one number from each equivalence class of $\mathbb{R}/\mathbb{Q}$. Let $R = \mathbb{Q} \cap [0, 1)$, and for each $r \in R$, let

$$E_r = \{x + r \mid x \in E \cap [0, 1 - r)\} \cup \{x + r - 1 \mid x \in E \cap [1 - r, 1)\};$$

see Figure 4.24. Clearly $E_r \subseteq [0, 1)$, and we claim that every $x \in [0, 1)$ belongs to some unique E_r.

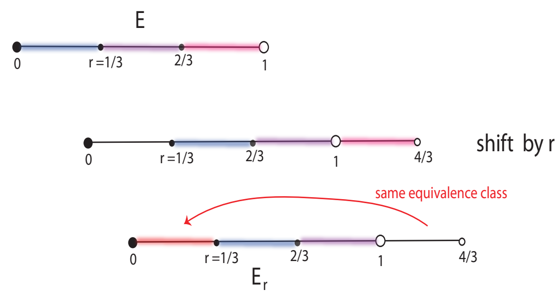


Figure 4.24: The construction of E_r.

Indeed, if $y \in E$ belongs to the equivalence class of $x \in [0, 1)$, then $x \in E_r$ where $r = x - y$ if $x \geq y$ or $r = x - y + 1$ if $x < y$. Furthermore, if $x \in E_r \cap E_s$ with $r \neq s$, then $x - r$ (or $x - r + 1$) and $x - s$ (or $x - s + 1$) would be distinct elements of E belonging to the same equivalence class, which is impossible (since $r, s \in R \subset \mathbb{Q}$). It follows that $[0, 1)$ is the countable disjoint union of the E_r. If a translation-invariant measure μ exists, then for any $r \in R$ we have

$$\mu(E) = \mu(E \cap [0, 1 - r)) + \mu(E \cap [1 - r, 1)) = \mu(E_r).$$

Since $[0, 1)$ is the countable disjoint union of the E_r,

$$\mu([0, 1)) = \sum_{r \in E} \mu(E_r) = \sum_{r \in E} \mu(E).$$

Now by assumption $\mu([0, 1)) \neq 0$ and $\mu([0, 1)) \neq +\infty$, but the sum on the right-hand side is either 0 if $\mu(E) = 0$ or $+\infty$ otherwise, a contradiction. $\square$

The above proof also implies that E is an uncountable subset of $[0, 1)$ which is not Lebesgue measurable (since the Lebesgue measure is translation-invariant).

We conclude by mentioning that if X is a topological space, given a function μ defined on the open subsets and the compact subsets of X, we can define the following maps for *every* subset A of X:

$$\begin{aligned}
\mu^*(A) &= \inf\{\mu(O) \mid A \subseteq O,\ A \text{ is open}\}\\
\mu_*(A) &= \sup\{\mu(K) \mid K \subseteq A,\ K \text{ is compact}\}.
\end{aligned}$$

Then the measurable subsets are those subsets A of X such that

$$\mu^*(A) = \mu_*(A).$$

It can be shown that these subsets form a σ-algebra $\mathcal{A}$, and that the map μ with domain $\mathcal{A}$ given by $\mu(A) = \mu^*(A) = \mu_*(A)$ is a measure. This is the approach using Radon measures.

Chapter 5

Integration

Given a measure space $(X, \mathcal{A}, \mu)$, we would like to define the integral of a real-valued function $f\colon X \to \mathbb{R}$, or more generally of a complex-valued function $f\colon X \to \mathbb{C}$, or even of a function $f\colon X \to F$, where F is a normed vector space. The key idea is that the integral of a very simple function f, such as a function taking only a finite number of nonzero values $y_1, \ldots, y_n$, should be "obvious." Namely, if $A_i = f^{-1}(y_i)$ is the subset of X over which f has the value y_i, then each A_i should be measurable (that is, $A_i \in \mathcal{A}$), and A_i should have finite measure, so that the expression

$$\sum_{i=1}^{n} y_i \mu(A_i) \in F$$

makes sense. Then we define the integral $\int f d\mu$ of our simple function f as

$$\int f d\mu = \sum_{i=1}^{n} \mu(A_i) y_i. \tag{$*$}$$

Observe that the function f can be written as

$$f = \sum_{i=1}^{n} y_i \chi_{A_i},$$

where χ_{A_i} is the characteristic function of the subset A_i. Such a function is called a *μ-step function*.

Observe that $(*)$ is a generalization of the notion of area under the curve. If the subsets A_i are closed adjacent intervals, then we are back to the notion of Riemann integral. However, in our new setting, the subsets A_i can be very complicated, but as long as they are measurable and have finite measure, the integral $(*)$ makes sense.

If we define $\|f\|$ as

$$\|f\| = \sum_{i=1}^{n} \|y_i\| \chi_{A_i},$$

(remember that our set F of values is a normed vector space), then the integral of $\|f\|$ is

$$\int \|f\|\, d\mu = \sum_{i=1}^{n} \mu(A_i)\, \|y_i\| \in \mathbb{R}_+.$$

If we define $N_1(f) = \int \|f\|\, d\mu$, then N_1 satisfies all the properties of a norm, except that $N_1(f) = 0$ does not necessarily imply that $f = 0$. However, $N_1(f) = 0$ iff $f = 0$ almost everywhere. The set $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ of μ-step functions is a vector space, and N_1 is almost a norm on it; it is a semi-norm. The integral given by $(*)$ is a linear continuous map on $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$. However, the space $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ is not Cauchy-complete under the semi-norm N_1 (there are Cauchy sequences with respect to N_1 that do not have a limit). The problem then is to complete the space $\mathcal{S}tep_\mu(X, \mathcal{A}, \mu)$ and to extend the integral $(*)$ to this bigger set of functions.

There are several ways to proceed.

(1) If we let $\mathcal{SN}$ be subspace of $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ consisting of the μ-step functions equal to 0 a.e., then the quotient space $\mathrm{Step}_\mu(X, \mathcal{A}, F) = \mathcal{S}tep_\mu(X, \mathcal{A}, \mu)/\mathcal{SN}$ is a vector space and N_1 induces a (true) norm on it. Therefore we can apply the general completion theorem (Theorem A.72) to obtain a complete normed vector space $(\mathrm{L}_\mu(X, \mathcal{A}, F), \|\,\|_1)$. Since integration is a linear continuous map on $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$, it extends uniquely to a linear continuous map on $\mathrm{L}_\mu(X, \mathcal{A}, F)$.

In theory we have achieved our goal of defining a complete normed vector space of functions containing the μ-step functions for which every function is integrable. However, the completion $(\mathrm{L}_\mu(X, \mathcal{A}, F), \|\,\|_1)$ is a very complicated object. It consists of equivalence classes of Cauchy sequences of functions in the quotient space $\mathrm{Step}_\mu(X, \mathcal{A}, F)$. It would be much more convenient if the objects in $\mathrm{L}_\mu(X, \mathcal{A}, F)$ could be described as functions, and this is indeed possible.

(2) The second approach is to first define a set $\mathcal{L}_\mu(X, \mathcal{A}, F)$ of *functions* using a limit process. Every function f in $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is the limit pointwise a.e. of a N_1-Cauchy sequence $(f_n)_{n\geq 1}$ (called an approximation sequence) of functions f_n in $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$. We also define the space $\mathcal{M}_\mu(X, \mathcal{A}, F)$ of *μ-measurable functions*, and $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is the subspace of $\mathcal{M}_\mu(X, \mathcal{A}, F)$ consisting of the functions for which the integral is well defined.

It turns out that $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is complete with respect to an extension $\|\,\|_1$ of the semi-norm N_1, and the integral $\int f d\mu$ of any function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ can be defined by a limit process. There are technical complications when F is infinite-dimensional, and it also takes some work to show that the integral of a function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ does not depend on the approximation sequence used to define f, but all difficulties can be overcome. Finally, the subspace $\mathcal{N}$ of functions f such that $\|f\|_1 = 0$ is the set of functions equal to 0 a.e., and we obtain the complete space $(\mathrm{L}_\mu(X, \mathcal{A}, F), \|\,\|_1)$ of

the first approach as the quotient space $\mathcal{L}_\mu(X, \mathcal{A}, F)/\mathcal{N}$. However, the construction of $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is much more informative.

We also investigate convergence properties of $\mathcal{L}_\mu(X, \mathcal{A}, F)$, as well as other related spaces (the spaces $\mathcal{L}^p_\mu(X, \mathcal{A}, F)$, $p = 1, 2, \infty$). We conclude with the construction of the integral on a product space.

The vector valued-integral defined in this chapter (where the space F of values is a Banach space) was first discovered by Bochner in 1933. The version discussed here is due to Dunford (1935), and is presented in detail in Dunford and Schwartz [30]. More recent expositions of this method are given in Lang [62] and Marle [69].

5.1 Measurable Maps

Measurable functions are functions between measurable spaces that are the analog of continuous functions between topological spaces, but as we will see, they are a lot more flexible, especially in terms of convergence properties. In this chapter our presentation follows Marle [69] and Lang [62] very closely.

Definition 5.1. Given any two measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, a function $f\colon X \to Y$ is *measurable* if $f^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$. A measurable function is also called a *measurable map*.

If $(X, \mathcal{A})$ is a measurable space, then obviously the identity $\mathrm{id}\colon X \to X$ is measurable.

The composition of two measurable maps is also measurable.

Proposition 5.1. *Given three measurable spaces $(X, \mathcal{A})$, $(Y, \mathcal{B})$, and $(Z, \mathcal{C})$, if $f\colon X \to Y$ and $g\colon Y \to Z$ are measurable maps, then $g \circ f\colon X \to Z$ is a measurable map.*

Proof. Recall that one of the properties of inverse images is that $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$, for any subset C of Z. But if $C \in \mathcal{C}$, since g is measurable, $g^{-1}(C) \in \mathcal{B}$, and since f is measurable, $f^{-1}(g^{-1}(C)) \in \mathcal{A}$, which shows that $g \circ f$ is measurable. $\square$

Remark: The above properties show that measurable spaces are the objects of a category whose morphisms are the measurable maps.

Proposition 5.2. *Let X and Y be any two nonempty sets, and let $f\colon X \to Y$ be a function between them.*

(1) If $\mathcal{A}$ is a σ-algebra on X, then we can define $\mathcal{A}_f$ as the family of subsets of Y given by

$$\mathcal{A}_f = \{B \in 2^Y \mid f^{-1}(B) \in \mathcal{A}\}.$$

Then $\mathcal{A}_f$ is the largest σ-algebra on Y which makes f measurable.

(2) If $\mathcal{B}$ is a σ-algebra on Y, then let $f^{-1}(\mathcal{B})$ be the family of subsets of X given by

$$f^{-1}(\mathcal{B}) = \{f^{-1}(B) \in 2^X \mid B \in \mathcal{B}\}.$$

Then $f^{-1}(\mathcal{B})$ is the smallest σ-algebra on X which makes f measurable.

The proof of Proposition 5.2 is left as an exercise.

Using Proposition 5.2 we obtain the following proposition which gives simple criteria to check that a map is measurable.

Proposition 5.3. *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable spaces.*

(1) If $\mathcal{S}$ generates the σ-algebra $\mathcal{B}$ (which means that the smallest σ-algebra containing $\mathcal{S}$ is $\mathcal{B}$), then a function $f \colon X \to Y$ is measurable iff $f^{-1}(S) \in \mathcal{A}$ for all $S \in \mathcal{S}$.

(2) If Y is a topological space and if $\mathcal{B}$ is its Borel σ-algebra of open subsets, then a function $f \colon X \to Y$ is measurable iff $f^{-1}(U) \in \mathcal{A}$ for every open subset U of Y (or $f^{-1}(U) \in \mathcal{A}$ for every closed subset U of Y).

(3) If X and Y are both topological spaces and if $\mathcal{A}$ and $\mathcal{B}$ are their respective Borel σ-algebras, then every continuous map $f \colon X \to Y$ is measurable.

Given any subset A of X, recall that the *characteristic function* χ_A of A is defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Then, as illustrated in Figure 5.1, it is easy to show that for any subset A of X, the function $\chi_A \colon X \to \mathbb{R}$ (where $\mathbb{R}$ is equipped with its σ-algebra of Borel sets) is measurable iff $A \in \mathcal{A}$, that is, A is measurable.

In the theory of integration, all maps of interest will be measurable maps[1] $f \colon X \to F$ where $(X, \mathcal{A})$ is a measurable space, and $(F, \mathcal{B})$ is a measurable space such that either $F = \mathbb{R}$, or $F = \mathbb{C}$, or more generally F is a Banach space (a complete normed vector space over $\mathbb{R}$ or $\mathbb{C}$), and $\mathcal{B}$ is the Borel σ-algebra of open subsets of F. In this case various operations can be performed on functions $f \colon X \to F$.

Assume that F is a normed vector space over the field K, where $K = \mathbb{R}$ or $K = \mathbb{C}$, and that $f \colon X \to F$ is any function, not necessarily measurable.

1. Given any function $f \colon X \to F$, for any $\lambda \in K$, let $\lambda f \colon X \to F$ be the function given by
$$(\lambda f)(x) = \lambda f(x), \quad x \in X.$$

[1] Actually, not quite in the most general case, but they will be equal to a measurable map a.e.

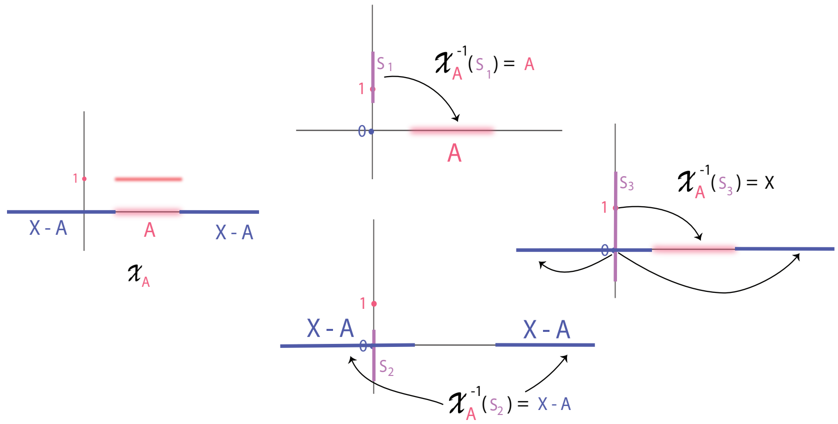


Figure 5.1: The left figure illustrates $\chi_A \colon X \to \mathbb{R}$. If $S_1 \subset \mathbb{R}$ contains 1 but not 0, $\chi_A^{-1}(S_1) = A$. If $S_2 \subset \mathbb{R}$ contains 0 but not 1, $\chi_A^{-1}(S_1) = X - A$. Finally, if $S_3 \subset \mathbb{R}$ contains both 0 and 1, $\chi_A^{-1}(S_1) = A \cup (X - A) = X$.

2. Given any function $f \colon X \to F$, let $\|f\| \colon X \to \mathbb{R}_+$ be the function given by

$$\|f\|(x) = \|f(x)\|, \quad x \in X;$$

see Figure 5.2.

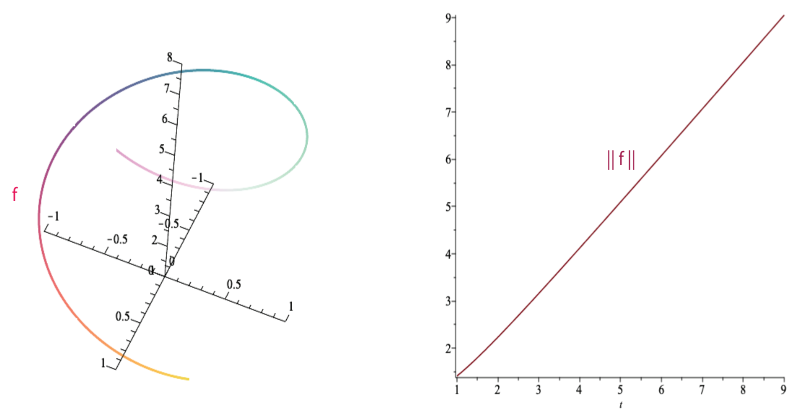


Figure 5.2: Let $f \colon \mathbb{R} \to \mathbb{R}^3$ be $f(t) = (\sin t, \cos t, t)$, the graph of which is the space curve in the left figure. If we use the Euclidean norm on $\mathbb{R}^3$, $\|f\|(t) = \sqrt{t^2 + 1}$, the graph of which is shown in the right figure.

Beware that $\|f\|$ is *not* the norm of the function f, where $\| \ \|$ is the norm on some function space consisting of functions from X to F. Instead, $\|f\|$ is the *function* defined pointwise as $\|f(x)\|$ for every $x \in X$, where $\|f(x)\|$ is the norm of $f(x)$ in F. This

notation is somewhat confusing but appears to be standard. Later on, we will equip our space of functions from X to F with a norm, but it will be denoted $\| \|_1$, or more generally $\| \|_p$, so there will be no risk of confusion.

3. For any two functions $f \colon X \to \mathbb{R}$ and $g \colon X \to \mathbb{R}$, let $\sup(f, g)$ and $\inf(f, g)$ be the functions given by

$$\begin{aligned} \sup(f,g)(x) &= \max(f(x), g(x)), \quad x \in X, \\ \inf(f,g)(x) &= \min(f(x), g(x)), \quad x \in X; \end{aligned}$$

see Figure 5.3.

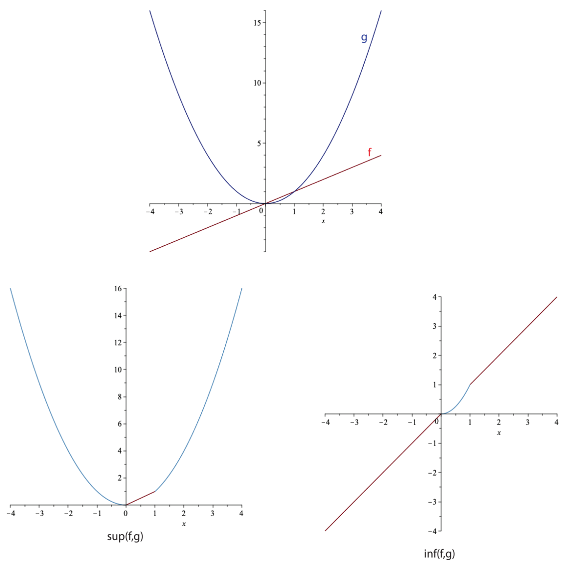


Figure 5.3: Let $f \colon \mathbb{R} \to \mathbb{R}$ be $f(x) = x$, and let $g \colon \mathbb{R} \to \mathbb{R}$ be $g(x) = x^2$. The graph of $\sup(f, g)$ is the lower left figure, while the graph of $\inf(f, g)$ is the lower right figure.

4. For any two functions $f \colon X \to \mathbb{R}$, let f^+ and f^- be the functions given by

$$f^+(x) = \begin{cases} 0 & \text{if } f(x) \leq 0 \\ f(x) & \text{if } f(x) > 0, \end{cases}$$

$$f^-(x) = \begin{cases} 0 & \text{if } f(x) \geq 0 \\ -f(x) & \text{if } f(x) < 0. \end{cases}$$

We also define $|f|$ as $|f| = f^+ + f^- = \sup(f, -f)$. Observe that $f = f^+ - f^-$. See Figures 5.4 through 5.6.

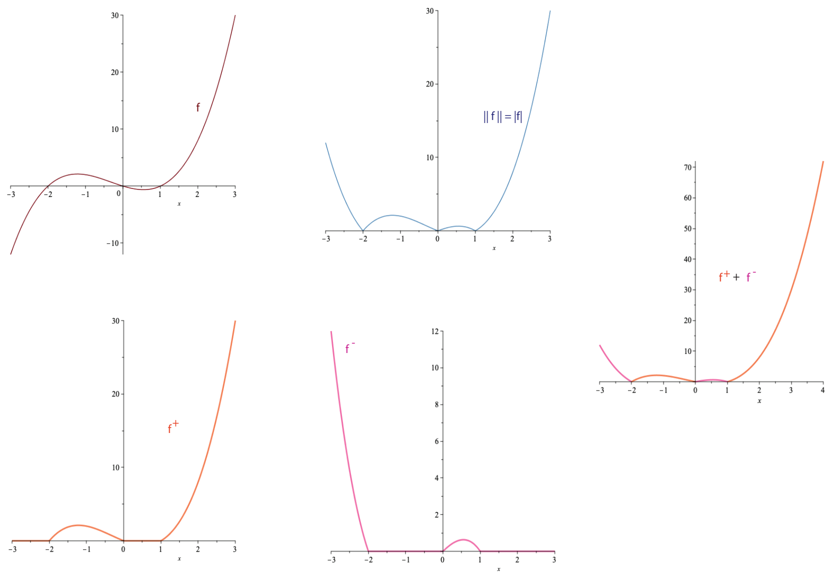


Figure 5.4: Let $f\colon \mathbb{R} \to \mathbb{R}$ be $f(x) = x(x-1)(x+2)$. The lower figures illustrate f^+ and f^-, while the right figures illustrate the identity $|f| = f^+ + f^-$.

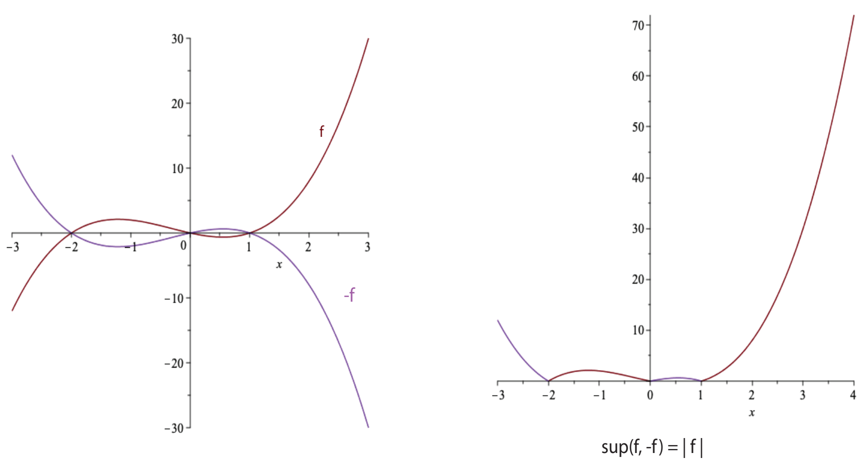


Figure 5.5: Let $f\colon \mathbb{R} \to \mathbb{R}$ be $f(x) = x(x-1)(x+2)$. The right figure illustrates the identity $|f| = \sup(f, -f)$.

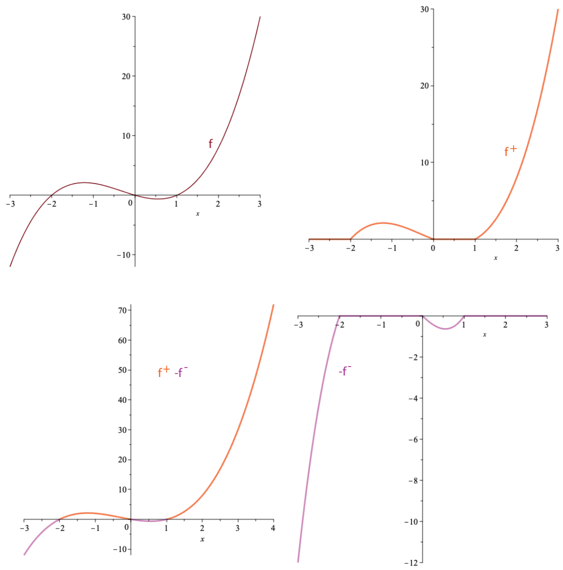


Figure 5.6: Let $f\colon \mathbb{R} \to \mathbb{R}$ be $f(x) = x(x-1)(x+2)$. The lower left figure, when combined with the two right figures, illustrates the identity $f = f^+ - f^-$.

5. For any two functions $f\colon X \to F$ and $g\colon X \to F$, let $f + g\colon X \to F$ be the function given by
$$(f+g)(x) = f(x) + g(x), \quad x \in X.$$

6. For any two functions $f\colon X \to K$ and $g\colon X \to K$, where $K = \mathbb{R}$ of $K = \mathbb{C}$, let $fg\colon X \to K$ be the function given by
$$(fg)(x) = f(x)g(x), \quad x \in X.$$

Definition 5.2. Let $(X, \mathcal{A})$ be a measurable space, and let $(F, \mathcal{B})$ be a measurable space such that $F = \mathbb{R}$, or $F = \mathbb{C}$, or more generally F is metric space (not necessarily complete), and $\mathcal{B}$ is the Borel σ-algebra of open subsets of F. The set of measurable maps $f\colon X \to F$ is denoted by $\mathcal{M}(X, \mathcal{A}, F)$.

The following technical result is needed.

Proposition 5.4. *Let $(X, \mathcal{A})$ be any measurable space, and let $(F_1, \mathcal{B}_1)$, $(F_2, \mathcal{B}_2)$, and $(G, \mathcal{G})$ be three measurable spaces, where F_1, F_2, G are topological spaces, and $\mathcal{B}_1, \mathcal{B}_2, \mathcal{G}$ are their respective Borel σ-algebras. Let $h\colon F_1 \times F_2 \to G$ be a continuous map, and let $f_1\colon X \to F_1$ and $f_2\colon X \to F_2$ be two measurable maps. If the subspace topologies on $f_1(X) \subseteq F_1$ and $f_2(X) \subseteq F_2$ are second-countable (which means that they have a countable basis of open subsets), then $h \circ (f_1, f_2)\colon X \to G$ is measurable.*

Recall that a topological space E is separable if it contains a countable subset which is dense in E (see Definition A.42). If E is a metric space, then by Proposition A.46, the space E is separable if and only if it is second-countable. Using Proposition 5.4 we obtain the following important result stating various closure properties of $\mathcal{M}(X, \mathcal{A}, F)$.

Proposition 5.5. *Let $(X, \mathcal{A})$ be any measurable space, and assume that F is a normed vector space over the field K, where $K = \mathbb{R}$ or $K = \mathbb{C}$. The following properties hold:*

1. *For any $f \in \mathcal{M}(X, \mathcal{A}, F)$ and any $\lambda \in K$, we have $\lambda f \in \mathcal{M}(X, \mathcal{A}, F)$.*
2. *For any $f \in \mathcal{M}(X, \mathcal{A}, F)$, we have $\|f\| \in \mathcal{M}(X, \mathcal{A}, \mathbb{R})$.*
3. *For any $f \in \mathcal{M}(X, \mathcal{A}, \mathbb{R})$ and any $g \in \mathcal{M}(X, \mathcal{A}, \mathbb{R})$, we have $\sup(f, g), \inf(f, g), f^+, f^-$, $|f| \in \mathcal{M}(X, \mathcal{A}, \mathbb{R})$.*
4. *For any $f \in \mathcal{M}(X, \mathcal{A}, F)$ and any $g \in \mathcal{M}(X, \mathcal{A}, F)$, if $f(X)$ and $g(X)$ are separable subsets of F, then $f + g \in \mathcal{M}(X, \mathcal{A}, F)$. In particular, if F is separable, then $\mathcal{M}(X, \mathcal{A}, F)$ is a vector space over K.*
5. *For any $f \in \mathcal{M}(X, \mathcal{A}, K)$ and any $g \in \mathcal{M}(X, \mathcal{A}, K)$, we have $fg \in \mathcal{M}(X, \mathcal{A}, K)$. This implies that $\mathcal{M}(X, \mathcal{A}, K)$ is actually a K-algebra.*

One will observe that in (4), if F is infinite-dimensional, the sum of two measurable maps may *not* be measurable. This is the first technical difficulty of the general theory of integration (with values in an infinite-dimensional vector space). As we will see, a second technical difficulty has to do with the approximation of a measurable map by step functions. Fortunately these technical difficulties can be overcome in a simple way.

The following important result shows that measurable maps behave better than continuous maps in terms of simple (pointwise) convergence.

Theorem 5.6. *Let $(X, \mathcal{A})$ and $(F, \mathcal{B})$ be two measurable spaces, where F is a metric space and $\mathcal{B}$ is the Borel σ-algebra on F. If $(f_n)_{n \geq 1}$ is a sequence of measurable maps $f_n \in \mathcal{M}(X, \mathcal{A}, F)$ which converges pointwise to a function $f \colon X \to F$, then $f \in \mathcal{M}(X, \mathcal{A}, F)$; that is, f is measurable.*

A proof of Theorem 5.6 can be found in Lang [62] (Chapter VI, Section 1, Property **M7**).

Our next goal is to generalize the notion of step function given in Definition 2.21 to the framework of measure spaces.

5.2 Step Maps on a Measurable Space

Let $(X, \mathcal{A})$ be a measurable space. The generalization of the notion of step map is obtained by replacing the intervals (a_i, a_{i+1}) by *arbitrary measurable sets*.

Definition 5.3. Let $(X, \mathcal{A})$ be a measurable space, and let F be any set. A function $f\colon X \to F$ is a *step map* (with respect to $\mathcal{A}$) if there is a finite partition $(A_1, \ldots, A_n)$ of X by pairwise disjoint nonempty subsets $A_i \in \mathcal{A}$ such that $X = \bigcup_{i=1}^n A_i$, and such that the restriction of f to each A_i is a constant function with some value $y_i \in F$. The partition $(A_1, \ldots, A_n)$ is said to be *adapted to* f; see Figure 5.7. The set of all step maps is denoted by $\mathcal{S}tep(X, \mathcal{A}, F)$.

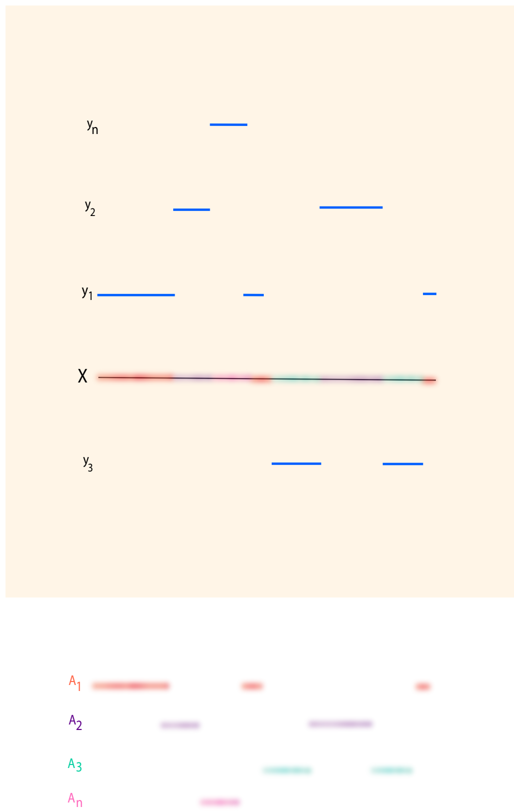


Figure 5.7: Let $(X, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $F = \mathbb{R}$. A step map is shown in blue with values $\{y_i\}_{i=1}^n$. The partition $(A_1, \ldots, A_n)$ adapted to f is shown underneath the peach box.

Observe that every constant function is a step map, and that $f(X)$ is a finite subset of F. At this stage, no measure μ is involved, but for the theory of integration, we will have a measure space $(X, \mathcal{A}, \mu)$ and we will need to require each A_i for which $y_i \neq 0$ to have finite measure (this makes sense since in this case F is a vector space).

We gather some useful properties of step maps in the following proposition.

Proposition 5.7. *Let $(X, \mathcal{A})$ be a measurable space, and let F be any set.*

1. *For any σ-algebra $\mathcal{B}$ on F, every step map $\mathcal{S}tep(X, \mathcal{A}, F)$ is measurable.*

2. *Let F_1, F_2, G be three sets, and let $h\colon F_1 \times F_2 \to G$ be any function. For any $f_1 \in \mathcal{S}tep(X, \mathcal{A}, F_1)$ and any $f_2 \in \mathcal{S}tep(X, \mathcal{A}, F_2)$, we have $h \circ (f_1, f_2) \in \mathcal{S}tep(X, \mathcal{A}, G)$.*

3. *If $K = \mathbb{R}$ or $K = \mathbb{C}$, then $\mathcal{S}tep(X, \mathcal{A}, K)$ is a vector space and a ring under pointwise multiplication of functions. Thus, $\mathcal{S}tep(X, \mathcal{A}, K)$ is an algebra over K.*

4. *If F is a vector space over K (with $K = \mathbb{R}$ or $K = \mathbb{C}$), then $\mathcal{S}tep(X, \mathcal{A}, F)$ is a vector space over K, and a module over $\mathcal{S}tep(X, \mathcal{A}, K)$, which means that if $f \in \mathcal{S}tep(X, \mathcal{A}, F)$ and $g \in \mathcal{S}tep(X, \mathcal{A}, K)$, then $gf \in \mathcal{S}tep(X, \mathcal{A}, F)$.*

5. *If F is a normed vector space, and if $f \in \mathcal{S}tep(X, \mathcal{A}, F)$, then $\|f\| \in \mathcal{S}tep(X, \mathcal{A}, \mathbb{R})$.*

6. *If $f \in \mathcal{S}tep(X, \mathcal{A}, \mathbb{R})$ and $g \in \mathcal{S}tep(X, \mathcal{A}, \mathbb{R})$, then we have $\sup(f, g), \inf(f, g), f^+, f^-$, $|f| \in \mathcal{S}tep(X, \mathcal{A}, \mathbb{R})$.*

Theorem 5.6 and Proposition 5.7 imply the following result.

Proposition 5.8. *Given a metric space F equipped with its σ-algebra of Borel sets, if a function $f\colon X \to F$ is the limit of a sequence $(f_n)_{n\geq 1}$ of step functions $f_n \in \mathcal{S}tep(X, \mathcal{A}, F)$ that converges pointwise, then the function $f\colon X \to F$ must be measurable.*

Unfortunately, in general, a measurable map $f\colon X \to F$ *may not* be the pointwise limit of a sequence of step maps if F has infinite dimension. For one thing, such a limit of steps maps has its image contained in the closure of a countable subset of F. This is the second technical difficulty of the general theory.

To overcome this second difficulty, we need to define a more refined notion of measurable map and of step map. We will do so shortly, but first we observe that if we only need to consider values in a finite-dimensional vector space, then there is no problem.

Proposition 5.9. *Let $(X, \mathcal{A})$ and $(F, \mathcal{B})$ be two measurable spaces, where F is a topological space and $\mathcal{B}$ is its Borel σ-algebra, and let $f\colon X \to F$ be a measurable map.*

1. *If F is either a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, or $F = \overline{\mathbb{R}}_+$, then there is a sequence (f_n) of step maps $f_n \in \mathcal{S}tep(X, \mathcal{A}, F)$ that converges pointwise to f. If $F = \overline{\mathbb{R}}_+$, we may assume that the f_n take finite values.*

2. *If $F = \mathbb{R}$ or $F = \overline{\mathbb{R}}_+$, and if $f \geq 0$, then we may assume that $f_n \geq 0$ and $f_n \leq f_{n+1}$ for all $n \geq 1$.*

A proof of Proposition 5.9 can be found in Lang [62] (Chapter VI, Section 1, Properties **M8** and **M9**).

5.3 μ-Step Maps

We explained in the previous sections that in general, the space $\mathcal{M}(X, \mathcal{A}, F)$ of measurable maps from X to F is not a vector space, and that a measurable map $f\colon X \to F$ may not be the pointwise limit of a sequence of step maps. This suggests modifying the notion of measurable map and the notion of step map to recover these properties. The second property is crucial in extending the notion of integral to more general functions.

So far, the space X was only a measurable space, but no measure was involved. The new ingredient is to define suitable notions of step maps and measurable maps relative to a *measure space* $(X, \mathcal{A}, \mu)$, where the measure μ plays a role.

The main trick is to relax the notion of pointwise convergence to pointwise convergence almost everywhere, and more generally, to consider that two functions are equivalent if they are equal almost everywhere (they differ on a null set). The plan is the following:

1. Define the space $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ of *μ-step maps*.
2. Define the space $\mathcal{M}_\mu(X, \mathcal{A}, F)$ of *μ-measurable maps*, where a μ-measurable map is the limit of a sequence (f_n) of μ-step maps $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ converging pointwise almost everywhere.
3. Prove that if F is a vector space, then $\mathcal{M}_\mu(X, \mathcal{A}, F)$ is a vector space.

Our presentation of the method that we just sketched follows Marle [69] and Lang [62] very closely. It is a generalization (with some simplifications) to functions with values in a Banach space of the approach followed by Halmos [44]. The results that we state without proof are proved either in Marle [69] or in Lang [62].

Definition 5.4. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let F be any vector space (over $\mathbb{R}$ or $\mathbb{C}$). A function $f\colon X \to F$ is a *μ-step map* if it is a step map, and if $\{x \in X \mid f(x) \neq 0\} \in \mathcal{A}$ and has finite measure; see Figure 5.8. The set of μ-step maps is denoted by $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$.

For technical reasons, it is useful to have the following equivalent characterization of a μ-step map.

Proposition 5.10. *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let F be any vector space (over $\mathbb{R}$ or $\mathbb{C}$). A function $f\colon X \to F$ is a μ-step map iff there is a nonempty subset $A \in \mathcal{A}$ of finite measure such that f vanishes outside A, that is, $f(x) = 0$ for all $x \in X - A$, and if there is a finite partition $(A_1, \ldots, A_n)$ of A of subsets $A_i \in \mathcal{A}$ (nonempty pairwise disjoint subsets) such that the restriction of f to each A_i has a constant value y_i.*

Proof. Let f be a μ-step map, that is, a step map with respect to a partition $(A_1, \ldots, A_n)$ of X such that $\{x \in X \mid f(x) \neq 0\} \in \mathcal{A}$ and has finite measure. Then any A_i on which f has value $y_i \neq 0$ must have finite measure. If $f = 0$ on X, then pick A to be any A_i and the

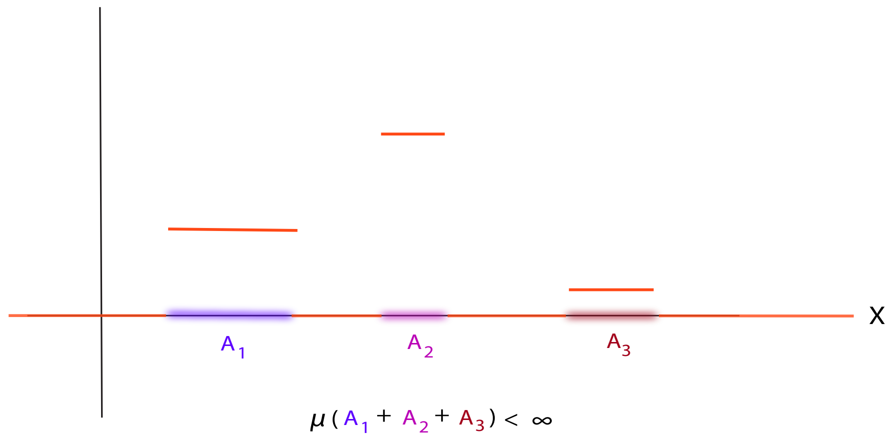


Figure 5.8: Let $(X, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $F = \mathbb{R}$. A μ-step map is shown in red where $A_1 \cup A_2 \cup A_3 = \{x \in X \mid f(x) \neq 0\} \in \mathcal{A}$.

partition to be (A_i). Otherwise, let $J = \{j \in \{1, \ldots, n\} \mid f \neq 0 \,\text{on}\, A_j\}$, and let $A = \bigcup_{j \in J} A_j$. Then, $(A_j)_{j \in J}$ is a partition of A with $A_j \in \mathcal{A}$, where A is a nonempty set of finite measure, and f vanishes on $X - A$; see Figure 5.9.

Conversely, since A has finite measure and since the A_i belongs to $\mathcal{A}$, each A_i has finite measure, so $\{x \in X \mid f(x) \neq 0\} \in \mathcal{A}$ is a set of finite measure. If $A = X$, then we already have a step map (as defined in Definition 5.3). Otherwise, $X - A \in \mathcal{A}$ and f vanishes on $X - A$, so $(A_1, \ldots, A_n, X - A)$ is partition of X, and f is a step map with respect to this partition; see Figure 5.10. $\square$

The condition that a μ-step map must vanish outside of a measurable set of finite measure is the measure-theoretic analog of the topological notion of compact support.

Proposition 5.10 suggests the following equivalent definition of a *μ-step map*.

Definition 5.5. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let F be any vector space (over $\mathbb{R}$ or $\mathbb{C}$). A function $f \colon X \to F$ is a *μ-step map* if there is a nonempty subset $A \in \mathcal{A}$ of finite measure such that f vanishes outside A, that is, $f(x) = 0$ for all $x \in X - A$, and if there is a finite partition $(A_1, \ldots, A_n)$ of A consisting of nonempty pairwise disjoint subsets in $\mathcal{A}$, such that the restriction of f to each A_i has a constant value y_i (possibly zero). The partition $(A_1, \ldots, A_n)$ of A is said to be *adapted to f*.

Technically, Definition 5.5 appears to be more convenient. Observe that a μ-step map can be expressed as a (necessarily finite) linear combination

$$f = \sum_{i=1}^{n} y_i \chi_{A_i},$$

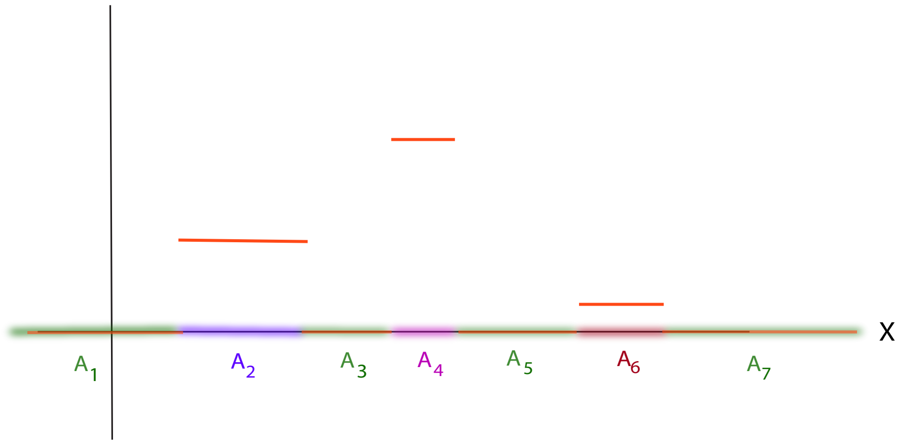


Figure 5.9: Let $(X, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $F = \mathbb{R}$. A step map is shown in red with adapted partition $\mathbb{R} = \bigcup_{i=1}^{7} A_i$. To interpret this step map as a μ-step map, let $A = A_2 \cup A_4 \cup A_6$, where $\mu(A) < \infty$, $A = \{x \in X \mid f(x) \neq 0\} \in \mathcal{A}$. Then $f(x) = 0$ on $X - A$ where $X - A = A_1 \cup A_3 \cup A_5 \cup A_7$.

for some $y_i \in F$ and for some nonempty pairwise disjoint measurable sets $A_i \in \mathcal{A}$ of finite measure, a concise and convenient representation.

Remark: The proof of Proposition 5.10 shows that if a μ-step function f is not identically zero, then we can find a subset A in $\mathcal{A}$ of finite measure, and a partition $(A_1, \ldots, A_n)$ of A of subsets in $\mathcal{A}$ such that the value of f on each A_i is nonzero, and f is zero outside of A. However, it turns out to be more convenient for certain proofs to allow f to be zero on some of the A_i, and this is why we allow this possibility in Definition 5.5.

Example 5.1. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$, given by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \text{ or } x > 1 \\ 1 & \text{if } x \in [0, 1/2] - \mathbb{Q} \\ 0 & \text{if } x \in [0, 1/2] \cap \mathbb{Q} \\ 2 & \text{if } x \in [1/2, 1] - \mathbb{Q} \\ 0 & \text{if } x \in [1/2, 1] \cap \mathbb{Q}; \end{cases}$$

see Figure 5.11. If we let $A_1 = [0, 1/2] - \mathbb{Q}$, $A_2 = [0, 1/2] \cap \mathbb{Q}$, $A_3 = [1/2, 1] - \mathbb{Q}$, $A_4 = [1/2, 1] \cap \mathbb{Q}$, and $A = [0, 1]$, with the Lebesgue measure μ_L on $\mathbb{R}$, then A_1, A_2, A_3, A_4 are Lebesgue measurable, $\mu(A_1) = 1/2$, $\mu(A_2) = 0$, $\mu(A_3) = 1/2$, $\mu(A_3) = 0$, (A_1, A_2, A_3, A_4) is a partition of A, a set of measure 1. Thus f is a μ_L-step function.

This example shows that a μ-step function can be very complicated, unlike the step functions of Definition 2.21.

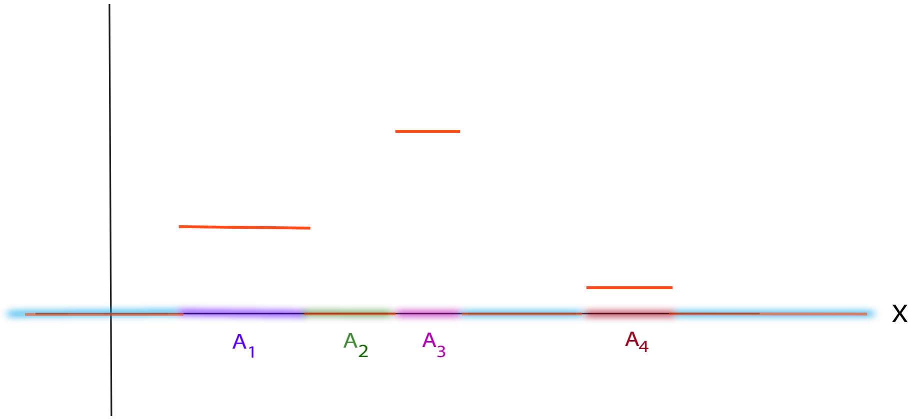


Figure 5.10: Let $(X, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $F = \mathbb{R}$. A μ-step map is shown in red with $A = A_1 \cup A_2 \cup A_3 \cup A_4$. The turquoise set is $X - A$ and $(A_1, A_2, A_3, A_4, X - A)$ forms an adapted partition for the corresponding step map.

Proposition 5.11. *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let F be any vector space*

1. *Let F_1, F_2, G be three Banach spaces over $\mathbb{R}$ or $\mathbb{C}$, and let $h\colon F_1 \times F_2 \to G$ be any function. If h satisfies $h(0,0) = 0$, then for any $f_1 \in \mathcal{S}tep_\mu(X, \mathcal{A}, F_1)$ and any $f_2 \in \mathcal{S}tep_\mu(X, \mathcal{A}, F_2)$, we have $h \circ (f_1, f_2) \in \mathcal{S}tep_\mu(X, \mathcal{A}, G)$.*

2. *If $K = \mathbb{R}$ or $K = \mathbb{C}$, then $\mathcal{S}tep_\mu(X, \mathcal{A}, K)$ is a subspace of $\mathcal{S}tep(X, \mathcal{A}, K)$, and for any $g \in \mathcal{S}tep(X, \mathcal{A}, K)$ and any $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, K)$ we have $gf \in \mathcal{S}tep_\mu(X, \mathcal{A}, K)$. Thus $\mathcal{S}tep_\mu(X, \mathcal{A}, K)$ is an ideal in $\mathcal{S}tep(X, \mathcal{A}, K)$.*

3. *If F is a vector space over K (with $K = \mathbb{R}$ or $K = \mathbb{C}$), then $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ is a subspace of $\mathcal{S}tep(X, \mathcal{A}, F)$ and a module over $\mathcal{S}tep(X, \mathcal{A}, K)$, which means that if $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ and $g \in \mathcal{S}tep(X, \mathcal{A}, K)$, then $gf \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$.*

4. *If F is a normed vector space, and if $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, then $\|f\| \in \mathcal{S}tep_\mu(X, \mathcal{A}, \mathbb{R})$. In fact, if $f = \sum_{i=1}^n y_i \chi_{A_i}$, then $\|f\| = \sum_{i=1}^n \|y_i\| \chi_{A_i}$.*

5. *If $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, \mathbb{R})$ and $g \in \mathcal{S}tep_\mu(X, \mathcal{A}, \mathbb{R})$, then $\sup(f, g), \inf(f, g), f^+, f^-, |f| \in \mathcal{S}tep_\mu(X, \mathcal{A}, \mathbb{R})$.*

We now come to the crucial notion of μ-measurable map.

5.4 μ-Measurable Maps

Definition 5.6. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let F be any vector space (over $\mathbb{R}$ or $\mathbb{C}$). A function $f\colon X \to F$ is a *μ-measurable* if there is a sequence $(f_n)_{n \geq 1}$ of μ-step

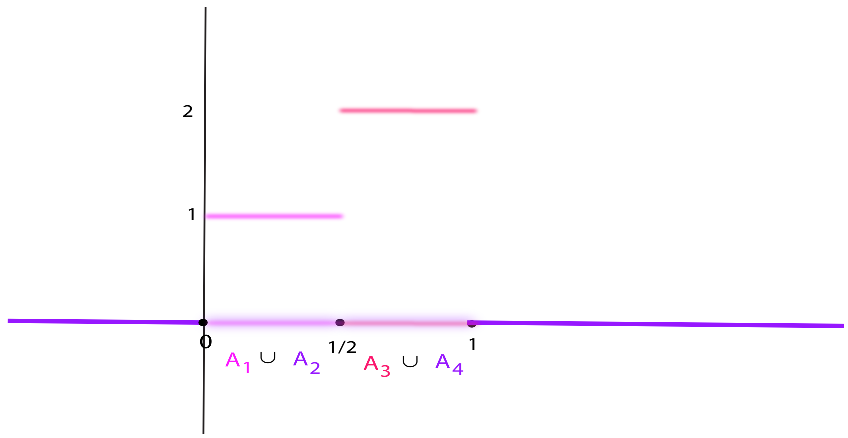


Figure 5.11: The μ-step map of Example 5.1.

maps $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ which converges pointwise to f almost everywhere. See Figures 5.12 and 5.13. Recall that this means that there is a null set $Z \subseteq X$ such that for every $x \in X - Z$, the sequence $(f_n(x))$ converges to $f(x)$. The set of μ-measurable maps is denoted by $\mathcal{M}_\mu(X, \mathcal{A}, F)$.

Observe that a μ-measurable map is not necessarily measurable, so $\mathcal{M}_\mu(X, \mathcal{A}, F)$ is *not* a subspace of $\mathcal{M}(X, \mathcal{A}, F)$. However, we will see shortly that a μ-measurable map is equal to a measurable map almost everywhere, and this is good enough to construct the Lebesgue integral. The following proposition can be proved using Proposition 5.11 by passing to the limit (carefully).

Proposition 5.12. *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let F be any vector space*

1. *Let F_1, F_2, G be three Banach spaces over $\mathbb{R}$ or $\mathbb{C}$, and let $h\colon F_1 \times F_2 \to G$ be any function. If h satisfies $h(0,0) = 0$, then for any $f_1 \in \mathcal{M}_\mu(X, \mathcal{A}, F_1)$ and any $f_2 \in \mathcal{M}_\mu(X, \mathcal{A}, F_2)$, we have $h \circ (f_1, f_2) \in \mathcal{M}_\mu(X, \mathcal{A}, G)$.*

2. *If $K = \mathbb{R}$ or $K = \mathbb{C}$, then $\mathcal{M}_\mu(X, \mathcal{A}, K)$ is a vector space, and for all $f, g \in \mathcal{M}_\mu(X, \mathcal{A}, K)$ we have $fg \in \mathcal{M}_\mu(X, \mathcal{A}, K)$. Thus $\mathcal{M}_\mu(X, \mathcal{A}, K)$ is an algebra over K. For any $g \in \mathcal{M}(X, \mathcal{A}, K)$ and any $f \in \mathcal{M}_\mu(X, \mathcal{A}, K)$ we have $gf \in \mathcal{M}_\mu(X, \mathcal{A}, K)$.*

3. *If F is a vector space over K (with $K = \mathbb{R}$ or $K = \mathbb{C}$), then $\mathcal{M}_\mu(X, \mathcal{A}, F)$ is a vector space over K and a module over $\mathcal{M}(X, \mathcal{A}, K)$, which means that if $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ and $g \in \mathcal{M}(X, \mathcal{A}, K)$, then $gf \in \mathcal{M}_\mu(X, \mathcal{A}, F)$. The space $\mathcal{M}_\mu(X, \mathcal{A}, F)$ is also a module over $\mathcal{M}_\mu(X, \mathcal{A}, K)$.*

4. *If F is a normed vector space, and if $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$, then $\|f\| \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})$.*

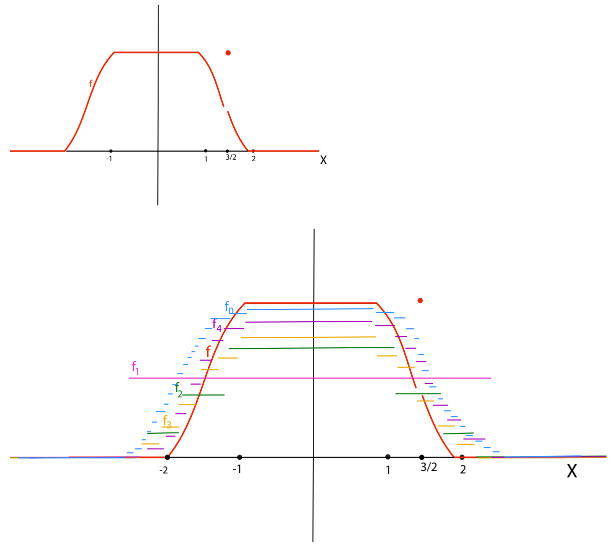


Figure 5.12: Let $X = F = \mathbb{R}$. Assume μ is the Lebesgue measure on $\mathbb{R}$. The graph of the μ-measurable f is shown in the upper left corner. The middle figure illustrates the sequence $(f_n)_{n\geq 1}$ of μ-step maps $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ which converges pointwise to f almost everywhere.

5. *If $f \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})$ and $g \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})$, then we have* $\sup(f, g), \inf(f, g), f^+, f^-, |f| \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})$.

The following result gives a characterization of a μ-measurable map which shows that a μ-measurable map is equal to a measurable map almost everywhere, and that there are strong countability restrictions on its domain and its range.

Proposition 5.13. *Let $(X, \mathcal{A}, \mu)$ be a measure space, and let F be any Banach space. A function $f\colon X \to F$ is μ-mesurable iff there is a null set Z such that the following three conditions hold:*

(1) There is a measurable map $g \in \mathcal{M}(X, \mathcal{A}, F)$ such that f and g are equal on $X - Z$.

(2) The function f vanishes outside of a measurable σ-finite subset of X (recall Definition 4.10).

(3) The image $f(X - Z)$ is separable in F, which means that $f(X - Z)$ contains a countable dense subset.

In particular, if μ is σ-finite and if F is separable, then $f\colon X \to F$ is μ-measurable iff f is measurable almost everywhere (there is a null set Z such that f agrees with a measurable map on $X - Z$).

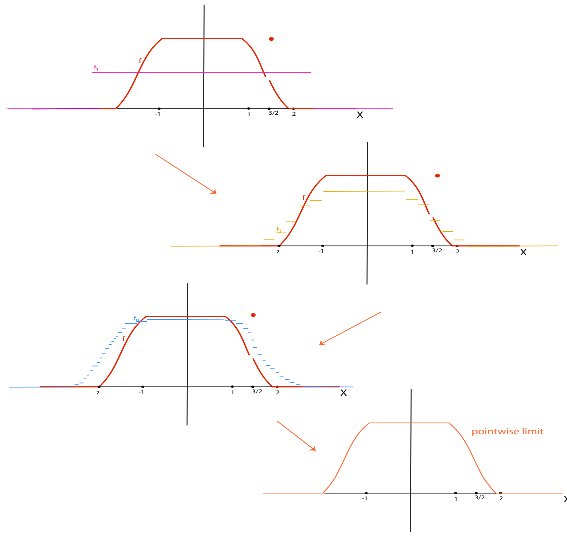


Figure 5.13: A more detailed look at the pointwise convergence of converges the $(f_n)_{n\geq 1}$. The pointwise limit only differs from f on a set of measure zero, namely $\{3/2\}$.

A proof of Proposition 5.13 can be found in Lang [62] (Chapter VI, Section 1, Property **M11**). Again, Condition (2) is a measure-theoretic analog of the notion of compact support.

The version of Theorem 5.6 for μ-measurable maps is stated below.

Theorem 5.14. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space, where F is a metric space and $\mathcal{B}$ is the Borel σ-algebra on F. If $(f_n)_{n\geq 1}$ is a sequence of μ-measurable maps $f_n \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ which converges pointwise to a function $f\colon X \to F$, then $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$; that is, f is μ-measurable.*

A proof of Theorem 5.6 can be found in Lang [62] (Chapter VI, Section 1, Property **M12**).

We are now ready construct a very general version of the integral. The original construction was first proposed by Lebesgue, but the more general version presented here applying to functions with values in a Banach space is due to Bochner and Dunford.

5.5 The Integral of μ-Step Maps

Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space consisting of a Banach space F and its Borel σ-algebra $\mathcal{B}$. There is an "obvious" definition of the integral of a μ-step

map $f = \sum_{i=1}^n y_i \chi_{A_i}$ (where $y_i \in F$), namely

$$I(f) = \int f d\mu = \sum_{i=1}^n \mu(A_i) y_i.$$

Since by definition the A_i belong to $\mathcal{A}$ and have finite measure, the linear combination $\sum_{i=1}^n \mu(A_i) y_i$ is a well-defined vector in F. The only problem is that $I(f)$ seems to depend on the subset A (and its partition) chosen to express f, but it is easy to show that $I(f)$ is independent of the representation of f. Then it is easy to show that $I \colon \mathcal{S}tep_\mu(X, \mathcal{A}, F) \to F$ is a linear map. Furthermore, by Proposition 5.12, we have $\|f\| \in \mathcal{S}tep_\mu(X, \mathcal{A}, \mathbb{R})$, so we can define

$$N_1(f) = \int \|f\|\, d\mu,$$

and we have

$$\left\| \int f d\mu \right\| \leq \int \|f\|\, d\mu = N_1(f).$$

It turns out that N_1 satisfies all the axioms of a norm, except that $N_1(f) = 0$ does not necessarily imply that $f = 0$. We say that N_1 is a *semi-norm*, see Definition A.3. Fortunately, for any $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, we have $N_1(f) = 0$ iff $f = 0$, except on a subset of measure zero.

We can define the notion of N_1-*Cauchy sequence* of a sequence (f_n) of functions $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ as follows: for all $\epsilon > 0$, there is some $N > 0$, such that for all $m, n \geq N$, we have $N_1(f_m - f_n) < \epsilon$. We can also define the notion of N_1-*convergence* of a sequence (f_n) of functions $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ to a limit $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ as follows: for all $\epsilon > 0$, there is some $N > 0$, such that for all $n \geq N$, we have $N_1(f - f_n) < \epsilon$. A convergent N_1-sequence does not necessarily have a unique limit, but we will see that any two limits are equal a.e.

The problem is that an N_1-Cauchy sequence *may not* have a limit in $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$. Thus we are led to completing $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ with respect to the semi-norm N_1. This can be done and we obtain a vector space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ which is a subspace of $\mathcal{M}_\mu(X, \mathcal{A}, F)$. The integral map I and the semi-norm N_1 can be extended to $\mathcal{L}_\mu(X, \mathcal{A}, F)$ as a semi-norm denoted $\| \,\|_1$, the space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is Cauchy-complete with respect to the semi-norm $\| \,\|_1$, and $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ is dense in $\mathcal{L}_\mu(X, \mathcal{A}, F)$ with respect to the semi-norm $\| \,\|_1$. This situation is schematically illustrated in Figure 5.14.

It also turns out that the subspace $\mathcal{SN}$ of $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ consisting of all functions f such that $N_1(f) = 0$ is the set of functions in $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ that are equal to 0 a.e. Similarly, the subspace $\mathcal{N}$ of $\mathcal{L}_\mu(X, \mathcal{A}, F)$ consisting of all functions f such that $\|f\|_1 = 0$ is the set of functions in $\mathcal{L}_\mu(X, \mathcal{A}, F)$ that are equal to 0 a.e. Thus, we can form the quotients spaces $\mathrm{Step}_\mu(X, \mathcal{A}, F) = \mathcal{S}tep_\mu(X, \mathcal{A}, F)/\mathcal{SN}$ and $\mathrm{L}_\mu(X, \mathcal{A}, F) = \mathcal{L}_\mu(X, \mathcal{A}, F)/\mathcal{N}$. In $\mathrm{Step}_\mu(X, \mathcal{A}, F)$ and in $\mathrm{L}_\mu(X, \mathcal{A}, F)$ the semi-norm $\| \,\|_1$ is really a norm, and $\mathrm{L}_\mu(X, \mathcal{A}, F)$ is the completion of $\mathrm{Step}_\mu(X, \mathcal{A}, F)$.

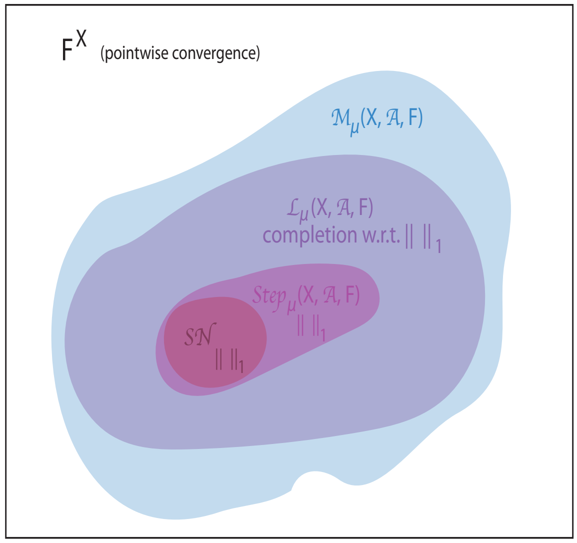


Figure 5.14: Let $(X, \mathcal{A}, \mu)$ be a measure space and $(F, \mathcal{B})$ be a Banach space with the Borel σ-algbera. The completion of $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ with respect to the semi-norm $\| \|_1$ is $\mathcal{L}_\mu(X, \mathcal{A}, F)$.

Theoretically, we could define $\mathrm{L}_\mu(X, \mathcal{A}, F)$ directly as the Cauchy completion (see Theorem A.62 and Theorem A.72) of $\mathrm{Step}_\mu(X, \mathcal{A}, F)$, but we obtain equivalence classes of Cauchy sequences of equivalence classes of functions in $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$, which are not easily interpretable as functions. The same space $\mathrm{L}_\mu(X, \mathcal{A}, F)$ is obtained, see the diagram below.

$$\begin{tikzcd}
\mathcal{S}tep_\mu(X, \mathcal{A}, F) \arrow[r, "completion"] \arrow[d, "quotient"'] & \mathcal{L}_\mu(X, \mathcal{A}, F) \arrow[d, "quotient"] \\
\mathrm{Step}_\mu(X, \mathcal{A}, F) = \mathcal{S}tep_\mu(X, \mathcal{A}, F)/\mathcal{SN} \arrow[r, "completion"'] & \mathrm{L}_\mu(X, \mathcal{A}, F) = \mathcal{L}_\mu(X, \mathcal{A}, F)/\mathcal{N}.
\end{tikzcd}$$

The construction that we alluded to, although involving some extra work, yields a very clear description of these equivalence classes in terms of functions (in $\mathcal{L}_\mu(X, \mathcal{A}, F)$). The completeness of $\mathrm{L}_\mu(X, \mathcal{A}, F)$ (under the $\| \|_1$-norm) is also immediately obtained.

As in the previous section the results that we state without proof are proved either in Marle [69] or in Lang [62].

We now return to the definition of the integral of a μ-step maps.

Proposition 5.15. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space, with F a Banach space and $\mathcal{B}$ its Borel σ-algebra. For any μ-step map $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, for any two partitions $(A_1, \ldots, A_m)$ and $(B_1, \ldots, B_n)$ adapted to f, so that $f = \sum_{i=1}^m y_i \chi_{A_i} = \sum_{j=1}^n z_j \chi_{B_j}$, we have*

$$\sum_{i=1}^{m} \mu(A_i) y_i = \sum_{j=1}^{n} \mu(B_j) z_j.$$

Proposition 5.15 justifies the following definition.

Definition 5.7. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space, with F a Banach space and $\mathcal{B}$ its Borel σ-algebra. For any μ-step map $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, the common value

$$\int f d\mu$$

of the expression

$$I(f) = \sum_{i=1}^{n} \mu(A_i) y_i$$

for any partition $(A_1, \ldots, A_n)$ adapted to f is called the *integral of f* (*relative to the measure μ*); see Figure 5.15.[2]

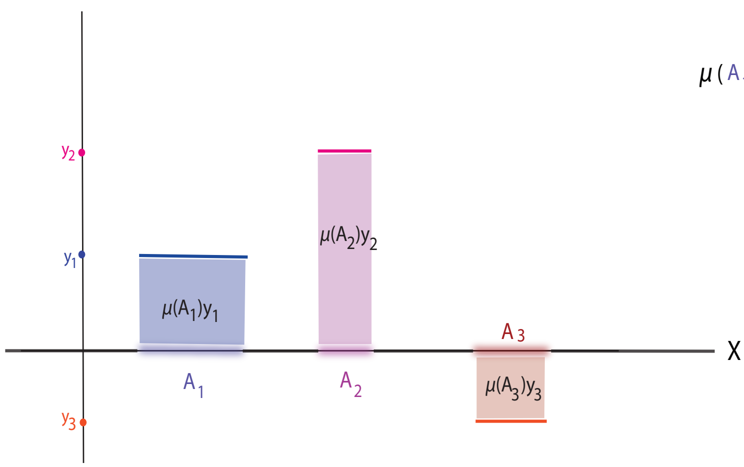


Figure 5.15: Let $(X, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $F = \mathbb{R}$. The integral of the μ-step map f is the signed area of the pastel "boxes".

Recall that if the μ-step map f is expressed as $f = \sum_{i=1}^n y_i \chi_{A_i}$, then the μ-step map $\|f\|$ is expressed as $\|f\| = \sum_{i=1}^n \|y_i\| \chi_{A_i}$.

[2]This integral is usually called the *Lebesgue integral* or *Bochner integral*. A more appropriate name might be the *Bochner–Dunford integral*.

Definition 5.8. We define the *semi-norm* $N_1(f)$ of the μ-step map $f = \sum_{i=1}^n y_i\chi_{A_i}$ as

$$N_1(f) = \int \|f\|\, d\mu = \sum_{i=1}^n \mu(A_i)\, \|y_i\|\,.$$

For any measurable subset $E \in \mathcal{A}$, since $\chi_E f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, we let

$$\int_E f d\mu = \int \chi_E f d\mu.$$

For simplicity of notation, we often write $\int_E f$ instead of $\int_E f d\mu$, and if $E = X$, we write $\int f$ instead of $\int f d\mu$.

We stress that the integral $\int f d\mu$ or $\int_E f d\mu$ is *always finite*; that is, an element of F, but not ∞. This is in contrast with the approach where the integral of a step function may have the value $+\infty$, as in Rudin [79] (Chapter 1). At some later stage, in defining the space $\mathrm{L}^1(X, \mathcal{A}, F)$, it is necessary to require the integral to be finite anyway. We find the approach where the integral is finite in the first place less confusing. It also yields a more explicit definition of $\mathrm{L}^1(X, \mathcal{A}, F)$.

Example 5.2. The special case in which X is a countable set, $\mathcal{A} = 2^X$, μ is the counting measure defined in Example 4.3, and $F = \mathbb{C}$ is of particular interest. Say $X = \mathbb{N}$. A μ-step function is of the form

$$f = \sum_{i=1}^n y_i\chi_{A_i},$$

where A_i must be a finite subset of $\mathbb{N}$, and $y_i \in \mathbb{C}$. By definition of μ, we have $\mu(A_i) = |A_i|$, so

$$\int f d\mu = \sum_{i=1}^n y_i|A_i|.$$

But f is the function with finite support $A = \bigcup_{i=1}^n A_i$, such that $f(j) = y_i$ for all $j \in A_i$, and $f(j) = 0$ for all $j \notin A$, so

$$\int f d\mu = \sum_{j\in A} f(j) = \sum_{j\in\mathbb{N}} f(j),$$

the sum of the (finite) sequence $(f(j))_{j\in\mathbb{N}}$. Similarly, if $X = \mathbb{Z}$, then for any sequence $(f_j)_{j\in\mathbb{Z}}$ with only finitely many nonzero entries,

$$\int f d\mu = \sum_{j\in\mathbb{Z}} f(j).$$

Example 5.3. Recall from Example 4.7 that for any $a \in X$, the Dirac measure δ_a is defined such that for any $A \subseteq X$,

$$\delta_a(A) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{if } a \notin A. \end{cases}$$

Here the σ-algebra is $\mathcal{A} = 2^X$. Then it is easy to check that for any μ-step function

$$f = \sum_{i=1}^{n} y_i \chi_{A_i},$$

we have

$$\int f \, d\delta_a = f(a).$$

So $f(a) = y_i$ iff $a \in A_i$, and $f(a) = 0$ otherwise.

Here are some of the main properties of the integral.

Proposition 5.16. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space, with F a Banach space and $\mathcal{B}$ its Borel σ-algebra. The following properties hold:*

1. *The integral map $\int : \mathcal{S}tep_\mu(X, \mathcal{A}, F) \to F$ is a linear map.*

2. *If A and B are any two disjoint measurable subsets, then*

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

3. *For any map $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, we have $\|f\| \in \mathcal{S}tep_\mu(X, \mathcal{A}, \mathbb{R})$, and*

$$\left\| \int f d\mu \right\| \leq \int \|f\| \, d\mu = N_1(f).$$

   We also have

$$\int \|f\| \, d\mu \leq \mu(\{x \in X \mid f(x) \neq 0\}) \, \|f\|_\infty .$$

4. *For any two maps $f, g \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, if $f = g$ a.e., then $\int f d\mu = \int g d\mu$.*

5. *For any two maps $f, g \in \mathcal{S}tep_\mu(X, \mathcal{A}, \mathbb{R})$, if $f \leq g$ a.e., then $\int f d\mu \leq \int g d\mu$. In particular, if $f \geq 0$ a.e., then $\int f d\mu \geq 0$.*

6. *N_1 is a semi-norm on $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$. Furthermore, for any $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, we have $N_1(f) = 0$ iff $f = 0$, except on a subset of measure zero.*

7. *If F_1 and F_2 are two Banach spaces over $\mathbb{R}$ or $\mathbb{C}$, and if $h\colon F_1 \to F_2$ is a continuous linear map, then for any $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F_1)$, we have $h \circ f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F_2)$, and*

$$\int (h \circ f) d\mu = h\left(\int f d\mu\right)$$

If $F = \mathbb{C}$, the above property holds for any semi-linear map.

Proof. We prove (3) and (6), leaving the other properties as exercises.

(3) If $f = \sum_{i=1}^n y_i \chi_{A_i}$ then $\int f d\mu = \sum_{i=1}^n \mu(A_i) y_i$. We also have $\|f\| = \sum_{i=1}^n \|y_i\| \chi_{A_i}$ and $\int \|f\| \, d\mu = \sum_{i=1}^n \mu(A_i) \|y_i\|$. It follows that

$$\left\|\int f d\mu\right\| = \left\|\sum_{i=1}^n \mu(A_i) y_i\right\| \leq \sum_{i=1}^n \|\mu(A_i) y_i\| = \sum_{i=1}^n \mu(A_i) \|y_i\| = \int \|f\| \, d\mu.$$

We also have

$$\begin{aligned}\int \|f\| \, d\mu = \sum_{i=1}^n \mu(A_i) \|y_i\| &\leq \mu(\{x \in X \mid f(x) \neq 0\}) \max_{1 \leq i \leq n} \|y_i\| \\ &= \mu(\{x \in X \mid f(x) \neq 0\}) \|f\|_\infty .\end{aligned}$$

(6) Since by (1) the integral is linear, we have

$$N_1(\lambda f) = \int \|\lambda f\| \, d\mu = \int |\lambda| \, \|f\| \, d\mu = |\lambda| \int \|f\| \, d\mu = |\lambda| N_1(f).$$

Since $\|(f+g)(x)\| \leq \|f(x)\| + \|g(x)\|$ for all $x \in X$, by (5) we have

$$N_1(f+g) = \int \|f+g\| \, d\mu \leq \int \|f\| \, d\mu + \int \|g\| \, d\mu = N_1(f) + N_1(g).$$

Assume that $N_1(f) = 0$, which means that $\int \|f\| \, d\mu = 0$. Since f is a μ-step function, we can write

$$f = \sum_{i=1}^n y_i \chi_{A_i},$$

for a finite sequence $(A_1, \ldots, A_n)$ of nonempty pairwise disjoint subsets $A_i \in \mathcal{A}$ of finite measure. Since

$$\|f\| = \sum_{i=1}^n \|y_i\| \chi_{A_i},$$

so

$$N_1(f) = \int \|f\| \, d\mu = \sum_{i=1}^n \|y_i\| \mu(A_i) = 0.$$

Since $\|y_i\| \geq 0$ and $\mu(A_i) \geq 0$, the following must hold:

$$\text{if } \mu(A_i) \neq 0, \text{ then } \|y_i\| = 0, \text{ that is } y_i = 0.$$
$$\text{if } y_i \neq 0, \text{ that is } \|y_i\| \neq 0, \text{ then } \mu(A_i) = 0.$$

Consequently

$$\{x \in X \mid f(x) \neq 0\} = \bigcup_{i \in I} A_i, \quad \text{with} \quad I = \{i \mid 1 \leq i \leq n \mid y_i \neq 0\},$$

where $\bigcup_{i \in I} A_i \in \mathcal{A}$ is a set of measure 0, since $i \in I$ implies that $\mu(A_i) = 0$. $\square$

By Proposition 5.16(6), the set

$$\mathcal{SN} = \big\{f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F) \mid N_1(f) = 0\big\} = \big\{f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F) \mid f = 0 \text{ a.e.}\big\}$$

is a subspace of $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$.

Definition 5.9. Let $\mathrm{Step}_\mu(X, \mathcal{A}, F)$ be the quotient space $\mathcal{S}tep_\mu(X, \mathcal{A}, F)/\mathcal{SN}$.

For every equivalence class $\mathbf{f} \in \mathrm{Step}_\mu(X, \mathcal{A}, F)$, we can define

$$\int \mathbf{f} d\mu = \int f d\mu$$

for any function $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ in the equivalence class of $\mathbf{f}$, because if $f = g$ a.e., then $\int f d\mu = \int g d\mu$, so $\int \mathbf{f} d\mu$ does not depend on the representative chosen in the equivalence class $\mathbf{f}$. Similarly, we define $N_1(\mathbf{f})$ by

$$N_1(\mathbf{f}) = N_1(f) = \int \|f\| \, d\mu,$$

for any function $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ in the equivalence class of $\mathbf{f}$. Again if $f = g$ a.e., then $\|f\| = \|g\|$ a.e., so $N_1(f) = N_1(g)$, which means that $N_1(\mathbf{f})$ is well defined. It is immediately verified that N_1 is a semi-norm, and in fact a norm, since $N_1(\mathbf{f}) = 0$ iff $N_1(f) = 0$ for any representative $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ in the equivalence class $\mathbf{f}$ iff $f = 0$ a.e., which means that $\mathbf{f} = 0$. Therefore, $(\mathrm{Step}_\mu(X, \mathcal{A}, F), N_1)$ is a normed vector space. It is easy to see that the inequality

$$\left\| \int \mathbf{f} d\mu \right\| \leq \int \|\mathbf{f}\| \, d\mu = N_1(\mathbf{f})$$

holds, which shows that the map $\int \colon \mathrm{Step}_\mu(X, \mathcal{A}, F) \to F$ is continuous (in fact, uniformly continuous). The space $(\mathrm{Step}_\mu(X, \mathcal{A}, F), N_1)$ is not complete, so we can apply Theorem A.72) to form its completion $\mathrm{L}_\mu(X, \mathcal{A}, F)$ and extend the map $\int$ to it. Theoretically we have achieved our goal of defining a notion of integral on a normed vector space $\mathrm{L}_\mu(X, \mathcal{A}, F)$ which is complete and in which $\mathrm{Step}_\mu(X, \mathcal{A}, F)$ is dense, but the elements in this abstract

completion are equivalence classes of Cauchy sequences, and are not easily identifiable with functions.

We will follow a different path, still very much inspired by the completion method involving Cauchy sequences, the twist being that we consider Cauchy sequences whose limit is known ahead of time, but where we use pointwise convergence *almost everywhere*, instead of pointwise convergence.

5.6 Integrable Functions; the Spaces $\mathcal{L}_\mu(X, \mathcal{A}, F)$ and $\mathrm{L}_\mu(X, \mathcal{A}, F)$

In this section we construct the completion $\mathcal{L}_\mu(X, \mathcal{A}, F)$ of the vector space $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ equipped with the semi-norm N_1, and construct the integral of a function in $\mathcal{L}_\mu(X, \mathcal{A}, F)$. The semi-norm N_1 is extended to $\mathcal{L}_\mu(X, \mathcal{A}, F)$ as a semi-norm $\| \|_1$ called the L^1-semi-norm, and we find that the space of functions such that $\|f\|_1 = 0$ is the set $\mathcal{N}$ of functions in $\mathcal{L}_\mu(X, \mathcal{A}, F)$ that are zero a.e. Then we define the quotient space $\mathrm{L}_\mu(X, \mathcal{A}, F) = \mathcal{L}_\mu(X, \mathcal{A}, F)/\mathcal{N}$. The space $\mathrm{L}_\mu(X, \mathcal{A}, F)$ is the completion of $\mathrm{Step}_\mu(X, \mathcal{A}, F)$; this is one of the most important results of this section (the Fischer–Riesz theorem).

As in the previous section the results that we state without proof are proved either in Marle [69] or in Lang [62].

Recall the following definitions.

Definition 5.10. A sequence (f_n) of functions $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ is a N_1-*Cauchy sequence* if for every $\epsilon > 0$, there is some $N > 0$, such that for all $m, n \geq N$, we have $N_1(f_m - f_n) < \epsilon$, where $N_1(f_m - f_n) = \int \|f_m - f_n\| \, d\mu$. A sequence (f_n) of maps $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ *converges pointwise almost everywhere* to a limit $f \colon X \to F$ if there is a null set Z such that for every $x \in X - Z$, for every $\epsilon > 0$, there is some $N > 0$, such that $\|f(x) - f_n(x)\| < \epsilon$ for all $n \geq N$.

We define the space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ as follows.

Definition 5.11. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space, with F a Banach space and $\mathcal{B}$ its Borel σ-algebra. The set $\mathcal{L}_\mu(X, \mathcal{A}, F)$ of *μ-integrable functions* consists of all functions $f \colon X \to F$ such that there is some N_1-Cauchy sequence $(f_n)_{n\geq 1}$ of μ-step maps $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ which converges pointwise almost everywhere to f. A sequence $(f_n)_{n\geq 1}$ of μ-step maps as above is called an *approximation sequence* for f.

Observe that not only do we require that the sequence $(f_n)_{n\geq 1}$ converges pointwise to f a.e., which makes f a μ-measurable map, but also that this sequence is N_1-Cauchy. This is the key to defining the notion of integral of the function f, as shown technically in Proposition 5.17.

We will see that $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is a vector space containing $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$, and a subspace of $\mathcal{M}_\mu(X, \mathcal{A}, F)$. Also, and this is the point of the construction, $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is complete with respect to the extension $\| \|_1$ of the semi-norm N_1 to $\mathcal{L}_\mu(X, \mathcal{A}, F)$, a fact that is not obvious at all from the definition.

The *crucial point* is that Definition 5.11 is designed so that the following fact holds.

Proposition 5.17. *For any N_1-Cauchy sequence $(f_n)_{n\geq 1}$ of μ-step maps, the sequence of integrals $\left(\int f_n d\mu\right)_{n\geq 1}$ is a Cauchy sequence in F.*

Proof. Indeed, by Proposition 5.16(3), we have

$$\left\| \int f_n d\mu - \int f_m d\mu \right\| = \left\| \int (f_n - f_m) d\mu \right\| \leq \int \|f_n - f_m\| \, d\mu = N_1(f_n - f_m),$$

and since by hypothesis (f_n) is an N_1-Cauchy sequence, the sequence $\left(\int f_n d\mu\right)_{n\geq 1}$ is a Cauchy sequence in F. Indeed, for every $\epsilon > 0$, since the sequence (f_n) is N_1-Cauchy, there is some $N > 0$ such that $N_1(f_n - f_m) < \epsilon$ for all $m, n \geq N$, which implies that $\left\| \int f_n d\mu - \int f_m d\mu \right\| < \epsilon$ for all $m, n \geq N$. ☐

Then, since F is complete, the sequence $\left(\int f_n d\mu\right)_{n\geq 1}$ converges to an element of F, and if $(f_n)_{n\geq 1}$ is an approximation sequence for $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, it is natural to define the integral of f as

$$\int f d\mu = \lim_{n \mapsto \infty} \int f_n d\mu.$$

The problem is that the definition of $\int f d\mu$ depends on the approximation sequence $(f_n)_{n\geq 1}$ chosen for f.

Actually, the definition of $\int f d\mu$ does *not depend* on the approximation approximation sequence $(f_n)_{n\geq 1}$ chosen for f, but proving this is nontrivial. The proof relies on a remarkable fact called the *fundamental lemma of integration* by Serge Lang; see [62], Chapter VI, §3.

Proposition 5.18. *Let $(f_n)_{n\geq 1}$ be any N_1-Cauchy sequence of maps $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$. There exists a subsequence (g_k) which converges pointwise almost everywhere to a limit $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$. Furthermore, for any $\epsilon > 0$, there is a measurable subset $Z_\epsilon \in \mathcal{A}$ such that $\mu(Z_\epsilon) \leq \epsilon$, and the subsequence (g_k) converges uniformly to f on $X - Z_\epsilon$ (recall Definition 2.6).*

Proof. We follow Lang's proof; see Lang [62] (Chapter VI, §3, Lemma 3.1). Since $(f_n)_{n\geq 1}$ is an N_1-Cauchy sequence, for every $k \geq 1$, there is some M_k such that if $m, n \geq M_k$, then

$$N_1(f_m - f_n) < \frac{1}{2^{2k}}.$$

By induction we can define the sequence (M_k) such that $M_k < M_{k+1}$ for all $k \geq 1$. We define the subsequence (g_k) such that $g_k = f_{M_k}$. By construction, we have

$$N_1(g_m - g_n) < \frac{1}{2^{2n}} \quad \text{if } m \geq n. \tag{$*_1$}$$

In particular, the sequence (g_k) is N_1-Cauchy.

Our next goal is to prove that the series

$$g_1(x) + \sum_{k=1}^{\infty}(g_{k+1}(x) - g_k(x)) \tag{$*_2$}$$

converges absolutely (thus pointwise) to a function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ outside a subset Z of measure 0, and in fact the convergence is uniform except on a set of arbitrary small measure. Observe that the partial sums of the series $(*_2)$ are $g_n(x)$, so this establishes the statement of the proposition.

Let Y_n be the set of all $x \in X$ such that

$$\|g_{n+1}(x) - g_n(x)\| \geq \frac{1}{2^n}.$$

Since the g_k are μ-step maps and since by Proposition 5.11(3) and 5.11(4) the function $\|g_{n+1} - g_n\|$ is measurable, the set Y_n has finite measure. Since

$$\|g_{n+1}(x) - g_n(x)\| \geq \frac{1}{2^n}$$

on Y_n, using $(*_1)$ we have

$$\frac{1}{2^n}\mu(Y_n) = \int_{Y_n} \frac{1}{2^n} d\mu \leq \int_X \|g_{n+1}(x) - g_n(x)\| \, d\mu = N_1(g_{n+1} - g_n) < \frac{1}{2^{2n}}.$$

The above implies that

$$\mu(Y_n) < \frac{1}{2^n}. \tag{$*_3$}$$

If we let

$$Z_n = \bigcup_{k \geq n} Y_n,$$

then we have

$$\mu(Z_n) \leq \sum_{k \geq n} \mu(Y_n) \leq \sum_{k \geq n} \frac{1}{2^k} = \frac{1}{2^n}\left(\sum_{k=0}^{\infty} \frac{1}{2^k}\right) < \frac{1}{2^{n-1}}.$$

If $x \notin Z_n$, then for all $k \geq n$ we have

$$\|g_{k+1}(x) - g_k(x)\| < \frac{1}{2^k},$$

and this implies that we have

$$\sum_{k=n}^{\infty} \|g_{k+1}(x) - g_k(x)\| < \sum_{k=n}^{\infty} \frac{1}{2^k} < \frac{1}{2^{n-1}},$$

so the series

$$\|g_1(x)\| + \sum_{k=1}^{\infty} \|g_{k+1}(x) - g_k(x)\|$$

converges, and since F is complete, the series in $(*_2)$ is uniformly convergent outside Z_n to a limit $f(x)$. For every $\epsilon > 0$ there is an n such that $\frac{1}{2^{n-1}} < \epsilon$, so the statement about uniform convergence holds with $Z = Z_n$. If we define Z by

$$Z = \bigcap_{n\geq 1} Z_n,$$

then $\mu(Z) = 0$ and since $x \in X - Z$ iff there is some n such that $x \in X - Z_n$, so the series $(*_2)$ converges to $f(x)$ and thus $g_n(x)$ converges to $f(x)$, which means that the sequence (g_n) converges pointwise to f outside the subset Z of measure zero. $\square$

It should be mentioned that in general, the original sequence (f_n) may not converge pointwise, even a.e. An example of such a sequence (f_n) which is N_1-Cauchy, yet $(f_n(x))$ diverges for every $x \in X$, is given in Schwartz [86] (Chapter 5, §6).

Using Proposition 5.18, the following result is obtained. This result implies that the integral $\int f d\mu$ is well defined.

Proposition 5.19. *Let $(f_n)_{n\geq 1}$ and $(g_n)_{n\geq 1}$ be two N_1-Cauchy sequences of μ-step maps $f_n, g_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$ which approximate the same function f. The sequences $\left(\int f_n d\mu\right)_{n\geq 1}$ and $\left(\int g_n d\mu\right)_{n\geq 1}$ converge to the same limit, and*

$$\lim_{n\mapsto\infty} \int \|f_n - g_n\|\, d\mu = 0,$$

that is, $\lim_{n\mapsto\infty} N_1(f_n - g_n) = 0$.

Proof. We follow Lang's proof; see Lang [62] (Chapter VI, §3, Lemma 3.2). The convergence of the sequences $\left(\int f_n d\mu\right)_{n\geq 1}$ and $\left(\int g_n d\mu\right)_{n\geq 1}$ follows from Proposition 5.17. Note that

$$\left\| \int f_n d\mu - \int f_m d\mu \right\| \leq N_1(f_n - f_m). \tag{$*$}$$

Next let $h_n = f_n - g_n$. Since the maps f_n and g_n approximate the same function f, the fact that

$$\int \|h_n - h_m\|\, d\mu = \int \|f_n - g_n - (f_m - g_m)\|\, d\mu \leq \int \|f_n - f_m\|\, d\mu + \int \|g_n - g_m\|\, d\mu$$

implies that the sequence (h_n) is N_1-Cauchy and converges almost everywhere to the zero function. We will prove that $N_1(h_n) = \int \|h_n\|\, d\mu$ converges to 0, and since

$$\left\| \int h_n d\mu \right\| \leq \int \|h_n\|\, d\mu,$$

the integral $\int h_n d\mu$ also converges to 0.

Since (h_n) is N_1-Cauchy, for every $\epsilon > 0$ there is some $N > 0$ such that for all $m, n \geq N$ we have

$$N_1(h_n - h_m) < \epsilon. \tag{$*_1$}$$

Since f_n and g_n are μ-step functions for all n there is some subset A of finite measure such that h_N vanishes outside A. Then for all $n \geq N$ we have

$$\int_{X-A} \|h_n\| \, d\mu = \int_{X-A} \|h_n - h_N\| \, d\mu \leq \int_X \|h_n - h_N\| \, d\mu = N_1(h_n - h_N) < \epsilon,$$

so

$$\int_{X-A} \|h_n\| \, d\mu < \epsilon, \quad n \geq N. \tag{$*_2$}$$

By Proposition 5.18, there is a subset Z of A such that

$$\mu(Z) < \frac{\epsilon}{1 + \|h_N\|_\infty}, \tag{$*_3$}$$

and a subsequence (h_m) that tends to 0 uniformly on $A - Z$. The reason for using $1 + \|h_N\|_\infty$ is to avoid division by zero. The point is that in all cases we have $\mu(Z) \|h_N\|_\infty < \epsilon$. Then for $m \geq N$ large enough we conclude that

$$\int_{A-Z} \|h_m\| \, d\mu < \epsilon. \tag{$*_4$}$$

Finally for m large enough we have

$$\begin{aligned} \int_Z \|h_m\| \, d\mu &\leq \int_Z \|h_n - h_N\| \, d\mu + \int_Z \|h_N\| \, d\mu \\ &\leq N_1(h_n - h_N) + \mu(Z) \|h_N\|_\infty < 2\epsilon, \end{aligned}$$

so

$$\int_Z \|h_m\| \, d\mu < 2\epsilon. \tag{$*_5$}$$

Using $(*_2)$, $(*_2)$ and $(*_5)$, we obtain

$$N_1(h_m) = \int_X \|h_m\| \, d\mu = \int_{X-A} \|h_m\| \, d\mu + \int_Z \|h_m\| \, d\mu + \int_{A-Z} \|h_m\| \, d\mu < \epsilon + \epsilon + 2\epsilon = 4\epsilon,$$

proving our result. $\square$

Proposition 5.19 justifies the following definition.

Definition 5.12. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space, with F a Banach space and $\mathcal{B}$ its Borel σ-algebra. For any function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, we define the *integral*[3] of f (*with respect to* μ) by

$$\int f d\mu = \lim_{n \mapsto \infty} \int f_n d\mu,$$

where $(f_n)_{n \geq 1}$ is any approximation sequence of f by μ-step maps.

Proposition 5.20. *For any function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ and any approximation sequence (f_n) of f with $f_n \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, we have $\|f\| \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, and the sequence $(\|f_n\|)$ is an approximation sequence of $\|f\|$ with $\|f_n\| \in \mathcal{S}tep_\mu(X, \mathcal{A}, \mathbb{R})$. Furthermore,*

$$\int \|f\| \, d\mu = \lim_{n \mapsto \infty} \int \|f_n\| \, d\mu = \lim_{n \mapsto \infty} N_1(f_n).$$

Proof. Since the sequence (f_n) converges pointwise to f a.e., we verify immediately that the sequence $(\|f_n\|)$ converges pointwise to $\|f\|$ a.e. Since

$$| \, \|f_n\| - \|f_m\| \, | \leq \|f_n - f_m\|$$

(see just after Definition A.3), by Proposition 5.16(5) we have

$$N_1(\|f_n\| - \|f_m\|) = \int | \, \|f_n\| - \|f_m\| \, | d\mu \leq \int \|f_n - f_m\| \, d\mu = N_1(f_n - f_m),$$

and since (f_n) is an N_1-Cauchy sequence, the sequence $(\|f_n\|)$ is an N_1-Cauchy sequence. Therefore $\|f\| \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, and $(\|f_n\|)$ is an approximation sequence of $\|f\|$. By definition of the integral,

$$\int \|f\| \, d\mu = \lim_{n \mapsto \infty} \int \|f_n\| \, d\mu = \lim_{n \mapsto \infty} N_1(f_n),$$

as claimed. ∎

Definition 5.13. For any function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, we define the *$L^1$-semi-norm* $\|f\|_1$ of f as

$$\|f\|_1 = \int \|f\| \, d\mu.$$

Observe that if $f \in \mathcal{S}tep_\mu(X, \mathcal{A}, F)$, then $\|f\|_1 = N_1(f)$. The following proposition is easily shown by passing to the limit.

Proposition 5.21. *The set $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is a vector space, and $\| \, \|_1$ is a semi-norm on $\mathcal{L}_\mu(X, \mathcal{A}, F)$. The space $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ is a subspace of $\mathcal{L}_\mu(X, \mathcal{A}, F)$, which is a subspace of $\mathcal{M}_\mu(X, \mathcal{A}, F)$.*

[3]This integral is usually called the *Lebesgue integral* or *Bochner integal*.

We are almost ready to prove that $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is complete with respect to the L^1-semi-norm, but first we need the following result.

Proposition 5.22. *The subspace $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ is dense in $\mathcal{L}_\mu(X, \mathcal{A}, F)$ with respect to the L^1-semi-norm $\| \; \|_1$. Furthermore, any approximation sequence $(f_n)_{n\geq 1}$ of f by μ-step maps converges to f according to the semi-norm $\| \; \|_1$.*

Proof. Pick any $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ and let (f_n) be any approximation sequence for f. This means that the sequence (f_n) is a N_1-Cauchy sequence of μ-step maps which converges pointwise to f a.e. We will prove that

$$\lim_{n \mapsto \infty} \|f - f_n\|_1 = 0,$$

which shows that the sequence (f_n) converges to f in the L^1-semi-norm.

First we claim that for any fixed $n \geq 1$, the sequence $(\|f_p - f_n\|)_{p\geq 1}$ is an N_1-Cauchy sequence which converges to $\|f - f_n\|$ a.e. Indeed, we have

$$\begin{aligned} \int | \, \|f_p - f_n\| - \|f_q - f_n\| \, | d\mu &\leq \int \|f_p - f_n - (f_q - f_n)\| \, d\mu \\ &= \int \|f_p - f_q\| \, d\mu = N_1(f_p - f_q), \end{aligned}$$

and since (f_n) is a N_1-Cauchy sequence, for every $\epsilon > 0$, there is some $N > 0$ such that $N_1(f_p - f_q) < \epsilon$ for all $p, q \geq N$, which shows that $(\|f_p - f_n\|)_{p\geq 1}$ is a N_1-Cauchy sequence (in $\mathbb{R}$). The fact that $(f_p)_{p\geq 1}$ converges pointwise a.e. to f immediately implies that $\|f_p - f_n\|$ converges to $\|f - f_n\|$ a.e. By definition of $\| \; \|_1$ and of the integral

$$\|f - f_n\|_1 = \int \|f - f_n\| \, d\mu = \lim_{p \mapsto \infty} \int \|f_p - f_n\| \, d\mu = \lim_{p \mapsto \infty} N_1(f_p - f_n).$$

Thus for every $\epsilon > 0$, there is some $M_1 > 0$ such that

$$| \, \|f - f_n\|_1 - N_1(f_p - f_n)| < \frac{\epsilon}{2} \quad \text{for all } p \geq M_1,$$

and since (f_n) is an N_1-Cauchy sequence, there is some $M_2 > 0$ such that

$$N_1(f_p - f_n) < \frac{\epsilon}{2} \quad \text{for all } n, p \geq M_2,$$

so for all $n, p \geq \max(M_1, M_2)$ we have

$$\|f - f_n\|_1 \leq | \, \|f - f_n\|_1 - N_1(f_p - f_n)| + N_1(f_p - f_n) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which proves that $\lim_{n\mapsto\infty} \|f - f_n\|_1 = 0$; that is, the sequence (f_n) converges to f in the L^1-semi-norm. ∎

Remark: It appears that Lang [62] skipped this step, which is used in the proof his Theorem 3.4, and the proof of the next theorem.

Now we can prove one of our main theorems.

5.7 The Fischer–Riesz Theorem

Theorem 5.23. *(Fischer–Riesz) The space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is complete with respect to the L^1-semi-norm. This means that for every sequence $(f_n)_{n\geq 1}$ of functions $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, if (f_n) is $\| \|_1$-Cauchy, then there is some function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ such that for every $\epsilon > 0$, there is some $N > 0$ such that $\|f - f_n\|_1 < \epsilon$ for all $n \geq N$.*

Proof. Let $(f_n)_{n\geq 1}$ be an $\| \|_1$-Cauchy sequence of functions $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, F)$. By Proposition 5.22, for every n there is an approximation sequence $(g_{n,m})_{m\geq 1}$ of μ-step maps that converges to f_n pointwise a.e. and in the $\| \|_1$-semi-norm. Thus, for every $n \geq 1$, there is some $m(n)$ such that

$$\left\| f_n - g_{n,m(n)} \right\|_1 \leq \frac{1}{n}. \tag{$*_6$}$$

Each sequence $(g_{n,m(n)})_{n\geq 1}$ is N_1-Cauchy, because

$$\begin{aligned} N_1(g_{p,m(p)} - g_{q,m(q)}) &= \left\| g_{p,m(p)} - g_{q,m(q)} \right\|_1 \\ &\leq \left\| g_{p,m(p)} - f_p \right\|_1 + \|f_p - f_q\|_1 + \left\| f_q - g_{q,m(q)} \right\|_1 \\ &\leq \frac{1}{p} + \frac{1}{q} + \|f_p - f_q\|_1 , \end{aligned}$$

and the right-hand side tends to 0 when p and q tend to $+\infty$, since the sequence (f_n) is $\| \|_1$-Cauchy. By Proposition 5.18, for each sequence $(g_{n,m(n)})_{m\geq 1}$, we can extract a subsequence $(g_{n_k,m(n_k)})_{k\geq 1}$ that converges pointwise a.e. to some function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, and is also N_1-Cauchy. By the second part of Proposition 5.22, the subsequence $(g_{n_k,m(n_k)})_{k\geq 1}$ converges to f for the semi-norm $\| \|_1$. Since $(g_{n,m(n)})_{n\geq 1}$ is N_1-Cauchy and has a subsequence $(g_{n_k,m(n_k)})_{k\geq 1}$ $\| \|_1$-convergent to f, it also $\| \|_1$-converges to the function f. Using ($*_6$) and the inequality

$$\begin{aligned} \|f - f_n\|_1 &\leq \left\| f - g_{n,m(n)} \right\|_1 + \left\| g_{n,m(n)} - f_n \right\|_1 \\ &\leq \left\| f - g_{n,m(n)} \right\|_1 + \frac{1}{n}, \end{aligned}$$

and since the sequence $(g_{n,m(n)})_{n\geq 1}$ $\| \|_1$-converges to the function f, we deduce that the sequence $(f_n)_{n\geq 1}$ converges to f for the semi-norm $\| \|_1$.

In the diagram below, the original sequence $(f_n)_{n\geq 1}$ is shown as the top horizontal row. Below each f_n, we have the approximation sequence $(g_{n,m})_{m\geq 1}$ shown as an ascending column. The sequence of $g_{n,m(n)}$ chosen for each n is shown in boldface, and its subsequence in red.

$$\begin{array}{ccccccccc}
f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & \cdots & f_n & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \\
g_{1,m(n)} & g_{2,m(n)} & g_{3,m(n)} & g_{4,m(n)} & g_{5,m(n)} & g_{6,m(n)} & \cdots & \mathbf{g_{n,m(n)}} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \\
g_{1,6} & g_{2,6} & \mathbf{g_{3,6}} & g_{4,6} & \mathbf{g_{5,6}} & g_{6,6} & \cdots & g_{n,6} & \cdots \\
g_{1,5} & \mathbf{g_{2,5}} & g_{3,5} & g_{4,5} & g_{5,5} & g_{6,5} & \cdots & g_{n,5} & \cdots \\
g_{1,4} & g_{2,4} & g_{3,4} & \mathbf{g_{4,4}} & g_{5,4} & g_{6,4} & \cdots & g_{n,4} & \cdots \\
\mathbf{g_{1,3}} & g_{2,3} & g_{3,3} & g_{4,3} & g_{5,3} & g_{6,3} & \cdots & g_{n,3} & \cdots \\
g_{1,2} & g_{2,2} & g_{3,2} & g_{4,2} & g_{5,2} & \mathbf{g_{6,2}} & \cdots & g_{n,2} & \cdots \\
g_{1,1} & g_{2,1} & g_{3,1} & g_{4,1} & g_{5,1} & g_{6,1} & \cdots & g_{n,1} & \cdots
\end{array}$$

This concludes the proof. $\square$

The following properties of the integral are easily obtained by passing to the limit.

Proposition 5.24. *Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space, with F a Banach space and $\mathcal{B}$ its Borel σ-algebra. The following properties hold:*

1. *For any $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, if $f = 0$ a.e., then $\int f d\mu = 0$. More generally, if $f, g \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ and if $f = g$ a.e., then $\int f d\mu = \int g d\mu$.*

2. *For any $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, and for any measurable subset $A \in \mathcal{A}$, the integral $\int_A f d\mu = \int f \chi_A d\mu$ exists, and*
$$\left\| \int_A f d\mu \right\| \leq \int_A \|f\| \, d\mu \leq \|f\|_\infty \mu(A).$$
Furthermore, if $A, B \in \mathcal{A}$ are disjoint, then
$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

3. *The integral $\int : \mathcal{L}_\mu(X, \mathcal{A}, F) \to F$ is linear.*

4. *For any $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, we have $\|f\| \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, and*
$$\left\| \int f d\mu \right\| \leq \int \|f\| \, d\mu = \|f\|_1 .$$

5. *If $f, g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, then $\sup(f, g), \inf(f, g), f^+, f^-, |f| \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$. Since $f^+ = (|f| + f)/2$ and $f^- = (|f| - f)/2$, we have $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ iff $f^+ \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ and $f^- \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$.*

6. *If $f, g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ and $f \leq g$ a.e., then $\int f d\mu \leq \int g d\mu$. In particular, if $f \geq 0$ a.e., then $\int f d\mu \geq 0$.*

7. *Let F_1 and F_2 be two Banach spaces, and let $h\colon F_1 \to F_2$ be a linear map (or semi-linear map when the field is $\mathbb{C}$). If $f \in \mathcal{L}_\mu(X, \mathcal{A}, F_1)$, then $h \circ f \in \mathcal{L}_\mu(X, \mathcal{A}, F_2)$, and*

$$\int (h \circ f) d\mu = h\left(\int f d\mu\right).$$

8. *Let F_1 and F_2 be two Banach spaces, and let $F_1 \times F_2$ be the product space (under any of the product norms defined just before Definition A.13). Then there is an isomorphism between $\mathcal{L}_\mu(X, \mathcal{A}, F_1 \times F_2)$ and $\mathcal{L}_\mu(X, \mathcal{A}, F_1) \times \mathcal{L}_\mu(X, \mathcal{A}, F_2)$, and if $f = (f_1, f_2)$, then*

$$\int f d\mu = \left(\int f_1 d\mu, \int f_2 d\mu\right).$$

In particular, since $\mathbb{C}$ is isomorphic to $\mathbb{R} \times \mathbb{R}$, a function $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$ corresponds uniquely to a function $f = u + iv$ with $u, v \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, and we have

$$\int f d\mu = \int u d\mu + i \int v d\mu.$$

Remark: Observe that in our approach, if f is a real-valued function or a complex-valued function, the integral $\int f d\mu$ is defined directly. There is another approach in which the integral is first defined for real-valued *positive* functions. Then the integral of a real-valued function f is defined in terms of the integrals of f^+ and f^-, and the integral of a complex valued function $f = u + iv$ is defined in terms of the integrals of u^+, u^-, v^+, v^-. See Rudin [79], Definition 1.31.

5.8 Characterizing Which Functions Satisfy $\|f\|_1 = 0$

The next step is to identify the functions f in $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ such that $\|f\|_1 = 0$. For this, we need two propositions.

Proposition 5.25. *For any function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, and for any real $a > 0$, the subset $E_a = \{x \in X \mid \|f(x)\| \geq a\}$ can be written as $E_a = (B - Z) \cup N$, with B a measurable subset of finite measure, and Z and N two null subsets. The function f vanishes outside of a σ-finite measurable set.*

Proof. We begin by showing that E_a is a measurable set with finite measure. Since $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, by Proposition 5.5(2), the function $\|f\|$ is measurable, so E_a is measurable. By Proposition 5.18, there is an N_1-Cauchy sequence (f_n) of μ-step maps which converges pointwise to f a.e., and for every $\epsilon > 0$, there is a measurable subset Z_1 of measure $\mu(Z_1) < \epsilon$ such that f_n converges uniformly to f on $X - Z_1$. Pick $\epsilon = a/2$. The uniform convergence implies that there is some $M > 0$ such that for all $n \geq M$ and all $x \in X - Z_1$,

$$\|f(x) - f_n(x)\| \leq a/2,$$

and since $\|f(x)\| \leq \|f(x) - f_n(x)\| + \|f_n(x)\|$, we have

$$\|f(x)\| \leq \|f_n(x)\| + a/2,$$

and thus $\|f_n(x)\| \geq \|f(x)\| - a/2$, so $\|f(x)\| \geq a$ implies $\|f_n(x)\| \geq a/2$, which implies that

$$E_a \subseteq \{x \in Z_1 \mid \|f(x)\| \geq a\} \cup \{x \in X - Z_1 \mid \|f_n(x)\| \geq a/2\},$$

where both sets on the right-hand side have finite measure (the second one because f_n is a μ-step function, and so is $\|f_n\|$, and a μ-step function vanishes outside of a set of finite measure). See Figure 5.16. Since both sets are measurable and the set on the right-hand side has finite measure, by Proposition 4.7(2) we deduce that E_a has finite measure. Since

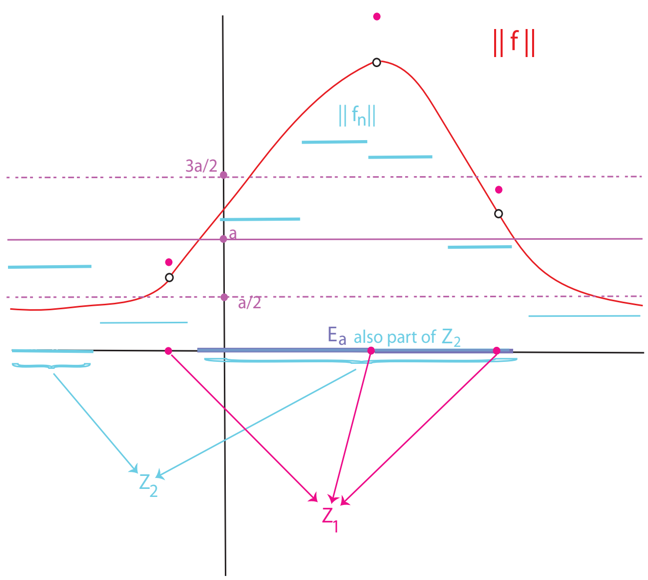


Figure 5.16: Let $X = \mathbb{R}$. The red curve is the graph of $\|f\|$, while the aqua graph is the μ-step function $\|f_n\|$. The set Z_1 corresponds to the three magenta dots on the x-axis. The purple horizontal line segment is E_a, the two horizontal aqua line segments are Z_2, and $E_a \subseteq \{x \in Z_1 \mid \|f(x)\| \geq a\} \cup Z_2$, where $Z_2 = \{x \in X - Z_1 \mid \|f_n(x)\| \geq a/2\}$.

the function $\|f\|$ is μ-measurable, by Proposition 5.13(1), it is equal a.e. to a measurable function g, so there is a null set such Z that $\|f\|(x) = g(x)$ for all $x \in X - Z$. Then we have

$$\begin{aligned}
E_a &= \{x \in X \mid \|f(x)\| \geq a\} \\
&= \{x \in X - Z \mid \|f(x)\| \geq a\} \cup \{x \in Z \mid \|f(x)\| \geq a\} \\
&= \{x \in X - Z \mid g(x) \geq a\} \cup \{x \in Z \mid \|f(x)\| \geq a\} \\
&= (\{x \in X \mid g(x) \geq a\} - Z) \cup \{x \in Z \mid \|f(x)\| \geq a\} \\
&= (B - Z) \cup N,
\end{aligned}$$

with $B = \{x \in X \mid g(x) \geq a\}$ and $N = \{x \in Z \mid \|f(x)\| \geq a\}$. Since g is measurable and $[a, \infty)$ is closed, B is measurable, and N as a subset of a null set is a null set; see Figure 5.17.

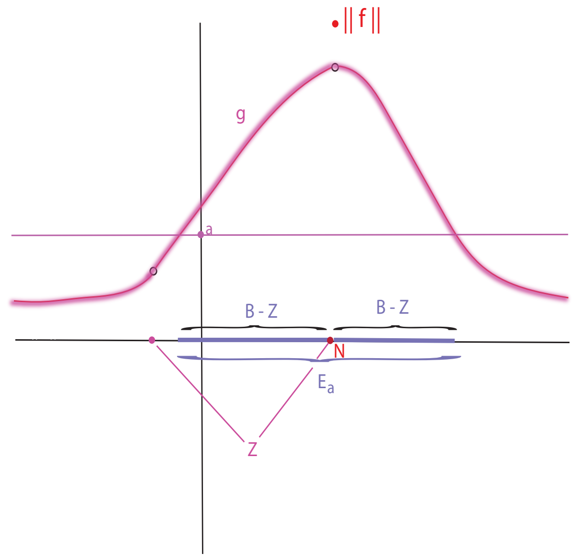


Figure 5.17: A continuation of Figure 5.16 where $\|f\|$ is replaced by the magenta bell curve g. Note that Z is the union of the two reddish dots while N is the darker red dot contained within E_a. In this particular illustration $B = E_a$.

What we showed above with $\|f\|$ replaced by g measurable implies that B has finite measure. The second statement of the proposition follows from Proposition 5.13(2). $\square$

Proposition 5.18 can be promoted to $\mathcal{L}_\mu(X, \mathcal{A}, F)$ as follows.

Proposition 5.26. *Let $(f_n)_{n\geq 1}$ be any $\| \|_1$-Cauchy sequence of maps $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ that converges to some function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ in the semi-norm $\| \|_1$. There exists a subsequence (g_k) which converges pointwise almost everywhere to f. Furthermore, for any $\epsilon > 0$, there is a measurable subset $Z_\epsilon \in \mathcal{A}$ such that $\mu(Z_\epsilon) \leq \epsilon$, and the subsequence $(g_k$ converges uniformly to f on $X - Z_\epsilon$ (recall Definition 2.6).*

Proposition 5.26 is proven in Lang [62] (Chapter VI, Theorem 5.2). The proof is very similar to the proof of Proposition 5.18. However, the f_n are no longer μ-step functions so we need Proposition 5.25 to justify the fact that the sets Y_n have finite measure.

Here are some corollaries of Proposition 5.26.

Proposition 5.27. *For any function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, we have $\|f\|_1 = 0$ iff $f = 0$ a.e.*

Proof. If $f = 0$ a.e., then $\|f\| = 0$ a.e., and by Proposition 5.24(1), we have $\|f\|_1 = 0$. Conversely, the sequence (f_n) where f_n is the zero function is $\| \|_1$-Cauchy and converges to f in the $\| \|_1$-norm. By Proposition 5.26 there is a subsequence that converges pointwise a.e. to f. But since f_n is the zero function for all n, this subsequence also converges pointwise a.e. to the zero function, so $f = 0$ a.e. ☐

Proposition 5.27 is the second main important result of this section because it provides a very natural characterization of the functions f such that $\|f\|_1 = 0$.

Proposition 5.28. *Let (f_n) be a sequence of functions $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, F)$. If (f_n) is an $\| \|_1$-Cauchy sequence which converges pointwise a.e. to a function $f\colon X \to F$, then $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, and (f_n) converges to f in the semi-norm $\| \|_1$.*

Proof. Since the sequence (f_n) is an $\| \|_1$-Cauchy sequence, by the Fischer–Riesz theorem (Theorem 5.23) it converges to some function $g \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ in the $\| \|_1$-semi-norm. By Proposition 5.26, some subsequence $(f_{n_k})_{k\geq 1}$ of (f_n) converges pointwise a.e. to g. Since (f_n) converges pointwise a.e. to f, the subsequence $(f_{n_k})_{k\geq 1}$ also converges pointwise a.e. to f, so $f = g$ a.e., and since $g \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ and (f_n) converges to g in the semi-norm $\| \|_1$, we also have $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, and (f_n) converges to f in the semi-norm $\| \|_1$. ☐

The main disadvantage of the space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is that it is not a normed vector space under the semi-norm $\| \|_1$. Thus it is natural to consider the quotient of $\mathcal{L}_\mu(X, \mathcal{A}, F)$ by the subspace $\mathcal{N}$ consisting of the functions such that $\|f\|_1 = 0$.

Definition 5.14. Let $\mathcal{N}$ be the subspace of $\mathcal{L}_\mu(X, \mathcal{A}, F)$ given by

$$\mathcal{N} = \{f \in \mathcal{L}_\mu(X, \mathcal{A}, F) \mid \|f\|_1 = 0\},$$

which is just the subspace of function equal to 0 a.e. Then we define $\mathrm{L}_\mu(X, \mathcal{A}, F)$ as the quotient space

$$\mathrm{L}_\mu(X, \mathcal{A}, F) = \mathcal{L}_\mu(X, \mathcal{A}, F)/\mathcal{N}.$$

For any equivalence class $\mathbf{f} \in \mathrm{L}_\mu(X, \mathcal{A}, F)$, since for any two representatives $f, g \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ in the equivalence class $\mathbf{f}$, we have $f = g$ a.e., by Proposition 5.24(1),

$$\int f d\mu = \int g d\mu,$$

so we can define $\int \mathbf{f} d\mu$ as

$$\int \mathbf{f} d\mu = \int f d\mu.$$

Similarly, $\|\mathbf{f}\|_1$ is defined as

$$\|\mathbf{f}\|_1 = \|f\|_1\,,$$

for any $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ in the equivalence class $\mathbf{f}$.

The following theorem is immediately obtained from Theorem 5.23 by passing to the quotient.

Theorem 5.29. *(Fischer–Riesz) The semi-norm* $\| \|_1$ *on* $\mathrm{L}_\mu(X, \mathcal{A}, F)$ *induced by the semi-norm* $\| \|_1$ *on* $\mathcal{L}_\mu(X, \mathcal{A}, F)$ *by passing to the quotient is a norm on* $\mathrm{L}_\mu(X, \mathcal{A}, F)$ *called the* L^1*-norm. With this norm, the space* $\mathrm{L}_\mu(X, \mathcal{A}, F)$ *is complete (it is a Banach space). The subspace* $\mathrm{Step}_\mu(X, \mathcal{A}, F)$ *is dense in* $\mathrm{L}_\mu(X, \mathcal{A}, F)$.

Finally, the following proposition confirms one of our earlier claims.

Proposition 5.30. *The space* $\mathrm{L}_\mu(X, \mathcal{A}, F) = \mathcal{L}_\mu(X, \mathcal{A}, F)/\mathcal{N}$ *is isomophic to the Cauchy completion of the space* $\mathcal{S}tep_\mu(X, \mathcal{A}, F)/\mathcal{SN}$*; see the diagram*

$$\begin{tikzcd}
\mathcal{S}tep_\mu(X, \mathcal{A}, F) \arrow[r, "completion"] \arrow[d, "quotient"'] & \mathcal{L}_\mu(X, \mathcal{A}, F) \arrow[d, "quotient"] \\
\mathrm{Step}_\mu(X, \mathcal{A}, F) = \mathcal{S}tep_\mu(X, \mathcal{A}, F)/\mathcal{SN} \arrow[r, "completion"'] & \mathrm{L}_\mu(X, \mathcal{A}, F) = \mathcal{L}_\mu(X, \mathcal{A}, F)/\mathcal{N}.
\end{tikzcd}$$

In the next section we consider some fundamental convergence theorems. A very useful corollary of these theorems is that a function f belongs to $\mathcal{L}_\mu(X, \mathcal{A}, F)$ iff it belongs to $\mathcal{M}_\mu(X, \mathcal{A}, F)$ (it is μ-measurable), and if $\int \|f\| \, d\mu$ exists. By Proposition 5.21, the space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is a subspace of $\mathcal{M}_\mu(X, \mathcal{A}, F)$, and we already know from Proposition 5.24(4) that if $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ then $\|f\| \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$. The converse is not trivial, but it will be shown as a corollary of the dominated convergence theorem discussed in Section 5.9.

5.9 Fundamental Convergence Theorems

Besides the fact that the Lebesgue–Bochner integral is defined for a much bigger class of functions than the regulated functions (or the Riemann-integrable functions), one of its main advantages is that it leads to simple and flexible criteria to tell whether the limit of a sequence of integrable functions is integrable. Most of these results allow interchanging a limit and an integral. We begin with criteria applying to real-valued functions. These results actually apply to extended functions with values in $\mathbb{R} \cup \{+\infty\}$, but for simplicity we stick to functions $f \colon X \to \mathbb{R}$. As in the previous section the results that we state without proof are proven either in Marle [69] or in Lang [62].

Theorem 5.31. *(Monotone Convergence Theorem) Let* $(f_n)_{n \geq 1}$ *be a sequence of functions* $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ *such that* $f_n \leq f_{n+1}$ *for all* $n \geq 1$*, and assume that there is some* $M > 0$ *such that*

$$\left| \int f_n d\mu \right| \leq M \quad \text{for all } n \geq 1.$$

Then the sequence $(f_n)_{n\geq 1}$ *converges pointwise a.e., and also in the* $\| \|_1$*-norm, to a function* $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$. *We also have* $\lim_{n\mapsto\infty} \|f_n\|_1 = \|f\|_1$ *and*

$$\lim_{n\mapsto\infty} \int f_n d\mu = \int f d\mu.$$

The same result applies to a nonincreasing sequence (f_n) *(with* $f_n \geq f_{n+1}$ *for all* $n \geq 1$*).*

Proof. We follow Lang [62] (Chapter VI, §5, Theorem 5.5). Let

$$\alpha = \sup_k \int f_k d\mu,$$

which is well defined since

$$\left| \int f_n d\mu \right| \leq M \quad \text{for all } n \geq 1.$$

For $n \geq m$, since $f_n \leq f_{n+1}$ for all $n \geq 1$, we have

$$\begin{aligned}
\|f_n - f_m\|_1 &= \int (f_n - f_m) d\mu \\
&= \int f_n d\mu - \int f_m d\mu \\
&\leq \alpha - \int f_m d\mu,
\end{aligned}$$

which implies that (f_n) is a $\| \|_1$-Cauchy sequence. By the Fischer–Riesz theorem (Theorem 5.23), the sequence converges to some limit $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ in the $\| \|_1$-norm. By Proposition 5.26, there is a subsequence $(f_{n_k})_{k\geq 1}$ of (f_n) that converges a.e. to $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, and since the sequence (f_n) is increasing, by a standard ϵ-argument, it also converges a.e. to f. Since

$$| \, \|f_n\|_1 - \|f\|_1 \, | \leq \|f_n - f\|_1$$

and (f_n) converges to f in the $\| \|_1$-norm, we deduce that $\lim_{n\mapsto\infty} \|f_n\|_1 = \|f\|_1$. We have

$$\begin{aligned}
\left| \int f_n d\mu - \int f d\mu \right| &= \left| \int (f_n - f) d\mu \right| \\
&\leq \int |f_n - f| d\mu \\
&= \|f_n - f\|_1 ,
\end{aligned}$$

and since (f_n) converges to f in the $\| \|_1$-norm, this implies that

$$\lim_{n\mapsto\infty} \int f_n d\mu = \int f d\mu,$$

as claimed. ∎

The following theorem has a different flavor. It asserts the existence of the sup of a sequence of functions.

Theorem 5.32. *(Beppo–Levi) Let (f_n) be a sequence of functions $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$. If there is a function $g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ such that $g \geq 0$ and $|f_n| \leq g$ for all $n \geq 1$, then* $\sup_{n\geq 1} f_n$ *and* $\inf_{n\geq 1} f_n$ *belong to $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, and we have*

$$\sup_{n\geq 1} \int f_n d\mu \leq \int (\sup_{n\geq 1} f_n) d\mu \quad \text{and} \quad \int (\inf_{n\geq 1} f_n) d\mu \leq \inf_{n\geq 1} \int f_n d\mu.$$

Proof. We follow Lang [62] (Chapter VI, §5, Corollary 5.6). By Proposition 5.24(5), the functions

$$g_n = \sup\{f_1, \ldots, f_n\}$$

belong to $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ and form an increasing sequence bounded by g. Since

$$\int g_n d\mu \leq \int g_{n+1} d\mu \quad \text{and} \quad \int g_n d\mu \leq \int g d\mu,$$

there is some $M > 0$ such that $|\int g_n d\mu| \leq M$ for all $n \geq 1$. Therefore, by the monotone convergence theorem (Theorem 5.31), the sequence (g_n) converges pointwise a.e. to some function in $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, but (g_n) converges pointwise to $\sup_{n\geq 1} f_n$, so the sequence (g_n) converges pointwise a.e. to $\sup_{n\geq 1} f_n$. Since $f_n \leq \sup_{n\geq 1} f_n$, we have

$$\int f_n d\mu \leq \int (\sup_{n\geq 1} f_n) d\mu,$$

which implies

$$\sup_{n\geq 1} \int f_n d\mu \leq \int (\sup_{n\geq 1} f_n) d\mu,$$

as claimed. ∎

Given a sequence $(f_n)_{n\geq 1}$ of functions $f_n \colon X \to \mathbb{R}$ such that $f_n \geq 0$, recall that

$$\liminf f_n = \lim_{k\mapsto\infty} \inf_{n\geq k} f_n.$$

Theorem 5.33. *(Fatou's Lemma) Let $(f_n)_{n\geq 1}$ be a sequence of functions $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ such that $f_n \geq 0$. If $\liminf \|f_n\|_1 = \liminf \int f_n d\mu$ exists, then there is a function $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ such that $\liminf f_n$ converges pointwise to f a.e., and*

$$\int f d\mu \leq \liminf \int f_n d\mu$$

A proof of Theorem 5.33 is given in Lang [62] (Chapter VI, §5).

The next theorem applies to functions with values in any Banach space F and is the most important convergence theorem.

Theorem 5.34. *(Lebesgue Dominated Convergence Theorem) Let $(f_n)_{n\geq 1}$ be a sequence of functions $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, F)$. If (f_n) converges pointwise a.e. to a function $f\colon X \to F$, and if there is some function $g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ such that $g \geq 0$ and $\|f_n\| \leq g$ for all $n \geq 1$, then $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ and $(f_n)_{n\geq 1}$ converges to f in the $\| \|_1$-norm. Consequently*

$$\lim_{n\mapsto\infty} \int f_n d\mu = \int f d\mu.$$

Proof. We follow Marle [69] (Chapter 2, Section 4, Theorem 2.4.7). For each $n, p \geq 1$, let

$$\begin{aligned} g_{n,p} &= \sup_{\substack{n\leq m\leq n+p \\ n\leq r\leq n+p}} \|f_m - f_r\| \\ g_n &= \sup_{m,r\geq n} \|f_m - f_r\| = \sup_{p\geq 1} g_{n,p}. \end{aligned}$$

By Proposition 5.24(4,5), for all $n, p \geq 1$, we have $g_{n,p} \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, $g_{n,p} \leq g_{n,p+1}$, and

$$0 \leq g_{n,p} \leq 2g.$$

We get

$$\left| \int g_{n,p}\, d\mu \right| \leq 2 \int g\, d\mu,$$

so by the monotone convergence theorem (Theorem 5.31), the sequence $(g_{n,p})_{p\geq 1}$ converges pointwise a.e. to a limit in $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$. However, by construction this limit is g_n. Thus $g_n \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, and we also have

$$\int g_n d\mu \leq 2 \int g\, d\mu.$$

The sequence $(g_n)_{n\geq 1}$ is nonincreasing and since by hypothesis (f_n) converges pointwise a.e. to f, the sequence (g_n) converges pointwise a.e. to 0. By the monotone convergence theorem (Theorem 5.31),

$$\lim_{n\mapsto\infty} \int g_n\, d\mu = 0.$$

Hence, by definition of g_n, the sequence (f_n) is actually an $\| \|_1$-Cauchy sequence, and by Proposition 5.28, we have $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ and (f_n) converges to f in the $\| \|_1$-norm. We have

$$\begin{aligned} \left\| \int f_n d\mu - \int f d\mu \right\| &= \left\| \int (f_n - f) d\mu \right\| \\ &\leq \int \|f_n - f\|\, d\mu \\ &= \|f_n - f\|_1, \end{aligned}$$

and since (f_n) converges to f in the $\| \|_1$-norm, this implies that

$$\lim_{n \mapsto \infty} \int f_n d\mu = \int f d\mu,$$

as claimed. $\square$

The first important application of Theorem 5.34 is to provide a characterization of the integrability of a function $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ in terms of $\int \|f\| \, d\mu$.

Theorem 5.35. *A function $f\colon X \to F$ is integrable, that is, $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, iff $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ and $\|f\| \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$. More generally, if $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ and if there is a function $g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ such that $g \geq 0$ and $\|f\| \leq g$, then $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$.*

Proof. By Proposition 5.21, the space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is a subspace of $\mathcal{M}_\mu(X, \mathcal{A}, F)$, and we already know from Proposition 5.24(4) that if $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ then $\|f\| \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$.

For the converse, we may assume that f and g are measurable, since a μ-measurable function is equal a.e. to a measurable function. There is a sequence $(h_n)_{n \geq 1}$ of μ-step maps that converges pointwise a.e. to f. For every $x \in X$ and every $n \geq 1$, let

$$h'_n(x) = \begin{cases} h_n(x) & \text{if } \|h_n(x)\| \leq 2g(x) \\ 0 & \text{if } \|h_n(x)\| > 2g(x). \end{cases}$$

For every $n \geq 1$, the function h'_n is a μ-step function and $\|h'_n\| \leq 2g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$. We claim that for every $x \in X$ such that $(h_n(x))$ converges to $f(x)$, the sequence $(h'_n(x))$ also converges to $f(x)$. If $g(x) = 0$, then $\|f(x)\| = 0$, so $f(x) = 0$, and then $h'_n(x) = 0$ for all $n \geq 1$. If $g(x) \neq 0$, since the sequence $(h_n(x))$ converges to $f(x)$ and since $\|f(x)\| < 2g(x)$, there is some $M > 0$ such that $\|h_n(x)\| \leq 2g(x)$ for all $n \geq M$, which implies that $h'_n(x) = h_n(x)$. It follows that the sequence $(h'_n)_{n \geq 1}$ converges pointwise a.e. to f. By Theorem 5.34, since $\|h'_n\| \leq 2g$, we conclude that $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$. $\square$

A useful corollary of Theorem 5.35 is the following result.

Proposition 5.36. *The following facts hold:*

(1) If $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, $g \in \mathcal{M}_\mu(X, \mathcal{A}, K)$ with $K = \mathbb{R}$ or $K = \mathbb{C}$, and $\|g\|$ is bounded, then $fg \in \mathcal{L}_\mu(X, \mathcal{A}, F)$.

(2) Let $h\colon E \times F \to G$ be a continuous bilinear map, where E, F, G are Banach spaces. If $f \in \mathcal{L}_\mu(X, \mathcal{A}, E)$ and $g \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ with $\|g\|$ bounded, then $h(f, g) \in \mathcal{L}_\mu(X, \mathcal{A}, G)$.

(3) Let $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, with $f \geq 0$, and let $g \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})$, with values in an interval $[m, M]$. Then $fg \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$ and we have

$$m \int f d\mu \leq \int g f d\mu \leq M \int f d\mu.$$

Another corollary involves series of functions in $\mathcal{L}_\mu(X, \mathcal{A}, F)$.

Proposition 5.37. *Let $(f_n)_{n\geq}$ be a sequence of functions $f_n \in \mathcal{L}_\mu(X, \mathcal{A}, F)$. If the series*

$$\sum_{n=1}^{\infty} \int \|f_n\| \, d\mu$$

converges, then the series

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

converges a.e., $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, and

$$\int f d\mu = \sum_{n=1}^{\infty} \int f_n \, d\mu$$

The following proposition is needed for the proof of several results stated in Chapter 7.

Proposition 5.38. *(Averaging Theorem) Let $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$ be any function and let S be any closed subset of F. If for any measurable subset A of finite measure $\mu(A) > 0$ we have*

$$\frac{1}{\mu(A)} \int_A f d\mu \in S,$$

and if $0 \in S$ or if X is σ-finite, then $f(x) \in S$ for all almost all $x \in X$.

Proposition 5.38 is proven in Lang [62] (Chapter VI, Theorem 5.15). By applying Proposition 5.38 to the set $S = \{0\}$, we obtain the following useful corollary.

Proposition 5.39. *For any function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, if*

$$\int_A f d\mu = 0$$

for every measurable subset A of finite measure, then $f = 0$ almost everywhere.

We conclude this section with two results about the continuity and the differentiability of a function defined by an integral.

Proposition 5.40. *Let $(X, \mathcal{A}, \mu)$ be a measure space, let U be metric space, let F be a Banach space (over $\mathbb{R}$ or $\mathbb{C}$), and let $f\colon U \times X \to F$ be a function.*

1. *(Continuity of the integral) Assume that f has the following properties:*

(a) For every $u \in U$, the map $f_{u,-} \colon X \to F$ given by

$$f_{u,-}(x) = f(u, x) \quad x \in X,$$

belongs to $\mathcal{L}_\mu(X, \mathcal{A}, F)$,

(b) For every $x \in X$, the map $f_{-,x} \colon U \to F$ given by

$$f_{-,x}(u) = f(u, x) \quad u \in U,$$

is continuous.

(c) There is some $g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, $g \geq 0$, such that

$$\|f(u, x)\| \leq g(x) \quad \text{for all } u \in U, \text{ and all } x \in X.$$

Then the map $h \colon U \to F$ given by

$$h(u) = \int f_{u,-}\, d\mu$$

is continuous.

2. *(Taking a derivative under the integral sign) Suppose U is an open subset of a Banach space G, and let $\mathcal{L}(G; F)$ be the space of linear continuous maps from G to F with the operator norm (see Definition A.50). Assume that f has the following properties:*

(d) For every $u \in U$, the map $f_{u,-} \colon X \to F$ given by

$$f_{u,-}(x) = f(u, x) \quad x \in X,$$

belongs to $\mathcal{L}_\mu(X, \mathcal{A}, F)$,

(e) For every $x \in X$, the map $f_{-,x} \colon U \to F$ is differentiable, and let $Df_{-,x}$ be this derivative (a map from U to $\mathcal{L}(G; F)$).

(f) For every $u \in U$, the map from X to $\mathcal{L}(G; F)$ given by

$$x \mapsto Df_{-,x}(u)$$

belongs to $\mathcal{L}_\mu(X, \mathcal{A}, \mathcal{L}(G; F))$, and there is some $g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, $g \geq 0$, such

$$\|Df_{-,x}(u)\| \leq g(x) \quad \text{for all } u \in U, \text{ and all } x \in X.$$

Then the map $h \colon U \to F$ given by

$$h(u) = \int f_{u,-}\, d\mu$$

is differentiable in U, and its derivative at $u \in U$ is given by

$$Dh_u = \int Df_{-,x}(u)\, d\mu.$$

More could be said about the applications of the convergence theorems, but we have everything we need.

Remark: There is another approach to the definition of the integral that applies only to real and complex-valued functions, presented in various texts such as Rudin [79]. In this approach, positive functions play a central role. This approach relies on the fact that for any measurable function $f\colon X \to [0, +\infty]$ there is a monotonic sequence (f_n) of positive step functions that converges pointwise to f; see Rudin [79] (Chapter 1, Theorem 1.17). The integral of a step function is defined in the usual way. Then given any *measurable* function $f\colon X \to [0, +\infty]$, the integral of f is defined as

$$\int f d\mu = \sup_{0 \leq s \leq f} \int s\, d\mu,$$

where s is a step function.

A main difference with the approach we followed is that this definition of the integral allows it to take the value $+\infty$. Of course, later on, in order to define what it means for a measurable complex-valued function $f\colon X \to \mathbb{C}$ to be integrable, the condition

$$\int |f| d\mu < +\infty$$

is required. Thus in this approach, the space $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$ is defined as the space of measurable functions such that the positive function $|f|$ has a finite integral.

In the approach that we followed, due to Bochner and Dunford, the space $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$ is defined in terms of various Cauchy sequences, and the fact that if a function $f\colon X \to \mathbb{C}$ is measurable and if $|f|$ has a finite integral, then $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$, is a *theorem* (Theorem 5.35). Ultimately, it is proved that $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$ is complete (see Rudin [79] Chapter 3, Theorem 3.11), and it is observed that as a corollary, from a $\| \, \|_1$-Cauchy sequence, one can extract a subsequence that converges pointwise a.e. (Rudin [79] Chapter 3, Theorem 3.12). It is also shown that the μ-step functions are dense in $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$ (Rudin [79] Chapter 3, Theorem 3.13).

The circle is closed. What we took as a definition of $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$ is obtained as a corollaries in the other approach, and the two approaches yield the same notion of integrability (the same space $\mathrm{L}_\mu(X, \mathcal{A}, \mathbb{C})$).

One might argue that the approach relying on the integral of positive functions is simpler, or at least takes less efforts. For one thing, it does not need the refined notion of μ-step maps and μ-measurable maps. However, our feeling is that the approach we followed provides a better understanding of the structure of $\mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$. Also, it can't be avoided if one wants to integrate functions with values in an infinite-dimensional vector space.

5.10 The Spaces $\mathcal{L}^p_\mu(X,\mathcal{A},F)$ and $\mathrm{L}^p_\mu(X,\mathcal{A},F)$; $p=1,2$

Theorem 5.35 suggests the definition of other families of integrable functions.

Definition 5.15. Let $(X,\mathcal{A},\mu)$ be a measure space and let $(F,\mathcal{B})$ be a measurable space, with F a Banach space and $\mathcal{B}$ its Borel σ-algebra. For any $p\geq 1$, the set of functions $\mathcal{L}^p_\mu(X,\mathcal{A},F)$ is the set of functions $f\in\mathcal{M}_\mu(X,\mathcal{A},F)$ such that $\|f\|^p\in\mathcal{L}_\mu(X,\mathcal{A},\mathbb{R})$, or equivalently

$$\int \|f\|^p\,d\mu<+\infty.$$

By Theorem 5.35, we have $\mathcal{L}^1_\mu(X,\mathcal{A},F)=\mathcal{L}_\mu(X,\mathcal{A},F)$, and we know that $\mathcal{L}^1_\mu(X,\mathcal{A},F)$ is a vector space. Although it is possible to develop a theory of $\mathcal{L}^p$ spaces for any $p\geq 1$, for our applications to harmonic analysis we only need the cases $p=1,2$. The case where $p=\infty$ arises when we consider duality, but we postpone the definition of $\mathcal{L}^\infty_\mu(X,\mathcal{A},F)$.

The space $\mathcal{L}^2_\mu(X,\mathcal{A},F)$ is particularly interesting because if F is a Hilbert space, then it can be given a Hilbert space structure which uses the inner product on F (not quite, because the Hermitian form $\langle -,-\rangle_\mu$ that we obtain is not positive definite, which means that $\langle f,f\rangle_\mu=0$ does necessarily imply that $f=0$).

Let us start with the simple case where $F=\mathbb{C}$. If $f\colon X\to\mathbb{C}$ is a complex-valued function, then by $|f|^2$ we mean the function defined such that

$$|f|^2(x)=f(x)\overline{f(x)}\quad\text{for all } x\in X.$$

For any two functions $f,g\colon X\to\mathbb{C}$, by $\langle f,g\rangle$ we mean the function defined such that

$$\langle f,g\rangle(x)=f(x)\overline{g(x)}\quad\text{for all } x\in X.$$

For the more general case where F is a Hilbert space with Hermitian inner product $\langle -,-\rangle_F$, for any two functions $f,g\colon X\to F$, then by $\langle f,g\rangle$ we mean the function defined such that

$$\langle f,g\rangle(x)=\langle f(x),g(x)\rangle_F\quad\text{for all } x\in X.$$

In particular, since $\langle f,f\rangle$ is the function given by $\langle f,f\rangle(x)=\langle f(x),g(x)\rangle_F$ and $\|f\|^2$ is the function given by

$$\|f\|^2(x)=\|f(x)\|^2=\langle f(x),f(x)\rangle_F,$$

we have

$$\|f\|^2=\langle f,f\rangle.$$

To simplify notation we will drop the subscript F when referring to the inner product on F.

From now on, *when dealing with* $\mathcal{L}^2_\mu(X,\mathcal{A},F)$, *we assume that* F *is a Hilbert space* (over $\mathbb{C}$). If the reader feels more comfortable, he/she may assume that $F=\mathbb{C}$, but significant simplifications do not arise.

Proposition 5.41. *The set $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$ is a vector space. For any two maps $f, g \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$, we have $\langle f, g\rangle \in \mathcal{L}^1_\mu(X, \mathcal{A}, \mathbb{C})$, and the map*

$$(f, g) \mapsto \int \langle f, g\rangle d\mu$$

is a Hermitian positive map (not necessarily positive definite).

Proof. It is easy to see that $\langle f, g\rangle$ is a limit of step maps a.e., so $\langle f, g\rangle \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{C})$, and by the Cauchy–Schwarz inequality, we have the standard inequality

$$2|\langle f, g\rangle| \leq \|f\|^2 + \|g\|^2,$$

with $\|f\|^2 + \|g\|^2 \in \mathcal{L}^1_\mu(X, \mathcal{A}, \mathbb{R})$ since $f, g \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$. By Theorem 5.35, we have $\langle f, g\rangle \in \mathcal{L}^1_\mu(X, \mathcal{A}, \mathbb{C})$, so $\int\langle f, g\rangle d\mu$ is well defined. If $f \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$ and $g \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$, then since

$$\|f + g\|^2 \leq \|f\|^2 + 2|\langle f, g\rangle| + \|g\|^2,$$

as all the functions on the right-hand side are in $\mathcal{L}^1_\mu(X, \mathcal{A}, \mathbb{R})$, we have $f + g \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$. For any $\lambda \in \mathbb{C}$, we have $|\lambda|^2 \|f\|^2 \in \mathcal{L}^1_\mu(X, \mathcal{A}, \mathbb{R})$, so $\lambda f \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$. Thus, $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$ is a vector space. Using the linearity of the integral, it is easy to check that the map

$$(f, g) \mapsto \int \langle f, g\rangle d\mu$$

is a Hermitian positive map. □

Definition 5.16. For any two functions $f, g \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$, the Hermitian map $\langle f, g\rangle_\mu$ is defined by

$$\langle f, g\rangle_\mu = \int \langle f, g\rangle d\mu$$

The L^2-semi-norm $\|f\|_2$ is given by

$$\|f\|_2 = \sqrt{\langle f, f\rangle_\mu} = \left(\int \langle f, f\rangle d\mu\right)^{1/2} = \left(\int \|f\|^2 \, d\mu\right)^{1/2}.$$

It is a standard result of linear algebra that the *Cauchy–Schwarz* inequality holds:

$$|\langle f, g\rangle_\mu| \leq \|f\|_2 \|g\|_2.$$

As a consequence $\| \, \|_2$ is a semi-norm.

Proposition 5.42. *For any $f \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$, we have $\|f\|_2 = 0$ iff $f = 0$ a.e.*

Proof. If $f = 0$ a.e., then $\langle f, f\rangle = 0$ a.e., so $\|f\|_2^2 = \int\langle f, f\rangle d\mu = 0$. Conversely, if $\|f\|_2 = 0$, then this means that $\int\langle f, f\rangle d\mu = 0$, but $\langle f, f\rangle \in \mathcal{L}^1_\mu(X, \mathcal{A}, \mathbb{R})$ is a positive function, so we know from Proposition 5.27 that $\langle f, f\rangle = 0$ a.e., that is, $f = 0$ a.e. □

If X has finite measure, then $\mathcal{L}^2_\mu(X,\mathcal{A},F)$ is contained in $\mathcal{L}^1_\mu(X,\mathcal{A},F)$.

Proposition 5.43. *If X has finite measure, then for any $f\in\mathcal{L}^2_\mu(X,\mathcal{A},F)$, we have $\|f\|_1\le \|f\|_2\,\|1_X\|_2$, and $\mathcal{L}^2_\mu(X,\mathcal{A},F)$ is contained in $\mathcal{L}^1_\mu(X,\mathcal{A},F)$.*

Proof. The function $\|f\|$ (namely $x\mapsto \|f(x)\|$) is complex-valued so we can apply the Cauchy–Schwarz inequality to $\|f\|$ and to the constant function 1_X equal to 1 on X. To be more specific, since

$$\langle \|f\|\,, 1_X\rangle = \|f\|\,\overline{1_X} = \|f\|\,,$$

we have

$$\langle \|f\|\,, 1_X\rangle_\mu = \int \langle \|f\|\,, 1_X\rangle d\mu = \int \|f\|\,d\mu = \|f\|_1\,,$$

and

$$\|\|f\|\|_2^2 = \langle \|f\|\,, \|f\|\rangle_\mu = \int \langle \|f\|\,, \|f\|\rangle d\mu = \int \|f\|^2\,d\mu = \|f\|_2^2\,,$$

and so the Cauchy–Schwarz inequality (for functions in $\mathcal{L}^2_\mu(X,\mathcal{A},\mathbb{C})$)

$$\langle \|f\|\,, 1_X\rangle_\mu \le \|f\|_2\,\|1_X\|_2$$

implies that $\|f\|_1\le\|f\|_2\,\|1_X\|_2$. Obviously, this inequality shows that $\mathcal{L}^2_\mu(X,\mathcal{A},F)$ is contained in $\mathcal{L}^1_\mu(X,\mathcal{A},F)$. $\square$

It should be noted that if X has finite measure then the inclusion can be strict, and if X has infinite measure, then in general there are no inclusion properties.

Example 5.4.

1. If $X=(0,1)$, with the Lebesgue measure, then $\frac{1}{\sqrt{x}}\in\mathcal{L}^1((0,1),\mu_L)$ but $\frac{1}{\sqrt{x}}\notin\mathcal{L}^2((0,1),\mu_L)$; see Figure 5.18.

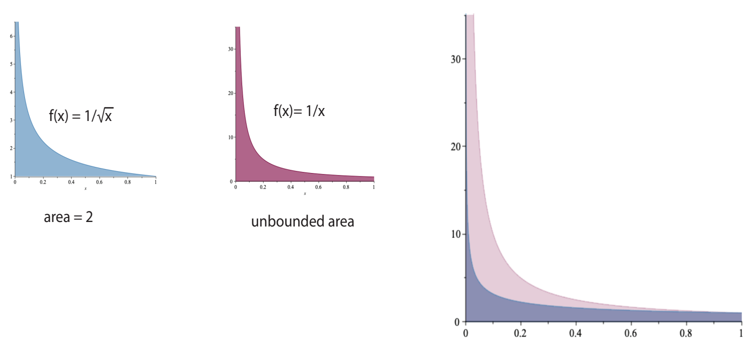


Figure 5.18: The first figure shows that $\frac{1}{\sqrt{x}}\in\mathcal{L}^1((0,1),\mu_L)$ since the area between f and the x-axis has finite value, while the second figure shows that $\frac{1}{\sqrt{x}}\notin\mathcal{L}^2((0,1),\mu_L)$. The third figure shows a direct comparison between the areas under the respective graphs.

2. If $X = (1, \infty)$ with the Lebesgue measure, then $\frac{1}{x} \in \mathcal{L}^2((1,\infty), \mu_L)$ but $\frac{1}{x} \notin \mathcal{L}^1((1,\infty), \mu_L)$; see Figure 5.19.

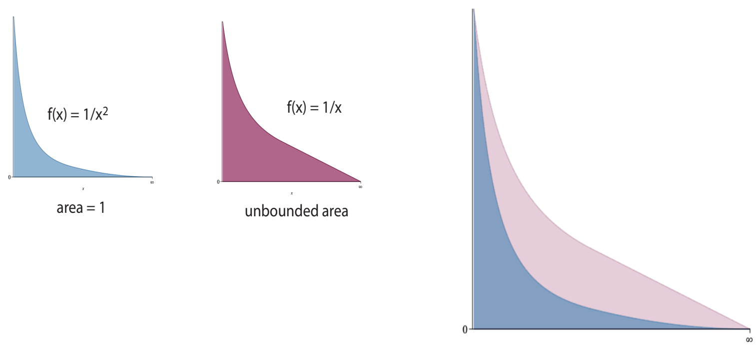


Figure 5.19: The first figure shows that $\frac{1}{x} \in \mathcal{L}^2((1,\infty), \mu_L)$ since the area between f^2 and the x-axis has finite value, while the second figure shows that $\frac{1}{x} \notin \mathcal{L}^1((1,\infty), \mu_L)$. The third figure shows a direct comparison between the areas under the respective graphs.

3. If $X = (0, \infty)$ with the Lebesgue measure, then $\frac{1}{(x+1)\sqrt{x}} \in \mathcal{L}^1((0,\infty), \mu_L)$ but $\frac{1}{(x+1)\sqrt{x}} \notin \mathcal{L}^2((0,\infty), \mu_L)$; see Figure 5.20.

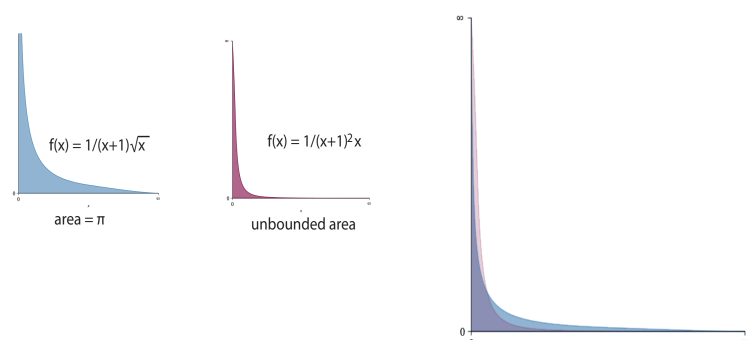


Figure 5.20: The first figure shows that $\frac{1}{(x+1)\sqrt{x}} \in \mathcal{L}^1((0,\infty), \mu_L)$s since the area between f and the x-axis has finite value, while the second figure shows that $\frac{1}{(x+1)\sqrt{x}} \notin \mathcal{L}^2((0,\infty), \mu_L)$. The third figure shows a direct comparison between the areas under the respective graphs.

One of the main properties of $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$ is that it is complete for the semi-norm $\| \, \|_2$. By taking the quotient of $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$ by the space of function equal to 0 a.e., we obtain a Hilbert space.

Theorem 5.44. *Let $(f_n)_{n \geq 1}$ be an $\| \, \|_2$-Cauchy sequence of functions $f_n \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$. Then there is a function $f \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$ with the following properties:*

1. *The sequence $(f_n)_{n\geq 1}$ converges to f in the $\| \, \|_2$-semi-norm. Thus $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$ is complete.*

   There is a subsequence $(f_{n_k})_{k\geq 1}$ of $(f_n)_{n\geq 1}$ with the following properties:

2. *The subsequence $(f_{n_k})_{k\geq 1}$ converges pointwise a.e. to f.*
3. *For every $\epsilon > 0$, there is a subset Z such that $\mu(Z) < \epsilon$ and the subsequence $(f_{n_k})_{k\geq 1}$ converges uniformly to f on $X - Z$.*

A proof of Theorem 5.44 is given in Lang [62] (Chapter VII, §1).

In view of Proposition 5.42, we make the following definition.

Definition 5.17. Let $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$ be the quotient of the vector space $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$ by the subspace of functions equal to 0 a.e. (which is the set of functions f such that $\|f\|_2 = 0$). The norm induced by the semi-norm $\| \, \|_2$ on $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$ is called the *L^2-norm*.

Obviously the positive Hermitian form $\langle f, g\rangle_\mu$ induces a positive definite Hermitian form on $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$. Theorem 5.44 immediately implies the following result.

Theorem 5.45. *(Fischer–Riesz) The space $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$ is a Hilbert space under the positive definite Hermitian form induced by $\langle -, -\rangle_\mu$.*

Definition 5.18. The norm $\| \, \|_2$ associated with the inner product $\langle -, -\rangle_\mu$ on $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$ is called the *L^2-norm*.

Example 5.5. In the special case where $X = \mathbb{N}$ (or $X = \mathbb{Z}$), $\mathcal{A} = 2^X$, μ is the counting measure, and $F = \mathbb{C}$, as in Example 5.2, we see that for $p = 1, 2$, we have

$$\mathrm{L}^p_\mu(X, \mathcal{A}, \mathbb{C}) = \{(x_n)_{n\in\mathbb{N}} \mid x_n \in \mathbb{C}, \sum_{n\in\mathbb{N}} |x_n|^p < \infty\}.$$

It is customary to denote this space by $\ell^p(\mathbb{N})$. We define $\ell^p(\mathbb{Z})$ similarly by replacing $\mathbb{N}$ by $\mathbb{Z}$.

We will show shortly that the space of μ-step functions is dense in $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$ (for the L^2-norm). First here is a corollary of Theorem 5.45.

Proposition 5.46. *If $(f_n)_{n\geq 1}$ is a $\| \, \|_2$-Cauchy sequence of functions $f_n \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$, and if $(f_n)_{n\geq 1}$ converges pointwise a.e. to a function $f \colon X \to F$, then $f \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$, and $(f_n)_{n\geq 1}$ converges to f in the $\| \, \|_2$-semi-norm.*

The Lebesgue dominated convergence theorem also holds for $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$.

Theorem 5.47. *(Lebesgue Dominated Convergence Theorem for $\mathcal{L}^2_\mu$) Let $(f_n)_{n\geq 1}$ be a sequence of functions $f_n \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$. If (f_n) converges pointwise a.e. to a function $f\colon X \to F$, and if there is some function $g \in \mathcal{L}^2_\mu(X, \mathcal{A}, \mathbb{R})$ such that $g \geq 0$ and $\|f_n\| \leq g$ for all $n \geq 1$, then $f \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$ and $(f_n)_{n\geq 1}$ converges to f in the $\|\,\|_2$-norm. Consequently*

$$\lim_{n\mapsto\infty} \int f_n d\mu = \int f d\mu.$$

A proof of Theorem 5.47 is given in Lang [62] (Chapter VII, §1).

The following version of Theorem 5.35 also holds for $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$.

Theorem 5.48. *A function $f\colon X \to F$ is $\mathcal{L}^2$-integrable, that is, $f \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$, iff $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ and $\|f\|_2 \in \mathcal{L}^2_\mu(X, \mathcal{A}, \mathbb{R})$.*

As a corollary of Theorem 5.47 we can show that the μ-step functions are dense in $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$.

Proposition 5.49. *The subspace $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ is dense in $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$ with respect to the L^2-semi-norm.*

Proof. Let $f \in \mathcal{L}^2_\mu(X, \mathcal{A}, F)$. Since f is μ-measurable, there is a sequence $(f_n)_{n\geq 1}$ of μ-step functions f_n that converges pointwise a.e. to f. For every $n \geq 1$ and every $x \in X$, define g_n by

$$g_n(x) = \begin{cases} f_n(x) & \text{if } \|f_n(x)\| \leq 2\,\|f(x)\| \\ 0 & \text{if } \|f_n(x)\| > 2\,\|f(x)\|. \end{cases}$$

We may assume that f is measurable since it differs from a measurable function on a set of measure zero. Then the functions g_n are μ-step functions, they satisfy the inequality $\|g_n\| \leq 2\,\|f\|$ with $2\,\|f\| \in \mathcal{L}^2_\mu(X, \mathcal{A}, \mathbb{R})$, and the sequence (g_n) converges a.e. to f. By Theorem 5.47, the sequence (g_n) converges to f in the $\|\,\|_2$-norm, which proves that $\mathcal{S}tep_\mu(X, \mathcal{A}, F)$ is dense in $\mathcal{L}^2_\mu(X, \mathcal{A}, F)$ with respect to the L^2-semi-norm ☐

We now would like to understand the duals of $\mathrm{L}^1_\mu(X, \mathcal{A}, F)$ and $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$, that is, the spaces of continuous linear forms on $\mathrm{L}^1_\mu(X, \mathcal{A}, F)$ and $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$ (with values in $\mathbb{C}$). In the case of $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$, it is a classical theorem (the *Riesz respresentation theorem*) that the dual of a Hilbert space is isomorphic to itself, so the dual of $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$ is isomorphic to $\mathrm{L}^2_\mu(X, \mathcal{A}, F)$. In the case of $\mathrm{L}^1_\mu(X, \mathcal{A}, F)$, it turns out that its dual is isomorphic to a space denoted $\mathrm{L}^\infty_\mu(X, \mathcal{A}, F)$. Here we assumed that F is a Hilbert space.

5.11 The Spaces $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$ and $\mathrm{L}^\infty_\mu(X, \mathcal{A}, F)$

To define $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$, we only need the fact that F is a Banach space. The space $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$ consists of all functions $f\colon X \to F$ that are equal to a bounded μ-measurable function a.e. We can define a semi-norm on $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$ as follows.

Definition 5.19. For any function $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$, define the *essential sup* or *semi-norm* $N_\infty(f)$ of f by

$$N_\infty(f) = \inf\{\alpha \in \mathbb{R}_+ \mid \mu(\{x \in X \mid \|f(x)\| \geq \alpha\}) = 0\};$$

see Figure 5.21. The space $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$ is the set of functions $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ such that $N_\infty(f) < +\infty$.

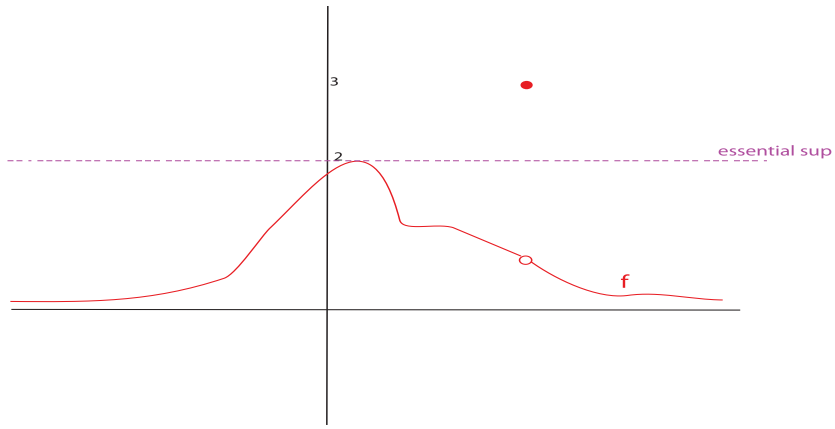


Figure 5.21: Let $X = F = \mathbb{R}$ with absolute value as norm and $\mathcal{A}$ as the Borel σ-algebra. The graph of $f \in \mathcal{M}_\mu(X, \mathcal{A}, F)$ is in red and has essential sup $N_\infty(f) = 2$. Note this is not the same as the sup norm for $f \in (F^X)_b$, which in this particular case is $\|f\|_\infty = 3$.

Remark: We decided to use the notation $N_\infty(f)$ for the essential sup semi-norm to avoid the confusion with the sup norm, $\|f\|_\infty$, since these norms differ in general. In the case of the semi-norms $\|f\|_1$ and $\|f\|_2$ there is little risk of confusion. A number of authors prefer the notation $N_p(f)$, but the notation $\| \|_p$ seems more prevalent (if $1 \leq p < \infty$). Another way to avoid confusion is to use the notation $\| \|_{\mathrm{L}^p}$ (even if $p = \infty$).

The definition of $N_\infty(f)$ makes is clear that $N_\infty(f) = 0$ iff $f = 0$ a.e. Observe that $N_\infty(f)$ is the greastest lower bound of the numbers $\alpha \geq 0$ such that a μ-measurable function f has the property that $\|f(x)\| \geq \alpha$ on a set of measure zero, in other words, such a μ-measurable function is bounded a.e.

The space $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$ is a vector space. We also have the following result showing that $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$ is complete in the semi-norm N_∞, but unless X has finite measure, the μ-step maps are *not* dense in $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$.

Theorem 5.50. *The following properties hold.*

1. *The space $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$ is complete in the semi-norm N_∞. Furthermore, if $(f_n)_{n\geq 1}$ is an N_∞-Cauchy sequence, then there is a set Z of measure zero such that $(f_n)_{n\geq 1}$ converges uniformly to f on $X - Z$.*

2. *If F is finite-dimensional, then the step maps (not the μ-step maps) are dense in $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$.*

3. *If X has finite measure, then for every $\epsilon > 0$ and every $f \in \mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$, there is a μ-step map s and a subset Z with $\mu(Z) < \epsilon$ such that*

$$\|f - s\| < \epsilon \quad on \quad X - Z.$$

Note that the constant with value 1 belongs to $\mathcal{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C})$, so if X has infinite measure, there is no way that it is a uniform limit of μ-step maps, since a μ-step map vanishes outside of a set of finite measure.

Remark: If X has finite measure, then we have the inclusion $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F) \subseteq \mathcal{L}^2_\mu(X, \mathcal{A}, F)$. In fact, $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F) \subseteq \mathcal{L}^p_\mu(X, \mathcal{A}, F) \subseteq \mathcal{L}^q_\mu(X, \mathcal{A}, F)$ for all $p, q \geq 1$ with $p > q$; see Marle [69] (Chapter 4, Proposition 4.5.7).

Definition 5.20. Let $\mathrm{L}^\infty_\mu(X, \mathcal{A}, F)$ be the quotient of the vector space $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$ by the subspace of functions equal to 0 a.e. (which is the set of functions f such that $N_\infty(f) = 0$). The norm induced by the semi-norm N_∞ on $\mathrm{L}^\infty_\mu(X, \mathcal{A}, F)$ is called the *L^∞-norm*.

It should be noted that both the monotone convergence theorem and the dominated convergence theorem *fail* for $\mathcal{L}^\infty_\mu(X, \mathcal{A}, F)$. Convergence in the N_∞-semi-norm fails; see Marle [69] (Chapter 4, Section 2).

We now consider the duality between the spaces $\mathrm{L}^1_\mu(X, \mathcal{A}, F)$ and $\mathrm{L}^\infty_\mu(X, \mathcal{A}, F)$. The field $\mathbb{C}$ is a Hilbert space, but for the general case we need to assume that F is a Hilbert space. The key point is that by Proposition 5.36(2), for any $f \in \mathrm{L}^1_\mu(X, \mathcal{A}, F)$ and any $g \in \mathrm{L}^\infty_\mu(X, \mathcal{A}, F)$, then $\langle f, g\rangle \in \mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C})$.

Definition 5.21. For any functions $f \in \mathrm{L}^1_\mu(X, \mathcal{A}, F)$ and $g \in \mathrm{L}^\infty_\mu(X, \mathcal{A}, F)$, define $[f, g]_\mu$ by

$$[f, g]_\mu = \int \langle f, g\rangle d\mu.$$

We obtain a map

$$[-, -]_\mu \colon \mathrm{L}^1_\mu(X, \mathcal{A}, F) \times \mathrm{L}^\infty_\mu(X, \mathcal{A}, F) \to \mathbb{C}$$

which is a sesquilinear pairing.

For simplicity, let us consider the special case where $F = \mathbb{C}$. In this case, we can define a bilinear (as opposed to sesquilinear) pairing $[-, -]_\mu \colon \mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C}) \times \mathrm{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C}) \to \mathbb{C}$ given by

$$[f, g]_\mu = \int f g \, d\mu.$$

Observe that we intentionally used fg instead of $f\overline{g}$, because we simply want a *bilinear* pairing.

Whenever we have a bilinear pairing $\varphi\colon E \times F \to \mathbb{C}$, recall that we define the linear maps $l_\varphi\colon E \to F^*$ and $r_\varphi\colon F \to E^*$ such that, for every $u \in E$,

$$l_\varphi(u)(y) = \varphi(u, y) \quad \text{for all } y \in F,$$

and for every $v \in F$,

$$r_\varphi(v)(x) = \varphi(x, v) \quad \text{for all } x \in E.$$

Definition 5.22. A bilinear pairing φ is *nondegenerate* if for every $u \in E$, if $\varphi(u, v) = 0$ for all $v \in F$, then $u = 0$, and for every $v \in F$, if $\varphi(u, v) = 0$ for all $u \in E$, then $v = 0$.

Then if φ is nondegenerate, then the maps l_φ and r_φ are injective. They are not surjective in general.

If E is a normed vector space, then its dual E' is the space of all continuous linear maps from E to $\mathbb{C}$. We have $E' \subseteq E^*$, and the inclusion is strict if E is infinite-dimensional.

The following result holds. For simplicity of notation, we drop φ when writing l_φ and r_φ.

Theorem 5.51. *Assume $(X, \mathcal{A}, \mu)$ is a measure space and that μ is σ-finite. Then the bilinear pairing*

$$[-,-]_\mu\colon \mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C}) \times \mathrm{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C}) \to \mathbb{C}$$

is nondegenerate. It satisfies the inequality

$$|[f, g]_\mu| \leq \|fg\|_1 \leq \|f\|_1 \|g\|_\infty .$$

The map l is a norm-preserving injective linear map between $\mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C})$ and the dual $\mathrm{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C})'$ of $\mathrm{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C})$, and the map r is a norm-preserving injective linear map between $\mathrm{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C})$ and the dual $\mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C})'$ of $\mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C})$. Furthermore, the map $r\colon \mathrm{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C}) \to \mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C})'$ is an isomorphism.

A proof of Theorem 5.51 is given in Lang [62] (Chapter VII, §2). Theorem 5.51 can be generalized to a Hilbert space F, one just has to exercise caution in defining l and r to deal with sequilinearity.

The map $l\colon \mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C}) \to \mathrm{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C})'$ is not surjective, and understanding which linear forms in $\mathrm{L}^\infty_\mu(X, \mathcal{A}, \mathbb{C})'$ can be represented by functions in $\mathrm{L}^1_\mu(X, \mathcal{A}, \mathbb{C})$ is a natural question. A partial answer to this question is the *Radon–Nikodym theorem*, but will this would lead us too far. The interested reader is referred to Lang [62] or Rudin [79].

5.12 Products of Measure Spaces and Fubini's Theorem

The purpose of this section is to define, given two measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$, the notion of product measure space and product measure. Then we will state *Fubini's theorem* (also known as *the theorem of Lebesgue–Fubini*), which allows us to compute the integral on a product space as two successive integrals. The technical details are surprisingly involved.

We begin by recalling what we did in Example 4.1. We defined the set $\mathcal{R}$ of *rectangles* in $X \times Y$ as follows:

$$\mathcal{R} = \{A \times B \in X \times Y \mid A \in \mathcal{A},\ B \in \mathcal{B}\}.$$

The set $\mathcal{R}$ is a semi-algebra, and it can be shown that the set $\mathcal{B}(\mathcal{R})$ of finite unions of pairwise disjoint sets in $\mathcal{R}$ is the smallest algebra containing the semi-algebra $\mathcal{R}$.

Definition 5.23. Let $\mathcal{A} \otimes \mathcal{B}$ be the smallest σ-algebra generated by $\mathcal{R}$ (and thus by $\mathcal{B}(\mathcal{R})$); see Proposition 4.3.[4]

The hard part is now to define a product measure λ on $\mathcal{A} \otimes \mathcal{B}$ which satisfies the natural identity

$$\lambda(A \times B) = \mu(A)\nu(B)$$

for all rectangles $A \times B$. Here as in Section 4.1 we use extended multiplication on $\overline{\mathbb{R}}_+$, where

$$a \cdot (+\infty) = (+\infty) \cdot a = +\infty$$

if $0 < a \leq +\infty$, and

$$0 \cdot (+\infty) = (+\infty) \cdot 0 = 0.$$

We need a few definitions.

Definition 5.24. Given any subset $E \subseteq X \times Y$, for any $x \in X$, we define the *section of E (determined by x)* as the subset E_x given by

$$E_x = \{y \in Y \mid (x, y) \in E\} \subseteq Y.$$

Similarly, for any $y \in Y$, we define the *section of E (determined by y)* as the subset E_y given by

$$E_y = \{x \in X \mid (x, y) \in E\} \subseteq X;$$

see Figure 5.22.

Proposition 5.52. *The sections of any subset $E \in \mathcal{A} \otimes \mathcal{B}$ are measurable.*

[4] The meaning of the tensor sign $\otimes$ in the notation $\mathcal{A} \otimes \mathcal{B}$ is a completely different from its meaning in a tensor product of vector spaces. Hopefully, the two notions will never appear together!

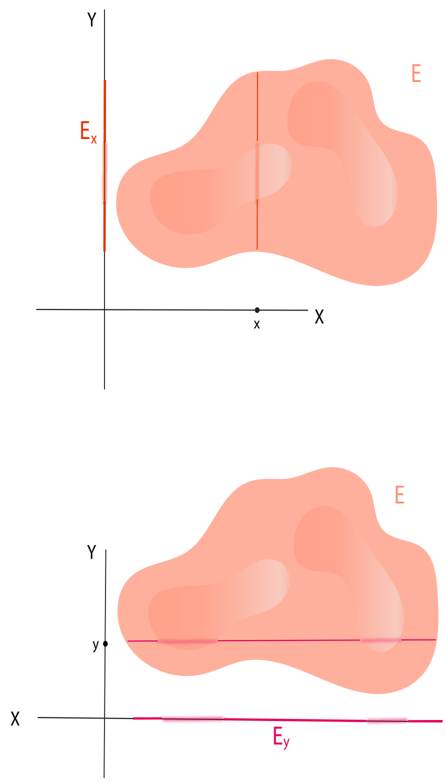


Figure 5.22: Let $X = Y = \mathbb{R}$. The top figure illustrates an x-section of the peach set E, while the bottom figure illustrates a y-section.

Proof idea. Let $\mathcal{E}$ be the family of subsets of $X \times Y$ defined as follows:

$$\mathcal{E} = \{F \subseteq X \times Y \mid F_x \in \mathcal{B} \quad \text{for all } x \in X, \quad \text{and} \quad F_y \in \mathcal{A} \quad \text{for all } y \in Y\}.$$

These are the subsets of $X \times Y$ whose sections are measurable. Then prove that $\mathcal{E}$ is a σ-algebra containing $\mathcal{R}$, which implies that $\mathcal{E} = \mathcal{A} \otimes \mathcal{B}$. ☐

Definition 5.25. Given any function $f \colon X \times Y \to F$ (where F is any set), for any $x \in X$, we define the *section of f (determined by x)* as the function $f_x \colon Y \to F$ given by

$$f_x(y) = f(x, y) \quad \text{for all } y \in Y.$$

Similarly, for any $y \in Y$, we define the *section of f (determined by y)* as the function $f_y \colon X \to F$ given by

$$f_y(x) = f(x, y) \quad \text{for all } x \in X.$$

See Figure 5.23.

Proposition 5.53. *If $f \colon X \times Y \to \mathbb{R}$ is a measurable function (on $(X \times Y, \mathcal{A} \otimes \mathcal{B})$), then every section of f is measurable.*

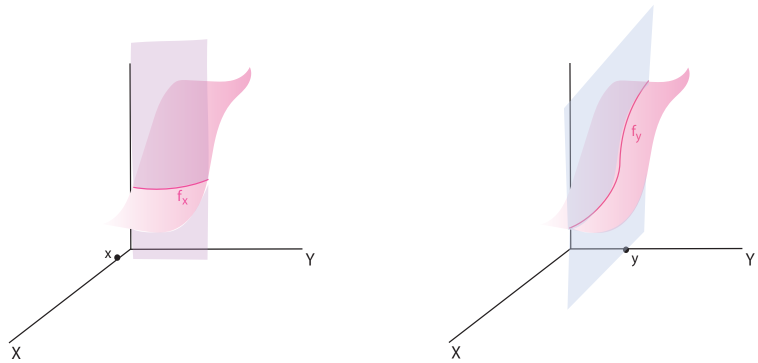


Figure 5.23: Let $X = Y = F = \mathbb{R}$. The graph of f is the pink surface. The left figure illustrates a section of f determined by x, while the right figure illustrates a section of f determined by y

Proof. By Proposition 4.12, it suffices to show that the inverse image of every open subset of the form $(-\infty, \alpha)$ is measurable.

For any $x \in X$, for any $\alpha \in \mathbb{R}$, we have

$$\begin{aligned}\{y \in Y \mid f_x(y) < \alpha\} &= \{y \in Y \mid f(x, y) < \alpha\}\\ &= \{(x, y) \in X \times Y \mid f(x, y) < \alpha\}_x,\end{aligned}$$

and this last subset is measurable by Proposition 5.52. The proof for f_y is similar. $\square$

Definition 5.26. Given an algebra $\mathfrak{A}$ of sets, a *measure* on $\mathfrak{A}$ satisfies the same axioms as a measure on a σ-algebra; see Definition 4.9.

The next two results take a lot more work.

Proposition 5.54. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two measure spaces, and assume that μ and ν are σ-finite. Then the map $\lambda \colon \mathcal{R} \to [0, +\infty]$ given by*

$$\lambda(A \times B) = \mu(A)\nu(B)$$

has a unique extension to a σ-finite measure on the algebra $\mathcal{B}(\mathcal{R})$.

A proof of Proposition 5.54 can be found in course notes given by Philippe G. Ciarlet in 1970-1971 at ENPC (Paris, France). Interestingly, the proof uses the monotone convergence theorem. A related treatment is given in Halmos [44] (Chapter VII); see also Lang [62] (Chapter VI, §8) and Marle [69] (Chapter 5, Section 2).

Theorem 5.55. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two measure spaces, and assume that μ and ν are σ-finite. Then the map $\lambda\colon \mathcal{R} \to [0, +\infty]$ given by*

$$\lambda(A \times B) = \mu(A)\nu(B)$$

has a unique extension to a measure $\lambda = \mu \otimes \nu$ is on the σ-algebra $\mathcal{A} \otimes \mathcal{B}$. The measure $\mu \otimes \nu$ is σ-finite.

The following properties hold for any measurable subset $E \in \mathcal{A} \times \mathcal{B}$:

(1) We have

$$(\mu \otimes \nu)(E) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i)\nu(B_i) \mid E \subseteq \bigcup_{i=1}^{\infty} (A_i \times B_i),\ A_i \in \mathcal{A},\ B_i \in \mathcal{B} \right\}. \qquad (*)$$

See Figure 5.24.

(2) The map ν_E from X to $\mathbb{R}_+$ given by $x \mapsto \nu(E_x)$ is measurable (w.r.t. $\mathcal{A}$), and the map μ_E from Y to $\mathbb{R}_+$ given by $y \mapsto \mu(E_y)$ is measurable (w.r.t. $\mathcal{B}$). One of these maps is integrable iff the other is integrable.

(3) We have

$$(\mu \otimes \nu)(E) = \begin{cases} \int \nu_E \, d\mu = \int \mu_E \, d\nu & \text{if both } \nu_E \text{ and } \mu_E \text{ are integrable} \\ +\infty & \text{otherwise.} \end{cases} \qquad (**)$$

See Figure 5.25.

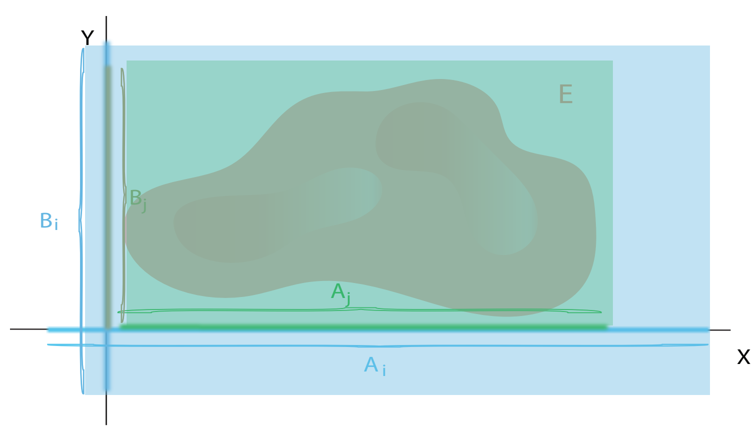


Figure 5.24: A schematic illustration of Equation $(*)$ in Theorem 5.55, where the measure of the peach set E is calculated by the "area" of the rectangles.

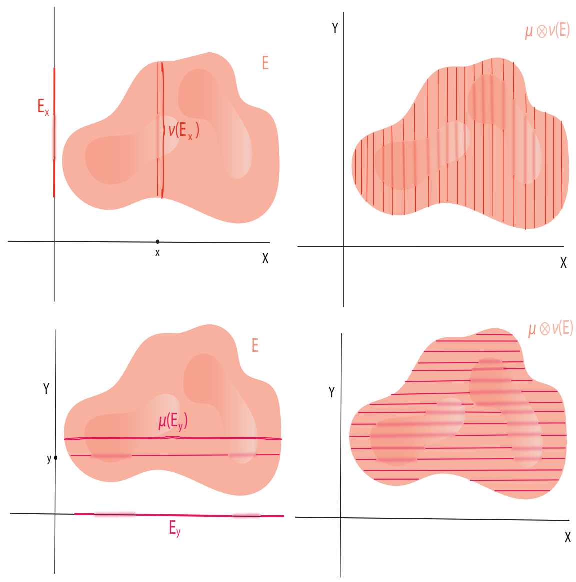


Figure 5.25: Two ways of calculating $(\mu \otimes \nu)(E)$. The top row of figures illustrates $(\mu \otimes \nu)(E) = \int \nu_E \, d\mu$, where the vertical slices represent $\nu(E_x)$. The bottom row of figures illustrates $(\mu \otimes \nu)(E) = \int \mu_E \, d\nu$, where the horizontal slices represent $\mu(E_y)$.

A proof of Theorem 5.55 can be found in course notes given by Philippe G. Ciarlet in 1970-1971 at ENPC (Paris, France). Again, the proof uses the monotone convergence theorem. A related treatment is given in Halmos [44] (Chapter VII); see also Lang [62] (Chapter VI, §8) and Marle [69] (Chapter 5, Section 2, Proposition 5.2.3).

If $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are two measure spaces with μ and ν both σ-finite, then for any Banach space F, we have the space of integrable functions $\mathcal{L}_{\mu\otimes\nu}(X \times Y, \mathcal{A} \otimes \mathcal{B}, F)$. The problem is to find a way to compute an integral $\iint f \, d(\mu \otimes \nu)$, also written $\iint f \, d\mu \otimes d\nu$, as two successive integrals. The answer is given by a theorem known as Fubini's theorem. The first version of this theorem was proved by Lebesgue and then was generalized by Fubini. For this reason some authors refer to this theorem as the Lebesgue–Fubini theorem, but it seems more common to call it simply Fubini's theorem.

Theorem 5.56. *(Fubini's Theorem, Part 1) Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two measure spaces with μ and ν both σ-finite. Consider a function $f \colon X \times Y \to F$, where F is a Banach space. If $f \in \mathcal{L}_{\mu\otimes\nu}(X \times Y, \mathcal{A} \otimes \mathcal{B}, F)$ then:*

1. *The section $f_x \colon Y \to F$ is ν-integrable for almost all $x \in X$, the section $f_y \colon X \to F$ is μ-integrable for almost all $y \in Y$.*

2. *The map from X to F defined a.e. by*

$$x \mapsto \int f_x \, d\nu$$

is μ-integrable, and the map from Y to F defined a.e. by

$$y \mapsto \int f_y \, d\mu$$

is ν-integrable.

Then

$$\iint f \, d\mu \otimes d\nu = \int_X \left(\int_Y f_x \, d\nu \right) d\mu = \int_Y \left(\int_X f_y \, d\mu \right) d\nu;$$

see Figure 5.26.

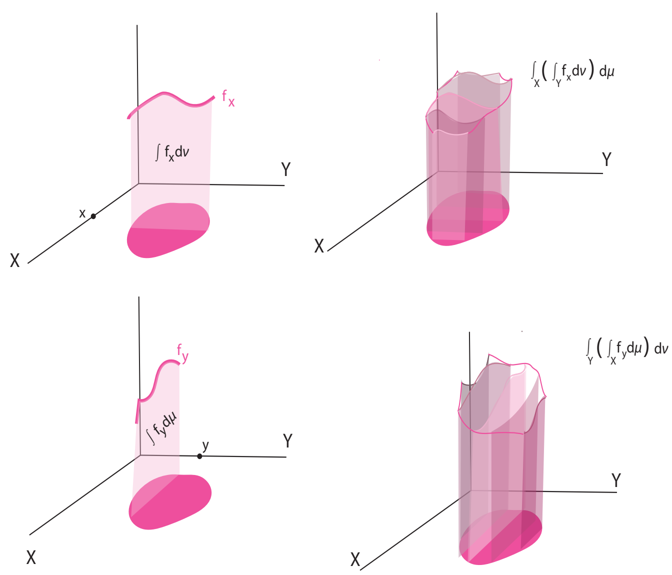


Figure 5.26: Two ways of calculating $\iint f d\mu \otimes d\nu$ when $F = \mathbb{R}$. The top row of figures illustrates $\iint f d\mu \otimes d\nu = \int_X \left(\int_Y f_x \, d\nu \right) d\mu$ as the "volume" under the graph calculated by the the stacked "areas" of $\int_Y f_x \, d\nu$ "sheets". The bottom row of figures illustrates $\iint f d\mu \otimes d\nu = \int_Y \left(\int_X f_y \, d\mu \right) d\nu$ as the "volume" under the graph calculated by the stacked "areas" of $\int_X f_y \, d\mu$ "sheets".

Theorem 5.56 is proved in Marle [69] (Chapter 5, Section 2, Theorem 5.2.10), and Lang [62] (Chapter VI, §8, Theorem 8.4).

Theorem 5.56 assumes that f is *integrable*. It is possible to weaken this assumption at the price of strengthening the other conditions. However, this is worth it in practice.

Theorem 5.57. *(Fubini's Theorem, Part 2) Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two measure spaces with μ and ν both σ-finite. Consider a function $f \colon X \times Y \to F$, where F is a Banach space. If $f \in \mathcal{M}_{\mu \otimes \nu}(X \times Y, \mathcal{A} \otimes \mathcal{B}, F)$ and if the following conditions hold:*

1. *The section $f_x\colon Y \to F$ is ν-integrable for almost all $x \in X$, the section $f_y\colon X \to F$ is μ-integrable for almost all $y \in Y$.*

2. *The map from X to $\mathbb{R}$ defined a.e. by*

$$x \mapsto \int \|f_x\|\, d\nu$$

is μ-integrable, and the map from Y to $\mathbb{R}$ defined a.e. by

$$y \mapsto \int \|f_y\|\, d\mu$$

is ν-integrable.

Then $f \in \mathcal{L}_{\mu\otimes\nu}(X \times Y, \mathcal{A} \otimes \mathcal{B}, F)$ and

$$\iint f d\mu \otimes d\nu = \int_X \left(\int_Y f_x\, d\nu\right) d\mu = \int_Y \left(\int_X f_y\, d\mu\right) d\nu.$$

Theorem 5.57 is proved in Lang [62] (Chapter VI, §8, Theorem 8.7); see also Marle [69] (Chapter 5, Section 2).

In practice, it is customary to use a less formal notation to express Fubini's theorem, namely

$$\iint f(x,y) d\mu(x) \otimes d\nu(y) = \int_X \left(\int_Y f(x,y)\, d\nu(y)\right) d\mu(x) = \int_Y \left(\int_X f(x,y)\, d\mu(x)\right) d\nu(y),$$

and the measure $d\mu(x) \otimes d\nu(y)$ is often denoted simply by $d\mu(x)d\nu(y)$.

As an application of the product measure, we define the Lebesgue measure in $\mathbb{R}^n$.

5.13 The Lebesgue Measure in $\mathbb{R}^n$

As an application of Theorem 5.55, since the Lebesgue measure μ_L on $\mathbb{R}$ is σ-finite, we see that the product measure $\mu_{L,n}$ of n copies of μ_L is a measure on $\mathbb{R}^n$. The completed σ-algebra (see Proposition 4.8) obtained from the product algebra $\underbrace{\mathcal{L}(\mathbb{R}) \otimes \cdots \otimes \mathcal{L}(\mathbb{R})}_{n}$ is called the σ-algebra of *Lebesgue measurable subsets of* $\mathbb{R}^n$; it is denoted $\mathcal{L}(\mathbb{R}^n)$. To simplify notation, we may write μ_n instead of $\mu_{L,n}$, and $\mathcal{L}^1(\mu_n)$ instead of $\mathcal{L}^1_{\mu_n}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n), \mathbb{C})$.

A crucial property of the Lebesgue integral is that the space $\mathcal{K}^\infty_{\mathbb{C}}(\mathbb{R}^n)$ of *smooth* functions with compact support is dense in $\mathcal{L}^1(\mu_n)$. To prove this, one needs to show the existence of smooth "bump functions" in order to approximate the characteristic function χ_A of a rectangle $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$. The following results are shown in Lang [62] (Chapter VI, Section 9).

Proposition 5.58. *For any function $f \in \mathcal{L}^1(\mu_n)$, if*

$$\int f\varphi\, d\mu_n = 0 \quad \textit{for all } \varphi \in \mathcal{K}^\infty_{\mathbb{C}}(\mathbb{R}^n),$$

then $f = 0$ a.e.

Proposition 5.59. *For every rectangle $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$, for every $\epsilon > 0$, there exits some functions $\varphi, \psi \in \mathcal{K}^\infty_{\mathbb{C}}(\mathbb{R}^n)$ such that*

(1) $0 \leq \varphi \leq \chi_A \leq \psi \leq 1$.

(2) $\int (\psi - \varphi) d\mu_n < \epsilon$.

Furthermore, ψ vanishes outside the rectangle $[a_1 - \epsilon, b_1 + \epsilon] \times \cdots \times [a_n - \epsilon, b_n + \epsilon]$, and $\varphi \equiv 1$ on the rectangle $[a_1 + \epsilon, b_1 - \epsilon] \times \cdots \times [a_n + \epsilon, b_n - \epsilon]$.

Using the above results, we obtain the following theorem.

Theorem 5.60. *The space $\mathcal{K}^\infty_{\mathbb{C}}(\mathbb{R}^n)$ is dense in $\mathcal{L}^1(\mu_n)$ (for the L^1-semi-norm).*

Theorem 5.60 is proven in Lang [62] (Chapter VI, Section 9).

The Lebesgue measure μ_n on $\mathbb{R}^n$ has the same regularity properties as the Lebesgue measure on $\mathbb{R}$, and we have the following version of Proposition 4.14.

Proposition 5.61. *For every Lebesgue-measurable set $A \in \mathcal{L}(\mathbb{R}^n)$, the following facts hold:*

(a)

$$\mu_n(A) = \inf\{\mu_n(O) \mid A \subseteq O,\ O \textit{ is open}\}$$
$$\mu_n(A) = \sup\{\mu_n(K) \mid K \subseteq A,\ K \textit{ is compact}\}.$$

(b) For every $\epsilon > 0$, if $\mu_n(A)$ has finite measure then there is some open subset O such that $A \subseteq O$ and $\mu_n(O - A) < \epsilon$, and there is some compact subset F such that $F \subseteq A$ and $\mu_n(A - F) < \epsilon$.

Proposition 5.61 is proven in Lang [62] (Chapter VI, Section 9).

The Lebesgue measure on $\mathbb{R}^n$ is translation-invariant, which means that $\mu_n(x + A) = \mu_n(A)$ for all $x \in \mathbb{R}^n$ and all $A \in \mathcal{L}(\mathbb{R}^n)$, where $x + A = \{x + a \mid a \in A\}$. This will be proved in Section 8.9.

We conclude this section with the change of variables formula. Without this formula, it would be basically impossible to compute the integrals of familiar functions. The proof is not really difficult but quite long and tedious. The interested reader is referred to Lang [62] (Chapter XXI, Section 2, Theorem 2.6).

Given an injective C^1 function $f \colon U \to \mathbb{R}^n$ where U is some open subset of $\mathbb{R}^n$, which means that the derivative $df \colon U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ is defined and continuous on U, we denote the Jacobian matrix of df_x at $x \in U$ (in the canonical basis of $\mathbb{R}^n$) by $J_f(x)$ (where $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ denotes the vector space of linear maps from $\mathbb{R}^n$ to itself).

Theorem 5.62. *(Change of variables formula, I) Let U be an open subset of $\mathbb{R}^n$, and let $f\colon U \to \mathbb{R}^n$ be an injective C^1 function. For every function $g \in \mathcal{L}^1(f(U), \mu_n)$, we have $(g \circ f)|\det(J_f)| \in \mathcal{L}^1(U, \mu_n)$, and*

$$\int_{f(U)} g(x) d\mu_n(x) = \int_U (g \circ f)(x) |\det(J_f(x))|\, d\mu_n(x).$$

In some cases, for example using polar coordinates, we deal with a C^1 function $f\colon U \to \mathbb{R}^n$ which is only injective on the interior of a measurable subset A of U whose boundary has measure zero. In this case, the following theorem can be used. For a proof, see Lang [62] (Chapter XXI, Section 2, Corollary 2.67).

Theorem 5.63. *(Change of variables formula, II) Let U be an open subset of $\mathbb{R}^n$, and let $f\colon U \to \mathbb{R}^n$ be an injective C^1 function. Let A be a measurable subset of U whose boundary has measure zero, and such that f is injective on the interior of A. For every function $g \in \mathcal{L}^1(f(A), \mu_n)$, we have $(g \circ f)|\det(J_f)| \in \mathcal{L}^1(A, \mu_n)$, and*

$$\int_{f(A)} g(x) d\mu_n(x) = \int_A (g \circ f)(x) |\det(J_f(x))|\, d\mu_n(x).$$

Chapter 6

The Fourier Transform and the Fourier Cotransform on $\mathbb{T}^n$, $\mathbb{Z}^n$, $\mathbb{R}^n$

Historically, trigonometric series were first used by D'Alembert (1747) to solve the equation of a vibrating string, elaborated by Euler a year later, and then solved in a different way essentially using Fourier series by D. Bernoulli (1753). However it was Fourier who introduced and developed Fourier series in order to solve the heat equation, in a sequence of works on heat diffusion, starting in 1807, and culminating with his famous book, *Théorie analytique de la chaleur*, published in 1822.

Originally, the theory of Fourier series is meant to deal with $\mathbb{T} = \mathbf{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\} \cong \mathbb{R}/(2\pi\mathbb{Z})$, say functions with period 2π. Remarkably (but we must apologize for the oversimplification), the theory of Fourier series is captured by the following two equations:

$$f(\theta) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta}, \tag{1}$$

$$c_m = \int_{-\pi}^{\pi} f(\theta) e^{-im\theta} \, \frac{d\theta}{2\pi}. \tag{2}$$

Equation (1) involves a series, and Equation (2) involves an integral. There are two ways of interpreting these equations.

The first way consists of starting with a convergent series as given by the right-hand side of (1) (of course $c_n \in \mathbb{C}$), and to ask what kind of function is obtained. A second question is the following: are the coefficients in (1) computable in terms of the formulae given by (2)?

Such questions were considered by Riemann and then Cantor and Lebesgue. Since they deal with the notion of integral, it is not surprising that they motivated the invention of the Riemann integral and then the Lebesgue integral.

The second way is to start with a periodic function f, apply Equation (2) to obtain the c_m, called *Fourier coefficients*, and then to consider Equation (1). Does the series $\sum_{m \in \mathbb{Z}} c_m e^{im\theta}$ (called *Fourier series*) converge at all? Does it converge to f?

Observe that the expression $f(\theta) = \sum_{m\in\mathbb{Z}} c_m e^{im\theta}$ may be interpreted as a countably infinite superposition of elementary periodic functions, intuitively representing simple wave functions, the functions $\theta \mapsto e^{im\theta}$. We can think of m as the frequency of this wave function.

The above questions were first considered by Fourier. Fourier boldly claimed that *every* function can be represented by a Fourier series. Of course this is false, and for several reasons. First, one needs to define what is an integrable function, and there are plenty of nonintegrable functions. Second, it depends on the kind of convergence that are we dealing with. The nth partial sum $S_{n,f}$ of the Fourier series $\sum_{m\in\mathbb{Z}} c_m e^{im\theta}$ for f (where the c_m are given by Equation (2)) is given by

$$S_{n,f}(\theta) = \sum_{k=-n}^{n} c_k e^{ik\theta}.$$

The most common type of convergence is *pointwise convergence*, which means that for every θ, we have $\lim_{n\mapsto\infty} |f(\theta) - S_{n,f}(\theta)| = 0$. Even if f is a continuous function, there are examples of Fourier series that do not converge pointwise for $\theta = 0$ (du Bois-Reymond). There is even a function in $\mathrm{L}^1(\mathbb{T})$ whose Fourier series diverges for all θ (Kolmogoroff). The convergence of Fourier series is a subtle matter.

But Fourier was almost right. If we consider a function f in $\mathrm{L}^2(\mathbb{T})$, a famous and deep theorem of Carleson states that its Fourier series converges to f pointwise almost everywhere. Other ways to ensure the convergence of the Fourier series of a function is to either restrict the class of functions being considered (Dirichlet, Jordan), or to use different kinds of summation (Abel, Cesàro). Abel summation leads to the Poisson kernel, and Cesàro summation leads to the Féjer kernel; see Example 8.10, Section 6.1, and Stein and Shakarchi [94] (Chapter 2).

In Section 6.1, as a motivation for Fourier analysis on $\mathbb{T}$, we solve the wave equation for a vibrating string. We are led immediately to the problem of Fourier inversion.

Given a periodic function f, the problem of determining when f can be reconstructed as the Fourier series (Equation (1)) given by its Fourier coefficients c_m (Equation (2)) is called the problem of *Fourier inversion*. To discuss this problem, it is useful to adopt a more general point of view of the correspondence between functions and Fourier coefficients, and Fourier coefficients and Fourier series.

Given a function $f \in \mathrm{L}^1(\mathbb{T})$, Equation (2) yields the $\mathbb{Z}$-indexed sequence $(c_m)_{m\in\mathbb{Z}}$ of Fourier coefficients of f, with

$$c_m = \int_{-\pi}^{\pi} f(\theta) e^{-im\theta} \frac{d\theta}{2\pi},$$

which we call the *Fourier transform* of f and denote by $\widehat{f}$, or $\mathcal{F}(f)$. We can view the Fourier transform $\mathcal{F}(f)$ of f as a function $\mathcal{F}(f)\colon \mathbb{Z} \to \mathbb{C}$ with domain $\mathbb{Z}$.

On the other hand, given a $\mathbb{Z}$-indexed sequence $c = (c_m)_{m\in\mathbb{Z}}$ of complex numbers c_m, we can define the Fourier series $\overline{\mathcal{F}}(c)$ associated with c, or *Fourier cotransform* of c, given by

$$\overline{\mathcal{F}}(c)(\theta) = \sum_{m\in\mathbb{Z}} c_m e^{im\theta}.$$

This time $\overline{\mathcal{F}}(c)$ is a function $\overline{\mathcal{F}}(c)\colon \mathbb{T} \to \mathbb{C}$ with domain $\mathbb{T}$. Of course there is an issue of convergence. If $c = (c_m) \in \ell^1(\mathbb{Z})$, then the series $\overline{\mathcal{F}}(c)$ converges uniformly. In general, if $c = (c_m) \notin \ell^1(\mathbb{Z})$, then $\overline{\mathcal{F}}(c)(\theta)$ may be undefined. If $(\overline{\mathcal{F}} \circ \mathcal{F})(f))(\theta)$ is defined, Fourier inversion can be stated as the equation

$$f(\theta) = ((\overline{\mathcal{F}} \circ \mathcal{F})(f))(\theta).$$

In general, even if $f \in \mathrm{L}^1(\mathbb{T})$, the above equation fails.

There are special cases for which Fourier inversion holds. One case is if $\mathcal{F}(f) \in \ell^1(\mathbb{Z})$, which means that the sum $\sum_{m\in\mathbb{Z}} |c_m|$ is finite. Another case is if $f \in \mathrm{L}^2(\mathbb{T})$. In fact, Plancherel's theorem asserts that the map $f \mapsto \widehat{f}$ is an isometric isomorphism between $\mathrm{L}^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$.

In Section 6.3 we return to the issue of pointwise convergence of Fourier series on $\mathbb{T}$. We give examples of functions for which the Fourier series does not converge pointwise, or worse. We show that for the class of functions of bounded variation there is a pointwise convergence theorem due to Dirichlet and Jordan.

In Section 6.4 we generalize the results of Section 6.1 to $\mathbb{T}^n$ and $\mathbb{Z}^n$. In addition to the definition of the Fourier transform on $\mathbb{T}^n$, we define the Fourier cotransform on $\mathbb{T}^n$, and in addition to the definition of the cotransform on $\mathbb{Z}^n$, we define the Fourier transform on $\mathbb{Z}^n$. We also generalize the Poisson kernel to $\mathbb{T}^n$ and prove generalizations of the results of Section 6.1 on spectral synthesis and Abel summation. Plancherel's theorem asserts that the map $f \mapsto \widehat{f}$ is an isometric isomorphism between $\mathrm{L}^2(\mathbb{T}^n)$ and $\ell^2(\mathbb{Z}^n)$.

In Section 6.5 we discuss the Fourier transform of functions defined on the entire real line $\mathbb{R}$ that are not necessarily periodic. Because $\mathbb{R}$ is not compact, $\mathrm{L}^1(\mathbb{R})$ and $\mathrm{L}^2(\mathbb{R})$ are incomparable (with respect to inclusion), and the theory of the Fourier transform on $\mathbb{R}$ is more delicate than the Fourier theory on $\mathbb{T}$.

In Section 6.6 we consider a classical problem in signal processing, which is to reconstruct a function $f\colon \mathbb{R} \to \mathbb{C}$ which is *band-limited*, which means that its Fourier transform $\widehat{f}$ vanishes outside some interval $[-\Omega, \Omega]$. Then f can be completely reconstructed by sampling at the points $t_n = n\pi/\Omega$, for $n \in \mathbb{N}$. We obtain the sampling theorem (Theorem 6.25).

The results of Section 6.5 are generalized to $\mathbb{R}^n$ in Section 6.7.

In Section 6.8 we define a class $\mathcal{S}(\mathbb{R}^n)$ of smooth functions that decay quickly when $\|x\|$ goes to infinity called the *Schwartz space*. The space $\mathcal{S}(\mathbb{R}^n)$ is not a normed vector space, but its topology can be defined by a countable family of semi-norms. It is a metrizable space that is complete, called a Fréchet space. The Schwartz space is closed under the Fourier transform and cotransform and is generally well-behaved. Fourier inversion holds and taking the Fourier transform of a derivative is just multiplication of the Fourier transform by a variable.

This last property can be exploited to solve certain partial differential equations by converting them to ordinary differential equations *via* the Fourier transform. We illustrate this method by solving the steady-state heat equation in the upper half-plane.

In Section 6.9 we discuss the Poisson summation formula, which is a way of finding the Fourier coefficients of the periodic function obtained from a nonperiodic function by applying the process of periodization.

In Section 6.10 we show that roughly, a function f and its Fourier transform $\widehat{f}$ can't be both highly localized. This can be stated precisely in terms of the *dispersion* of f about the point a given by

$$\Delta_a f = \int (x-a)^2 |f(x)|^2 \, dx \Big/ \int |f(x)|^2 \, dx.$$

The Heisenberg inequality states that if f is a function in $\mathrm{L}^2(\mathbb{R})$, then for all $a, b \in \mathbb{R}$, we have

$$(\Delta_a f)(\Delta_b \widehat{f}) \geq \frac{1}{4}.$$

We briefly discuss the interpretation of this inequality in quantum mechanics, called the *Heisenberg uncertainty principle*.

In the last section, Section 6.11, we give a brief summary of Fourier's captivating life.

6.1 Fourier Analysis on $\mathbb{T}$

We begin this chapter with a preview of Fourier analysis on one of the simplest locally compact abelian groups, namely $\mathbb{T}$.

Definition 6.1. The *circle group* $\mathbb{T} = \mathbf{U}(1)$ is the group $\{z \in \mathbb{C} \mid |z| = 1\}$ of complex numbers of unit length under multiplication. We give $\mathbb{T}$ the subspace topology induced by $\mathbb{C}$.

The circle group $\mathbb{T}$ is abelian (commutative). As a set,

$$\mathbb{T} = \{e^{i\theta} \mid \theta \in [-\pi, \pi)\}.$$

Geometrically, this is the unit circle $S = S^1$.

The map $\sigma \colon \mathbb{R} \to \mathbb{T}$ given by

$$\sigma(\theta) = e^{i\theta}$$

is clearly a surjective group homomorphism (with $\mathbb{R}$ under addition, and $\mathbb{T}$ under multiplication); see Figure 6.1. Since $e^{i\theta} = 1$ iff $\theta = k2\pi$ with $k \in \mathbb{Z}$, we see that the kernel of σ is $2\pi\mathbb{Z}$, so by the first isomorphism theorem the *additive group* $\mathbb{R}/(2\pi\mathbb{Z})$ *is isomorphic to the multiplicative group* $\mathbb{T}$. This isomorphism allows to view a complex number of unit length as $e^{i\theta}$, with θ defined modulo 2π, which is often more convenient than picking a representative of the equivalence class of $\theta \pmod{2\pi}$ in $[-\pi, \pi)$.

Functions on the unit circle $\mathbb{T}$ are equivalent to periodic functions on $\mathbb{R}$ as defined next.

Definition 6.2. A function $f \colon \mathbb{R} \to \mathbb{C}$ is *periodic with period* T (for some $T \in \mathbb{R}$, with $T > 0$), if $f(x + T) = f(x)$ for all $x \in \mathbb{R}$.

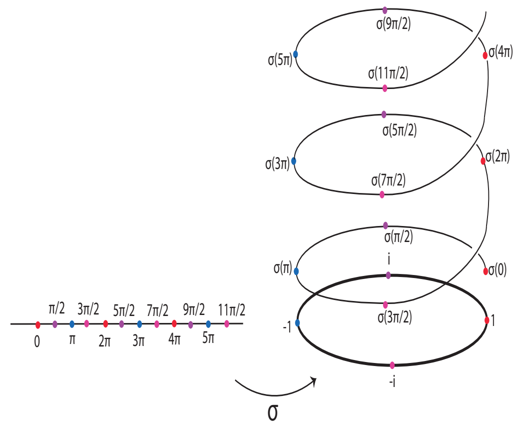


Figure 6.1: The map $\sigma\colon \mathbb{R} \to \mathbb{T}$ which "wraps" the line around the unit circle.

Obviously, a periodic function is completely defined by its restriction to the interval $[-T/2, T/2)$. In most cases, the periods $T = 1$ or $T = 2\pi$ are considered, and which is picked is a matter of taste. We pick $T = 2\pi$. Then we have the following two transformations.

Given a periodic function $f\colon \mathbb{R} \to \mathbb{C}$ (with period 2π), let $f_{\mathbb{T}}\colon \mathbb{T} \to \mathbb{C}$ be the function given by

$$f_{\mathbb{T}}(e^{i\theta}) = f(\theta), \quad -\pi \leq \theta < \pi.$$

Given a function $g\colon \mathbb{T} \to \mathbb{C}$, let $g_{\mathbb{R}}\colon \mathbb{R} \to \mathbb{C}$ be the periodic function (with period 2π) given by

$$g_{\mathbb{R}}(\theta) = g(e^{i\theta}), \quad \theta \in \mathbb{R}.$$

Observe that because the map $\theta \mapsto e^{i\theta}$ is a bijection between $[-\pi, \pi)$ and $\mathbb{T}$, we have

$$(f_{\mathbb{T}})_{\mathbb{R}} = f, \qquad (g_{\mathbb{R}})_{\mathbb{T}} = g,$$

which shows that there is a bijection between the space of periodic function $f : \mathbb{R} \to \mathbb{C}$ (with period 2π), and the space of functions $g\colon \mathbb{T} \to \mathbb{C}$. This bijection restricts to the space of periodic L^p functions that are integrable over $[-\pi, \pi]$, and the space $\mathrm{L}^p(\mathbb{T})$, for $p = 1, 2, \infty$.

The identification between $\mathbb{R}/(2\pi\mathbb{Z})$ and $\mathbb{T}$, and between the space of functions defined on $\mathbb{T}$ and the space of periodic function on $\mathbb{R}$ is often implicit, and in what follows, we take the view that functions on $\mathbb{T}$ are periodic (with period 2π). The reader should be cautioned that other authors use the period 1, so the factor $1/(2\pi)$ showing up in our formulae is missing in the other version (assuming period 1).

To be completely rigorous, we need to equip the abelian group $\mathbb{T}$ with an invariant measure called a Haar measure. This will be done very thoroughly in Chapter 8. For the time being, it suffices to know that in Example 8.10, we show that a normalized Haar measure on $\mathbb{T}$ is given by $dx/2\pi$, where dx is the Lebesgue on $\mathbb{R}$ (so that $\mathbb{T}$ has measure

1). Readers not familiar with the Lebesgue theory of integration should not be concerned, and they should replace this fancy notion with the notion of integral that they are familiar with. After reading this motivating chapter, they should return to the chapters presenting measure theory and integration.

The solution of the wave equation for a vibrating string provides an excellent motivation for using Fourier series on $\mathbb{T}$.

Consider a homogeneous string in the (x, y)-plane, stretched along the x-axis between $x = 0$ and $x = \pi$. The constant π is chosen for mathematical convenience; we could use any constant $L > 0$, but by a change of units, we may assume that it is equal to π. If the string is set to vibrate, its displacement $u(x, t)$ is then a function of x and t. We assume that its endpoints are fixed, so that we have the initial conditions

$$u(0, t) = u(\pi, t) = 0 \qquad \text{for all } t.$$

We also assume that the initial position and velocity of the string are given by two functions f and g defined on $[0, \pi]$ (with $f(0) = f(\pi) = 0$), so that

$$u(x, 0) = f(x), \qquad \frac{\partial u}{\partial t}(x, 0) = g(x).$$

We extend the functions f and g to $[-\pi, \pi]$ by making them odd, namely, we set $f(-x) = -f(x)$ and $g(-x) = -g(x)$ for $x \in [0, \pi]$, and then we extend f and g to $\mathbb{R}$ by making them periodic of period 2π (so, $f(x + 2\pi k) = f(x)$ and $g(x + 2\pi k) = g(x)$, for all $k \in \mathbb{Z}$ and $x \in [-\pi, \pi]$).

Using some physics, it can be shown that u is a solution of the *one-dimensional wave-equation*,

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2},$$

for some constant c. Again, by a change of units, we may assume that $c = 1$, so the wave equation becomes

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2}. \tag{$*$}$$

Equation $(*)$ can be solved by two methods:

1. Using *traveling waves*.

2. Using *standing waves*.

The method of traveling waves was used by d'Alembert, and the method of standing waves by D. Bernoulli; see Stein and Shakarchi [94] (Chapter 1).

The method of standing waves leads immediately to Fourier series. In this method we use the technique of *separation of variables*, which means that we express the solution $u(x, t)$

as the product $u(x,t) = \varphi(x)\psi(t)$, where $\varphi(x)$ and $\psi(t)$ are functions of the two independent variables x and t. Equation $(*)$ yields the equation

$$\varphi(x)\psi''(t) = \varphi''(x)\psi(t),$$

which can be written as

$$\frac{\psi''(t)}{\psi(t)} = \frac{\varphi''(x)}{\varphi(x)}.$$

Since the left-hand side depends only on t, and the right-hand side depends only on x, the above equation can hold only if both sides are equal to the same constant, say λ, so we deduce that

$$\begin{aligned}\varphi''(x) - \lambda\varphi(x) &= 0\\ \psi''(t) - \lambda\psi(t) &= 0.\end{aligned}$$

These equations have well-known solutions. If $\lambda > 0$, then

$$\varphi(x) = Ae^{\sqrt{\lambda}x} + Be^{-\sqrt{\lambda}x},$$

and we obtain a solution which is not physically possible since the displacement of the string is unbounded, so we must have $\lambda \leq 0$, say $\lambda = -m^2$. The solution is given by

$$\varphi(x) = \alpha e^{imx} + \beta e^{-imx}$$

with $\alpha, \beta \in \mathbb{C}$, or equivalently

$$\begin{aligned}\varphi(x) &= \frac{\alpha}{2}(\cos mx + i\sin mx) + \frac{\beta}{2}(\cos mx - i\sin mx)\\ &= \frac{(\alpha+\beta)}{2}\cos mx + i\frac{(\alpha-\beta)}{2}\sin mx.\end{aligned}$$

Since we are seeking real functions as solutions, the solutions are given by

$$\begin{aligned}\varphi(x) &= C\cos mx + D\sin mx\\ \psi(t) &= A\cos mt + B\sin mt,\end{aligned}$$

with $A, B, C, D \in \mathbb{R}$. Since $\varphi(0) = \varphi(\pi) = 0$, we get $C = 0$, and if $D \neq 0$, then m must be an integer in order to have $\sin m\pi = 0$. If $m = 0$, then $\varphi(x) = 0$ for all x, and if $m \leq -1$, we can rename the constants and reduce to the case $m \geq 1$ (since cos is even and sin is odd). Finally, we arrive at the solution

$$u_m(x,t) = (A_m \cos mt + B_m \sin mt)\sin mx, \quad m \geq 1. \tag{$**$}$$

Since the wave equation is linear, any linear combination of the functions in $(**)$ is also a solution, so we are led to the fact that a general solution $u(x,t)$ of the wave equation is a superposition of the solutions u_m, that is,

$$u(x,t) = \sum_{m=1}^{\infty}(A_m \cos mt + B_m \sin mt)\sin mx.$$

There is obviously an issue of convergence, but we will not worry about this yet. The last step is to impose the boundary conditions

$$u(x,0) = f(x), \qquad \frac{\partial u}{\partial t}(x,0) = g(x).$$

which yield the equations

$$\begin{aligned} f(x) &= \sum_{m=1}^{\infty} A_m \sin mx \\ g(x) &= \sum_{m=1}^{\infty} mB_m \sin mx. \end{aligned}$$

Thus we arrived at the following question: given a "reasonable" periodic function $f\colon \mathbb{T} \to \mathbb{C}$ (say $f \in \mathbb{L}^1(\mathbb{T})$), can we find some coefficients $c_m \in \mathbb{C}$ such that

$$f(\theta) = \sum_{m\in\mathbb{Z}} c_m e^{im\theta},$$

where the series on the right-hand side is the Fourier series associated with c_m?

This is the basic problem that motivated Fourier in his quest for solving the heat equation on various domains.

The integer $m \geq 1$ is the *frequency* of the wave component $(A_m \cos mt + B_m \sin mt) \sin mx$, which is called a *harmonic* or *tone*. The general solution is thus a superpositions of harmonics. The case $m = 1$ corresponds to the *first harmonic* or *fundamental tone*. If the vibrating string is the string of a violin, then the first harmonic is the sound of lowest pitch.

If $f \in \mathbb{L}^1(\mathbb{T})$, then we can compute the Fourier coefficients c_m by the formula

$$c_m = \int_{-\pi}^{\pi} f(t) e^{-imt}\, \frac{dx(t)}{2\pi},$$

and then the question is whether the Fourier series $\sum_{m\in\mathbb{Z}} c_m e^{im\theta}$ converges to f (and in what sense).

Recall that the nth partial sum $S_{n,f}$ of the Fourier series $\sum_{m\in\mathbb{Z}} c_m e^{im\theta}$ for f is given by

$$S_{n,f}(\theta) = \sum_{k=-n}^{n} c_k e^{ik\theta},$$

and the average $A_{n,f}$ of these partial sums is given by

$$A_{n,f} = \frac{1}{n}(S_{0,f} + \cdots + S_{n-1,f}).$$

It would be desirable that the partial sums $S_{n,f}$ converge pointwise to f, but in general, this is not the case, even for continuous functions. We will see that if $f \in \mathrm{L}^2(\mathbb{T})$, then $S_{n,f}$ converge to f in the L^2-sense (Proposition 6.2), but the convergence may fail to be pointwise.

In Example 8.10 we discuss Cesàro sums and Féjer's theorem. We show that the average sums $A_{n,f}$ converge uniformly to f if f is continuous. We now discuss *Abel's sums* and Poisson kernels, which yield another kind of convergence.

The *Poisson kernel* on the unit disk is the family of functions $P_r(\theta)$, parametrized by $r \in [0, 1)$, and given by

$$P_r(\theta) = \sum_{n=-\infty}^{n=\infty} r^{|n|} e^{in\theta} = \frac{1 - r^2}{1 - 2r\cos\theta + r^2};$$

see Example 8.11 for the derivation of this formula. Also see Figures 6.2 and 6.3.

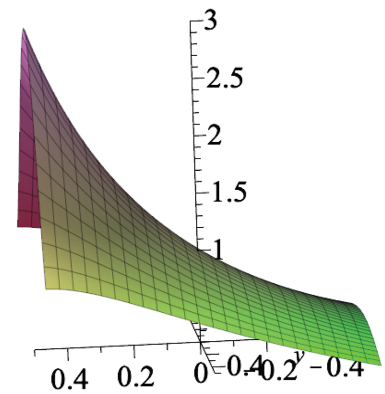


Figure 6.2: The graph $P_r(\theta) = \frac{1-r^2}{1-2r\cos\theta+r^2} = \frac{1-x^2-y^2}{1-2x+x^2+y^2}$ over the region $-1/4 \leq x \leq 1/4$ and $-1/4 \leq y \leq 1/4$. When $r = 0$, the z-coordinate is 1.

A key concept in Fourier analysis is the notion of convolution. To discuss convolution rigorously requires some work so in this chapter we content ourselves with a definition leaving justifications to Section 8.12.

Definition 6.3. The *convolution* $f * g$ of two functions $f, g \in \mathrm{L}^1(\mathbb{T})$ is given by

$$(f * g)(\theta) = \int_{\mathbb{T}} f(\theta - \varphi)g(\varphi)\, \frac{dx(\varphi)}{2\pi} = \int_{\mathbb{T}} f(\varphi)g(\theta - \varphi)\, \frac{dx(\varphi)}{2\pi},$$

where dx is the Lebesgue measure on $\mathbb{R}$.

By Proposition 8.48, we have $f * g \in \mathrm{L}^1(\mathbb{T})$.

We have the following result using the Poisson kernel which gives a preview of Fourier analysis.

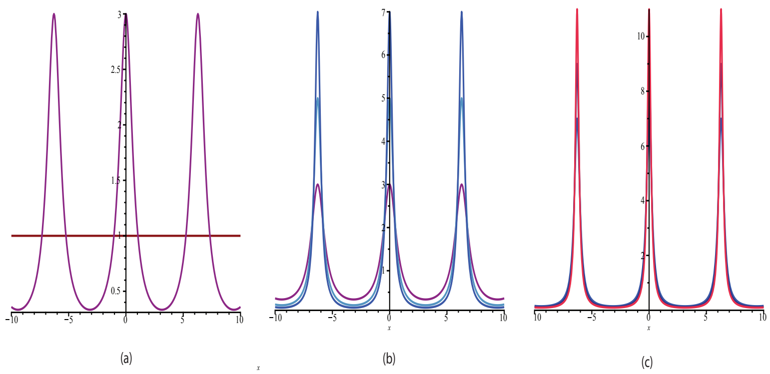


Figure 6.3: Another graphical interpretation of $P_r(\theta) = \frac{1-r^2}{1-2r\cos\theta+r^2}$ when r is fixed. Figure (a) shows the graphs of $P_0(\theta) = 1$ and $P_{1/2}(\theta)$. Figure (b) shows the graphs of $P_{1/2}(\theta)$, $P_{2/3}(\theta)$, and $P_{3/4}(\theta)$, while Figure (c) shows the graphs of $P_{3/4}(\theta)$, $P_{4/5}(\theta)$, and $P_{5/6}(\theta)$. As $r \to 1$, the sinusoid curves have "narrower" peaks centered at $\theta = 2\pi k$, $k \in \mathbb{Z}$, and outside of those peaks, the function limits to the constant value of zero.

Proposition 6.1. *For any $r \in [0, 1)$, if $f \in \mathrm{L}^1(\mathbb{T})$ and if P_r is the Poisson kernel, then for all $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, we have*

$$(P_r * f)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta},$$

where c_m is the mth Fourier coefficient of f,

$$c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \, \frac{dx(t)}{2\pi}.$$

Proof. For $0 \le r < 1$, the series defining P_r is absolutely convergent, so

$$\begin{aligned}
(P_r * f)(\theta) &= \int_{-\pi}^{\pi} P_r(\theta - \varphi) f(\varphi) \, \frac{dx(\varphi)}{2\pi} \\
&= \int_{-\pi}^{\pi} \sum_{m=-\infty}^{m=\infty} r^{|m|} e^{im(\theta-\varphi)} f(\varphi) \, \frac{dx(\varphi)}{2\pi} \\
&= \sum_{m=-\infty}^{m=\infty} \int_{-\pi}^{\pi} r^{|m|} e^{im(\theta-\varphi)} f(\varphi) \, \frac{dx(\varphi)}{2\pi} \\
&= \sum_{m=-\infty}^{m=\infty} r^{|m|} e^{im\theta} \int_{-\pi}^{\pi} f(\varphi) e^{-im\varphi} \, \frac{dx(\varphi)}{2\pi} \\
&= \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta},
\end{aligned}$$

as claimed. ∎

The functions D_n and K_n are defined as

$$D_n(x) = \sum_{k=-n}^{n} e^{ikx}$$
$$K_n(x) = \frac{1}{n}\sum_{m=0}^{n-1}\sum_{k=-m}^{m} e^{ikx} = \frac{1}{n}(D_0(x) + \cdots + D_{n-1}(x)).$$

It can be shown that

$$D_n(x) = \frac{\sin((2n+1)x/2)}{\sin(x/2)}$$
$$K_n(x) = \frac{1}{n}\left(\frac{\sin(nx/2)}{\sin(x/2)}\right)^2.$$

The functions D_n are known as *Dirichlet kernels*, and the functions K_n are *Fejér kernels*. See Figures 6.4 and 6.5. Also see Section 8.15 for applications.

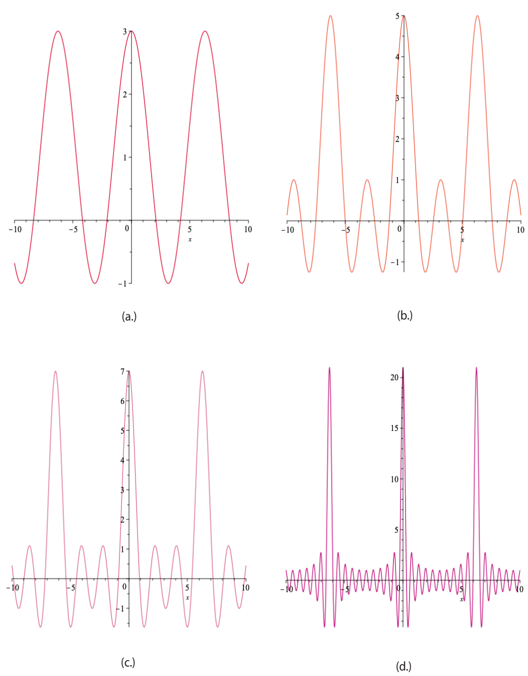


Figure 6.4: Figure (a) is the graph of $D_1(x)$, Figure (b) is the graph of $D_2(x)$, Figure (c) is the graph of $D_3(x)$, while Figure (d) is the graph of $D_{10}(x)$. In all cases the "spike" at $x = 0$ has y-value $2n + 1$.

Observe that for $r = 1$, the partial sum $\sum_{m=-n}^{m=n} c_m r^{|m|} e^{im\theta}$ is the partial sum $S_{n,f}$ of the Fourier series for f, and the partial sum $\sum_{m=-n}^{m=n} r^{|m|} e^{im\theta}$ of the Poisson kernel is the Dirichlet

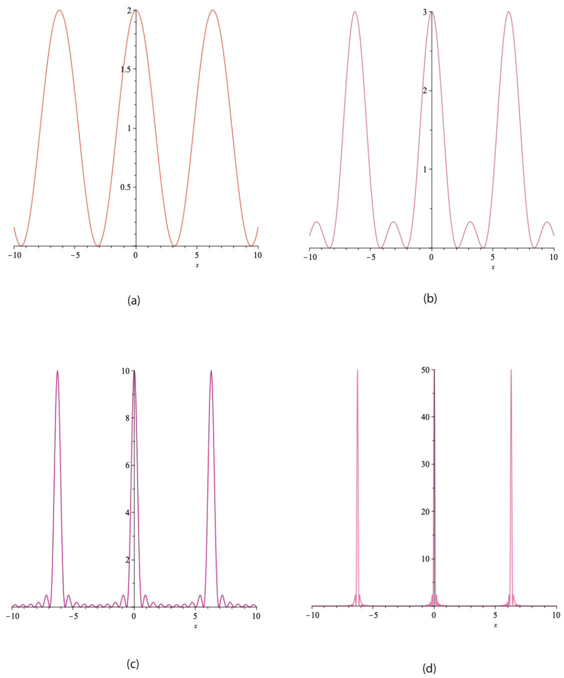


Figure 6.5: Figure (a) is the graph of $K_2(x)$, Figure (b) is the graph of $K_3(x)$, Figure (c) is the graph of $K_{10}(x)$, while Figure (d) is the graph of $K_{50}(x)$. In all cases the "spike" at $x = 0$ has y-value n.

kernel D_n. A slight modification of the proof of Proposition 6.1 shows that

$$D_n * f = S_{n,f},$$

and this immediately implies that

$$K_n * f = A_{n,f}.$$

Recall that for any $p \geq 1$, the space $\ell^p(\mathbb{Z})$ is the set of sequences $x = (x_n)_{n\in\mathbb{Z}}$ with $x_n \in \mathbb{C}$ such that $\sum_{n\in\mathbb{Z}} |x_n|^p < \infty$. Also, if $1 \leq p < q$, then $\ell^p(\mathbb{Z}) \subseteq \ell^q(\mathbb{Z})$; see Figure 6.6.

Indeed, since the sequence $|x_m|^p$ converges, for some $M > 0$ we have $|x_m| < 1$ for all $|m| \geq M$, and since if $q > p$ we have $|x_m|^q \leq |x_m|^p$ (because $|x_m|^p - |x_m|^q = |x_m|^p(1 - |x_m|^{q-p}) \geq 0$ since $|x_m| < 1$), thus $\sum_{|m|\geq M} |x_m|^q \leq \sum_{|m|\geq M} |x_m|^p$, and since $\sum_{n\in\mathbb{Z}} |x_n|^p < \infty$, we also have $\sum_{n\in\mathbb{Z}} |x_n|^q < \infty$.

Each space $\ell^p(\mathbb{Z})$ $(p \geq 1)$ is a normed vector space with the norm

$$\|(x_m)_{m\in\mathbb{Z}}\| = \left(\sum_{m\in\mathbb{Z}} |x_m|^p\right)^{1/p}.$$

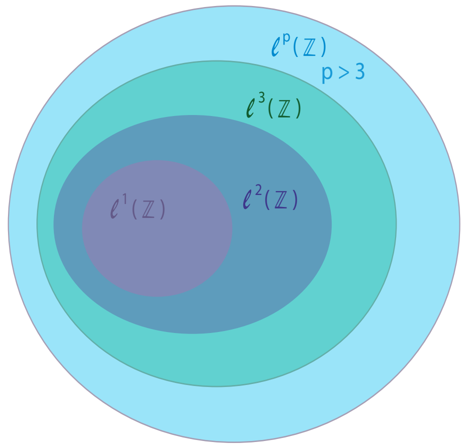


Figure 6.6: A Venn diagram of the containments $\ell^1(\mathbb{Z}) \subseteq \ell^2(\mathbb{Z}) \subseteq \ell^3(\mathbb{Z}) \subseteq \ell^p(\mathbb{Z})$, where $p > 3$.

The space $\ell^p(\mathbb{Z})$ ($p \geq 1$) is a Banach space (it is complete). This is proven by a simple modification of the proof of Proposition D.14.

In general, given a function $f \in \mathrm{L}^1(\mathbb{T})$, the Fourier series $\sum_{m\in\mathbb{Z}} c_m e^{im\theta}$ *does not* converge pointwise. However, if $0 \leq r < 1$, then $f_r(\theta) = (P_r * f)(\theta) = \sum_{m\in\mathbb{Z}} c_m r^{|m|} e^{im\theta}$, so the series on the right-hand side converges pointwise. The following results shows that if r tends to 1, then f_r is an approximation of f that tends to f (in a technical sense). Since $\mathbb{T}$ is compact, we have $\mathrm{L}^2(\mathbb{T}) \subseteq \mathrm{L}^1(\mathbb{T})$. Then if $f \in \mathrm{L}^2(\mathbb{T})$, the partial sums of the Fourier series of f converge to f in the L^2-norm.

Theorem 6.2. *(Spectral Synthesis)*

(1) If $f \in \mathrm{L}^p(\mathbb{T})$ for $p = 1, 2$, and if $r \in [0, 1)$, for all $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, write

$$f_r(\theta) = (P_r * f)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta},$$

with

$$c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \frac{dx(t)}{2\pi}.$$

Then $\lim_{r \mapsto 1} \|f - f_r\|_p = 0$.

(2) If $f \in \mathcal{C}(\mathbb{T})$, then $\lim_{r\mapsto 1} \|f - f_r\|_\infty = 0$.

(3) If $f \in \mathrm{L}^2(\mathbb{T})$, then

$$\lim_{n\mapsto\infty} \left\| f - \sum_{m=-n}^{m=n} c_m e^{im\theta} \right\|_2 = 0.$$

Furthermore, we have the Parseval theorem:

$$\|f\|_2^2 = \sum_{m=-\infty}^{m=\infty} |c_m|^2.$$

The above implies that $c = (c_m)_{m\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$.

Theorem 6.2 is proven in Malliavin [68] (Chapter 3, Section 2.2.5). The function

$$f_r(\theta) = (P_r * f)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta}$$

is known as the r*th Abel mean* of the Fourier series

$$\sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$$

of f. The Fourier series does not always converge pointwise, but the rth Abel mean f_r converges uniformly for all $r < 1$ ($r \geq 0$).

The results of Theorem 6.2 are examples of *spectral synthesis*, namely, the reconstruction of a function from its Fourier coefficients. Facts (1) and (2) are not very practical because they require first summing the series f_r. Fact (2) for continuous functions is better because it shows uniform convergence. Fact (3) is very satisfactory since it shows convergence of the partial sums of the Fourier series in the L^2-sense, but convergence pointwise generally fails. If $(c_m) \in \ell^2(\mathbb{Z})$, for some $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$, the sums $\sum_{m=-n}^{m=n} c_m e^{im\theta}$ may not converge. For more about this phenomenon, see Section 6.3.

Remark: Lennart Carleson showed in 1966 that for any function $f \in \mathrm{L}^2(\mathbb{T})$, the partial sums of the Fourier series of f converge pointwise almost everywhere to f, putting a close to a problem that had been open for fifty years.

6.2 Fourier Inversion on $\mathbb{T}$

Recall that for any $p \geq 1$, the space $\ell^p(\mathbb{Z})$ is the set of sequences $x = (x_n)_{n\in\mathbb{Z}}$ with $x_n \in \mathbb{C}$ such that $\sum_{n\in\mathbb{Z}} |x_n|^p < \infty$. For $p = 2$, the space $\ell^2(\mathbb{Z})$ is a Hilbert space with the inner product

$$\langle (x_m)_{m\in\mathbb{Z}}, (y_m)_{m\in\mathbb{Z}} \rangle = \sum_{m\in\mathbb{Z}} x_m \overline{y_m}$$

and norm

$$\|(x_m)_{m\in\mathbb{Z}}\| = \left(\sum_{m\in\mathbb{Z}} |x_m|^2 \right)^{1/2};$$

see Proposition D.14.

The following result shows that if the sequence $c = (c_m)_{m\in\mathbb{Z}}$ of Fourier coefficients of f is well-behaved, then f can be reconstructed from c.

Theorem 6.3. *(Fourier inversion formula) Let $f \in \mathrm{L}^1(\mathbb{T})$. If $c = (c_m)_{m\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$, that is, if the series $\sum_{m=-\infty}^{m=\infty} |c_m|$ converges, where c_m is the Fourier coefficient*

$$c_m = \int_{-\pi}^{\pi} f(t)e^{-imt}\,\frac{dx(t)}{2\pi},$$

then

$$f(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$$

for all almost all $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Furthermore, if f is continuous, then equality holds everywhere.

Proof. Write

$$\varphi(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$$

and recall that

$$f_r(\theta) = \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta}.$$

Since the series $\sum_{m=-\infty}^{m=\infty} |c_m|$ converges, the series defining φ converges absolutely, so φ is continuous. We claim that

$$\lim_{r\mapsto 1} \|\varphi - f_r\|_\infty = 0.$$

We have

$$\|\varphi - f_r\|_\infty \le \sum_{m=-\infty}^{m=\infty} |c_m|(1 - r^{|m|}).$$

Given $\epsilon > 0$, we can find p so that $\sum_{|m|>p} |c_m| \le \epsilon/2$. Then $\sum_{|m|\le p} |c_m|(1 - r^{|m|})$ is the sum of $2p + 1$ terms that tend to 0 as r tends to 1, so for r close enough to 1 so that $\sum_{|m|>p} |c_m|(1 - r^{|m|}) < \epsilon/2$, we have

$$\sum_{|m|>p} |c_m|(1 - r^{|m|}) + \sum_{|m|\le p} |c_m|(1 - r^{|m|}) < \epsilon/2 + \epsilon/2 = \epsilon,$$

which shows that $\lim_{r\mapsto 1} \|\varphi - f_r\|_\infty = 0$.

Since by Proposition 5.24(2), $\|\varphi - f_r\|_1 \le 2\pi \|\varphi - f_r\|_\infty$, we also have

$$\lim_{r\mapsto 1} \|\varphi - f_r\|_1 = 0.$$

Since $f \in \mathrm{L}^1(\mathbb{T})$, by Theorem 6.2(1),

$$\lim_{r \mapsto 1} \|f - f_r\|_1 = 0,$$

and since

$$\|f - \varphi\|_1 \leq \|f - f_r\|_1 + \|f_r - \varphi\|_1,$$

we deduce that

$$\|f - \varphi\|_1 = 0,$$

which means that $f = \varphi$ almost everywhere. If f is continuous, since φ is also continuous, $f - \varphi = h$ is continuous. But if $h \neq 0$, then h is nonzero on some interval, which contradicts the fact that $f = \varphi$ almost everywhere. $\square$

Definition 6.4. Given any function $f \in \mathrm{L}^1(\mathbb{T})$, the function $\mathcal{F}(f)\colon \mathbb{Z} \to \mathbb{C}$ given by $\mathcal{F}(f)(m) = c_m$, where c_m is the *Fourier coefficient*

$$c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \, \frac{dx(t)}{2\pi},$$

is called the *Fourier transform* of f. We identify the sequence $\mathcal{F}(f)$ with the sequence $(c_m)_{m \in \mathbb{Z}}$, which is also denoted by $\widehat{f}$.

Theorem 6.2(3) (Parseval's theorem) implies that if $f \in \mathrm{L}^2(\mathbb{T})$, then $\widehat{f} \in \ell^2(\mathbb{Z})$. However, if $f \in \mathrm{L}^1(\mathbb{T})$, then it *may not* be the case that $\widehat{f} \in \ell^1(\mathbb{Z})$.

Theorem 6.3 says that if $f \in \mathrm{L}^1(\mathbb{T})$ and if $\widehat{f} \in \ell^1(\mathbb{Z})$, then f can be reconstructed by its Fourier series $\sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$. Theorem 6.2(3) says that if $f \in \mathrm{L}^2(\mathbb{T})$ then the partial sums $S_{n,f}$ (with $S_{n,f}(\theta) = \sum_{m=-n}^{m=n} c_m e^{im\theta}$) of the Fourier series of f converge to f in the L^2-norm. It also shows that if $(c_m) \in \ell^2(\mathbb{Z})$, then the partial sums $S_{n,f}$ of the series $\sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$ form a Cauchy sequence in $\mathrm{L}^2(\mathbb{T})$ (with respect to the L^2-norm), and so the series $\sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$ converges to a function in $\mathrm{L}^2(\mathbb{T})$ (in the L^2-norm). In fact, we have a stronger result.

Theorem 6.4. *(Plancherel) The map $\mathcal{F}\colon f \mapsto \widehat{f}$ is an isometric isomorphism of the Hilbert spaces $\mathrm{L}^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$.*

Proof sketch. Theorem 6.4 is classical theorem of Hilbert theory. Its proof can be found in Rudin [79] (Chapter 4) or Malliavin [68] (Chapter 3, Section 2.2.5). Consider the map $\mathcal{F}\colon \mathrm{L}^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$ given by $\mathcal{F}f = \widehat{f}$. The fact that the linear map $\mathcal{F}$ is an isometry is an immediate consequence of Parseval's theorem. This fact implies that $\mathcal{F}$ is injective. We prove the surjectivity of the map $\mathcal{F}$ by a density argument. Since $\mathcal{F}$ is an isometry, its image $\mathcal{F}(\mathrm{L}^2(\mathbb{T}))$ is complete in $\ell^2(\mathbb{Z})$, and thus closed. Consider the subset W of $\ell^2(\mathbb{Z})$ given by

$$W = \{(c_m) \in \ell^2(\mathbb{Z}) \mid c_m = 0 \text{ for all but finitely many } m \in \mathbb{Z}\}.$$

The subset W is dense in $\ell^2(\mathbb{Z})$, and obviously $W \subseteq \ell^1(\mathbb{Z})$. For any $c = (c_m) \in W$, the series $\varphi(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$ only has finitely many nonzero terms $c_m e^{im\theta}$, so φ is continuous, and thus in $\mathrm{L}^2(\mathbb{T})$. It is also immediately verified that $\mathcal{F}(\varphi) = c$. It follows that $W \subseteq \mathcal{F}(\mathrm{L}^2(\mathbb{T}))$, and since W is dense in $\ell^2(\mathbb{Z})$ and $\mathcal{F}(\mathrm{L}^2(\mathbb{T}))$ is closed in $\ell^2(\mathbb{Z})$, we have $\mathcal{F}(\mathrm{L}^2(\mathbb{T})) = \ell^2(\mathbb{Z})$. $\square$

Definition 6.5. Given a sequence $c = (c_m)_{m\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$, we define the *Fourier cotransform* $\overline{\mathcal{F}}(c)$ of c as the function $\overline{\mathcal{F}}(c)\colon \mathbb{T} \to \mathbb{C}$ defined on $\mathbb{T}$ given by

$$\overline{\mathcal{F}}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta} = \sum_{m=-\infty}^{m=\infty} c_m (e^{i\theta})^m,$$

the Fourier series associated with c (with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$). Given a function $f \in \mathrm{L}^1(\mathbb{T})$, if $\widehat{f}$ is the Fourier transform of f, then the Fourier cotransform $\overline{\mathcal{F}}(\widehat{f}) = \sum_{m=-\infty}^{m=\infty} \widehat{f}_m e^{im\theta}$ of $\widehat{f}$ is called the *the Fourier series* of f.

Note that $e^{im\theta}$ is used instead of the term $e^{-im\theta}$ occurring in the Fourier transform.

Since $c \in \ell^1(\mathbb{Z}) \subseteq \ell^2(\mathbb{Z})$, the series $\overline{\mathcal{F}}(c)(\theta)$ converges uniformly. On the other hand, if $c \in \ell^2(\mathbb{Z})$, the series $\overline{\mathcal{F}}(c)$ may not converge pointwise, although it converges to a function in $\mathrm{L}^2(\mathbb{T})$ in the L^2-norm. As a consequence the Fourier cotransform $\overline{\mathcal{F}}(c)$ is not defined on $\ell^2(\mathbb{Z})$. However, Theorem 6.4 shows that the Fourier transform $\mathcal{F}$ has an inverse and Theorem 6.3 shows that $\overline{\mathcal{F}}$ is the inverse of $\mathcal{F}$ on $\ell^1(\mathbb{Z})$. This allows to view $\mathcal{F}^{-1}$ as the extension of $\overline{\mathcal{F}}$ from $\ell^1(\mathbb{Z})$ to $\ell^2(\mathbb{Z})$ in such a way that $\mathcal{F}$ and $\overline{\mathcal{F}}$ are mutual inverses between $\mathrm{L}^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$. In general, for an arbitrary function $f \in \mathrm{L}^1(\mathbb{T})$, there is *no guarantee* that $\widehat{f} \in \ell^1(\mathbb{Z})$, so the Fourier series $\overline{\mathcal{F}}(\widehat{f}) = \sum_{m=-\infty}^{m=\infty} \widehat{f}_m e^{im\theta}$ may not converge to f in the L^1 sense or pointwise.

Remark: The maps $e^{i\theta} \mapsto e^{im\theta} = (e^{i\theta})^m$, for $m \in \mathbb{Z}$ and $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$, are continuous homomorphisms of the group $\mathbb{T} = \mathbf{U}(1)$ into itself. In fact, it can be shown that they are the only ones of this kind. They are called the *characters* of $\mathbb{T}$; see Section 10.1 and more generally Chapter 10 for a detailed treatment. Obviously the set of characters of $\mathbb{T}$ is in bijection with $\mathbb{Z}$. Thus the Fourier transform $\mathcal{F}(f)$ of a function $f \in \mathrm{L}^2(\mathbb{T})$, a sequence of complex numbers indexed by $\mathbb{Z}$, can be viewed as a function of the characters of $\mathbb{T}$.

The characters of $\mathbb{Z}$ are the group homomorphisms $\varphi\colon \mathbb{Z} \to \mathbb{T}$. Since $\mathbb{Z}$ is generated by 1, a homomorphism satisfies the equation

$$\varphi(m) = (\varphi(1))^m, \qquad m \in \mathbb{Z},$$

so it is uniquely determined by picking $\varphi(1) = e^{i\theta} \in \mathbb{T}$ (with $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$), and is of the form $\varphi(m) = (e^{i\theta})^m = e^{im\theta}$ for all $m \in \mathbb{Z}$. Thus the set of characters of $\mathbb{Z}$ is in bijection with $\mathbb{T}$. Then the Fourier cotransform $\overline{\mathcal{F}}(c)$ of a "function" $c \in \ell^2(\mathbb{Z})$ ($\overline{\mathcal{F}}(c)$ is the Fourier series associated with c) can also be viewed as a function on the characters of $\mathbb{Z}$, namely a function on $\mathbb{T}$. This fact generalizes to an arbitrary abelian locally compact group and is the key to the definition of the Fourier transform on such a group; see Chapter 10.

Sometimes it is more convenient to express the Fourier series $\overline{\mathcal{F}}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$ in terms of $\cos m\theta$ and $\sin m\theta$ instead of the complex exponentials $e^{im\theta}$. Here we are assuming that $c = (c_m)_{m\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$, so the series $\sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$ is absolutely convergent and it is

permissible to permute terms. Since $e^{im\theta} = \cos m\theta + i \sin m\theta$, the Fourier series $\overline{\mathcal{F}}(c)(\theta)$ can be expressed as

$$\begin{aligned}
\overline{\mathcal{F}}(c)(\theta) &= \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta} = \sum_{m=-\infty}^{m=\infty} c_m(\cos m\theta + i \sin m\theta) \\
&= \sum_{m=-\infty}^{-1} c_m(\cos m\theta + i \sin m\theta) + \sum_{m=0}^{\infty} c_m(\cos m\theta + i \sin m\theta) \\
&= c_0 + \sum_{m=1}^{\infty} c_{-m}(\cos m\theta - i \sin m\theta) + \sum_{m=1}^{\infty} c_m(\cos m\theta + i \sin m\theta) \\
&= c_0 + \sum_{m=1}^{\infty} ((c_m + c_{-m}) \cos m\theta + i(c_m - c_{-m}) \sin m\theta).
\end{aligned}$$

Therefore, if we let

$$a_0 = 2c_0, \qquad a_m = c_m + c_{-m}, \qquad b_m = i(c_m - c_{-m}), \qquad m \geq 1,$$

then we have

$$\overline{\mathcal{F}}(c)(\theta) = \frac{1}{2}a_0 + \sum_{m=1}^{\infty} (a_m \cos m\theta + b_m \sin m\theta). \tag{$\dagger$}$$

Equation (†) makes it very clear that the function $\overline{\mathcal{F}}(c)(\theta)$ can be viewed as the countably infinite superposition of the basic periodic functions $\cos m\theta$ and $\sin m\theta$, often called *harmonics*. The number $m \in \mathbb{N} - \{0\}$ is called a *frequency*.

Conversely, if

$$\overline{\mathcal{F}}(c)(\theta) = \frac{1}{2}a_0 + \sum_{m=1}^{\infty} (a_m \cos m\theta + b_m \sin m\theta),$$

then if we let

$$c_0 = \frac{1}{2}a_0, \qquad c_m = \frac{1}{2}(a_m - ib_m), \qquad c_{-m} = \frac{1}{2}(a_m + ib_m), \qquad m \geq 1,$$

then

$$\overline{\mathcal{F}}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}.$$

From

$$c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \, \frac{dx(t)}{2\pi},$$

and

$$a_0 = 2c_0, \qquad a_m = c_m + c_{-m}, \qquad b_m = i(c_m - c_{-m}), \qquad m \geq 1,$$

for $m \geq 1$, we get

$$a_m = c_m + c_{-m} = \int_{-\pi}^{\pi} f(t)(e^{-imt} + e^{imt})\,\frac{dx(t)}{2\pi} = 2\int_{-\pi}^{\pi} f(t)\cos mt\,\frac{dx(t)}{2\pi}$$

and

$$b_m = i(c_m - c_{-m}) = \int_{-\pi}^{\pi} f(t)i(e^{-imt} - e^{imt})\,\frac{dx(t)}{2\pi} = 2\int_{-\pi}^{\pi} f(t)\sin mt\,\frac{dx(t)}{2\pi},$$

that is, for $m \geq 1$, we have

$$\begin{aligned} a_m &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos mt\,dt \\ b_m &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin mt\,dt. \end{aligned}$$

We also have

$$a_0 = 2c_0 = 2\int_{-\pi}^{\pi} f(t)\,\frac{dx(t)}{2\pi} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\,dt.$$

Therefore we can combine the above equations and we obtain

$$\begin{aligned} a_m &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos mt\,dt \qquad (m \geq 0) \\ b_m &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin mt\,dt \qquad (m \geq 1). \end{aligned}$$

Note that the equation for a_m also holds for $m = 0$. This is the reason for the term $(1/2)a_0$ in equation (†). The numbers a_m and b_m are also called the *Fourier coefficients* of f. If the function f is real-valued, then the coefficients a_m and b_m are real.

Observe that

$$c_0 = \frac{1}{2}a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\,dt$$

is the mean value of f over the interval $[-\pi, \pi]$.

Here are a few examples of Fourier transforms. Many more examples can be found in Folland [32].

Example 6.1. Let $f\colon [-\pi, \pi] \to \mathbb{R}$ be the periodic function given by

$$f(\theta) = |\theta|, \qquad -\pi \leq \theta \leq \pi.$$

The graph of $f(\theta)$ is shown in Figure 6.7.

Let us compute the coefficients a_m and b_m. Since the function f is even, we have $b_m = 0$ for all $m \geq 1$, and

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)\cos m\theta\,d\theta = \frac{2}{\pi}\int_{0}^{\pi} f(\theta)\cos m\theta\,d\theta.$$

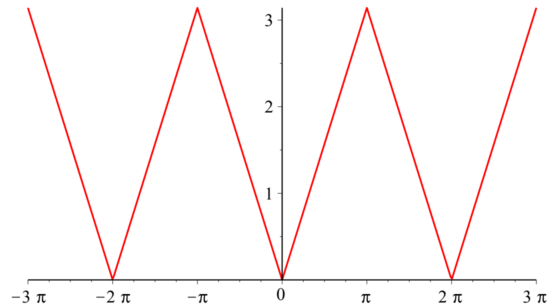


Figure 6.7: The graph of the periodic function $f(\theta) = |\theta|$, where $-\pi \leq \theta \leq \pi$.

Thus for $m = 0$ we have

$$a_0 = \frac{2}{\pi}\int_0^{\pi} \theta\, d\theta = \frac{1}{\pi}[\theta^2]_0^{\pi} = \pi,$$

and for $m \geq 1$, integrating by parts we have

$$a_m = \frac{2}{\pi}\left[\frac{\theta \sin m\theta}{m}\right]_0^{\pi} - \frac{2}{\pi}\int_0^{\pi} \frac{\sin m\theta}{m}\, d\theta = \frac{2}{\pi}\left[\frac{\cos m\theta}{m^2}\right]_0^{\pi} = \frac{2}{\pi}\frac{(-1)^m - 1}{m^2}$$

since $\sin m\pi = 0$ and $\cos m\pi = (-1)^m$. Now $(-1)^m - 1 = -2$ when m is odd and 0 when m is even, so we find that the Fourier series for f is given by

$$\frac{\pi}{2} - \frac{4}{\pi}\sum_{m \text{ odd}} \frac{1}{m^2}\cos m\theta = \frac{\pi}{2} - \frac{4}{\pi}\sum_{k=1}^{\infty} \frac{\cos(2k-1)\theta}{(2k-1)^2}.$$

If we plot the graphs of the partial sums for a few terms (say five terms), we see that they provide a very good approximation to f. See Figures 6.8 and 6.9. The series converges uniformly to f due to the presence of the term $1/(2k-1)^2$.

Example 6.2. Let $f\colon [-\pi, \pi] \to \mathbb{R}$ be the periodic function given by

$$f(\theta) = \theta, \qquad -\pi < \theta \leq \pi.$$

The graph of $f(\theta)$ is shown in Figure 6.10.

This time let us compute the coefficients c_m. We have

$$c_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \theta\, d\theta = 0,$$

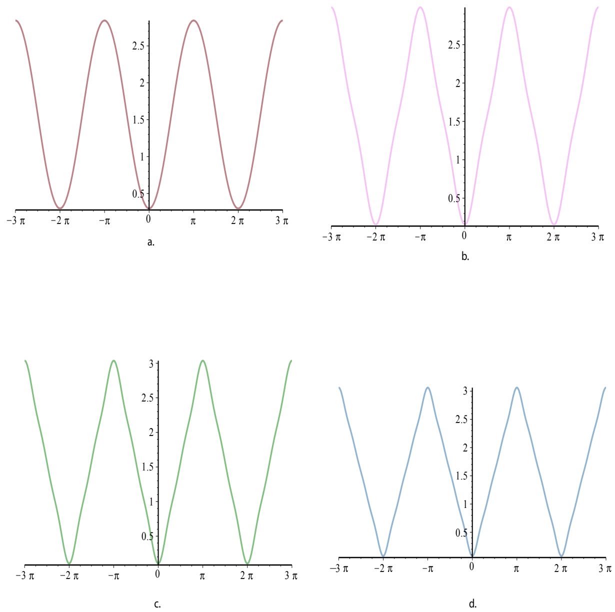


Figure 6.8: Let $S_M = \frac{\pi}{2} - \frac{4}{\pi}\sum_{k=1}^{M}\frac{\cos(2k-1)\theta}{(2k-1)^2}$. Figure (a) is the graph of S_1; Figure (b) is the graph of S_2; Figure (c) is the graph of S_3, and Figure (d) is the graph of S_4.

and for $m \neq 0$, by integrating by parts, we have

$$\begin{aligned} c_m &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\theta e^{-im\theta}\,d\theta \\ &= \frac{1}{2\pi}\left[\frac{\theta e^{-im\theta}}{-im}\right]_{-\pi}^{\pi} - \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-im\theta}}{-im}\,d\theta \\ &= \frac{1}{2\pi}\left[e^{-im\theta}\left(\frac{\theta}{-im}+\frac{1}{m^2}\right)\right]_{-\pi}^{\pi} \\ &= \frac{(-1)^{m+1}}{im} \end{aligned}$$

since $e^{-im\pi} = e^{im\pi} = (-1)^m$. Hence the Fourier series for f is

$$\sum_{m\neq 0}\frac{(-1)^{m+1}}{im}e^{im\theta}.$$

Since $(-1)^m = (-1)^{-m}$, the mth and the $(-m)$th term can be combined to give

$$(-1)^{m+1}\left(\frac{e^{im\theta}}{im}+\frac{e^{-im\theta}}{-im}\right) = \frac{2(-1)^{m+1}}{m}\sin m\theta,$$

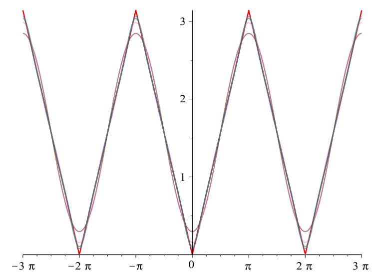


Figure 6.9: The partial sums S_1 through S_4 approximating $f(\theta)$ of Example 6.1.

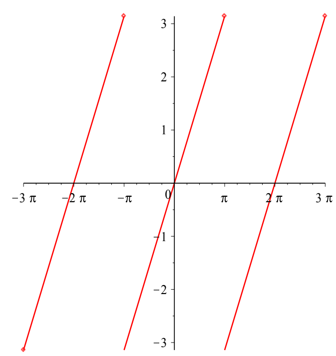


Figure 6.10: The graph of the periodic function $f(\theta) = \theta$, where $-\pi < \theta \leq \pi$.

and we obtain the Fourier series

$$2 \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \sin m\theta. \tag{$*$}$$

Here to be rigorous we should consider the partial sums

$$S_{m,f}(\theta) = \sum_{k=-m}^{m} c_k e^{ik\theta},$$

in which the terms corresponding to the indices $-m$ and m can be combined. The details are left as an exercise. Note that $c_m \in \ell^2(\mathbb{Z})$ but $c_m \notin \ell^1(\mathbb{Z})$, so we can only claim that the Fourier series $(*)$ belongs to $\mathrm{L}^2(\mathbb{T})$.

This time if we plot the graphs of the partial sums, we see that they approximate the function f, but the quality of the approximation is inferior to that of Example 6.1. See Figures 6.11 and 6.12.

This is due to the fact that the function of Example 6.1 is continuous, but the function of Example 6.2 has jump discontinuities. The other reason why the quality of approximation

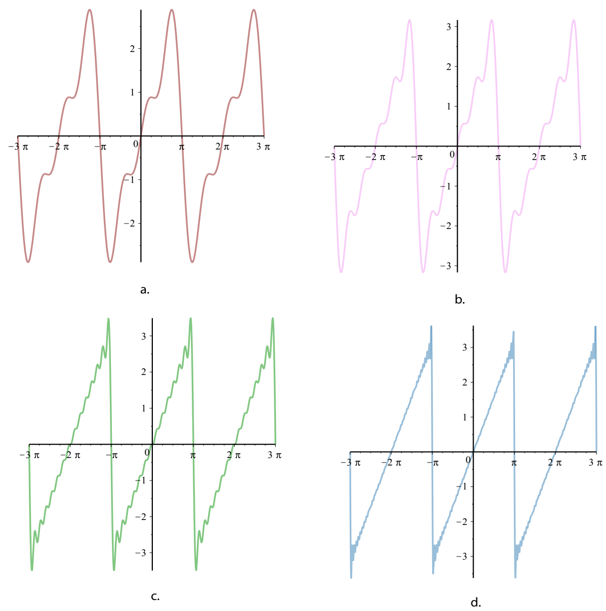


Figure 6.11: Let $S_M = 2\sum_{m=1}^{M} \frac{(-1)^{m+1}}{m} \sin m\theta$. Figure (a) is the graph of S_3; Figure (b) is the graph of S_5; Figure (c) is the graph of S_{14}, and Figure (d) is the graph of S_{40}.

is not as good as in Example 6.1 is that the terms of the series in Example 6.1 tend to zero faster than the terms of the series in Example 6.2. Thus, in Example 6.2, the influence of the higher order terms is much more significant in Example 6.2. The point is that the rougher the function is, the more difficult it is to approximate it by smooth functions such as $\cos m\theta$ and $\sin m\theta$. In fact, it is not obvious that the series

$$2\sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \sin m\theta$$

converges pointwise. It does, with

$$2\sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \sin m\theta = \begin{cases} \theta & \text{if } -\pi < \theta < \pi \\ 0 & \text{if } \theta = \pm\pi. \end{cases}$$

This series converges pointwise to the function f of Example 6.2, except for $\theta = (2k+1)\pi$, according to a theorem of Dirichlet (see Section 6.3).

A phenomenon that shows up in Example 6.2 is the *Gibbs phenomenon*. Even for a partial sum of 40 terms, we observe some spikes near the discontinuities. These spikes tend to zero in width, but not in height; see Folland [32] (Chapter 2, Section 2.6).

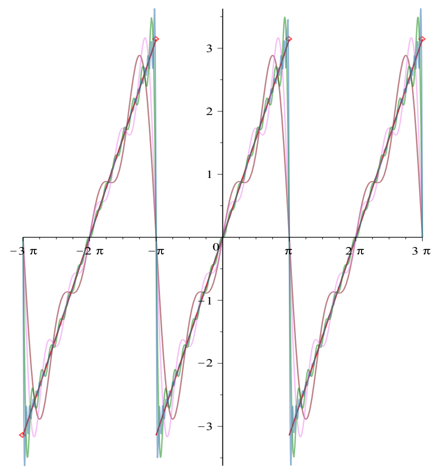


Figure 6.12: The partial sums S_3,S_5,S_{14},S_{40} approximating $f(\theta)$ of Example 6.2.

6.3 Pointwise Convergence of Fourier Series on $\mathbb{T}$

By Theorem 6.2, if $f \in \mathrm{L}^2(\mathbb{T})$, then

$$\lim_{n \mapsto \infty} \left\| f - \sum_{m=-n}^{m=n} c_m e^{im\theta} \right\|_2 = 0,$$

where c_m is the mth Fourier coefficient of f,

$$c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \, \frac{dx(t)}{2\pi}.$$

Thus the partial sums

$$S_{m,f}(\theta) = \sum_{k=-m}^{k=m} c_k e^{ik\theta}$$

converge to f in the L^2-norm. However, even if f is continuous, the partial sums $S_{m,f}$ may not converge to f pointwise.

The first example of a function whose Fourier series diverges at 0 was given by du Bois–Reymond in 1873. This is a fairly complicated example involving a piecewise monotone function that oscillates indefinitely near 0. Simpler examples were given later by Fejér and Lebesgue.

Fejér's example makes use of the functions

$$F_n(x) = \sin x + \frac{1}{2} \sin 2x + \cdots + \frac{1}{n} \sin nx,$$

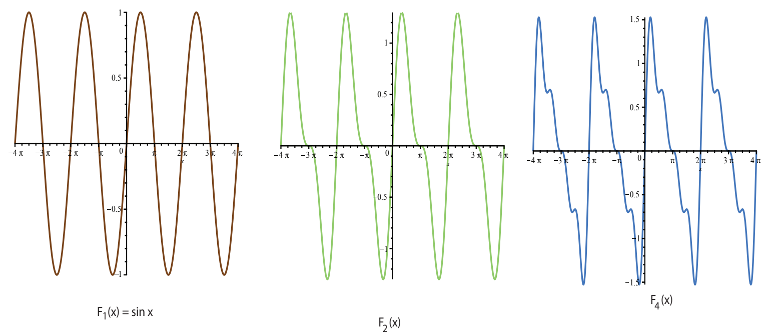


Figure 6.13: The graphs of $F_1(x)$, $F_2(x)$, and $F_4(x)$.

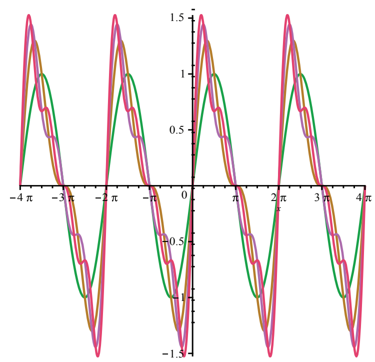


Figure 6.14: The graphs of $F_1(x)$ through $F_4(x)$ superimposed on each other.

which are uniformly bounded; see Figures 6.13 and 6.14.

The series

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \sin(3^{n^2} x) F_{2^{n^2}}(x)$$

defines a continuous function f, but it can be shown that its Fourier series diverges for $x = 0$. See also the example in Stein and Shakarchi [94] (Chapter 3, Section 2.2).

In 1926 Kolmogoroff gave an example of a function $f \in \mathrm{L}^1(\mathbb{T})$ whose Fourier series diverges for all x.

Later it was found that a systematic method for producing functions with a "bad" Fourier series was to use the Banach–Steinhaus theorem.

Definition 6.6. For any fixed $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ and for any continuous function $f \in \mathcal{C}(\mathbb{T}; \mathbb{C})$, let

$$S^*(f, \theta) = \sup_{m \in \mathbb{N}} |S_{m,f}(\theta)|.$$

Also recall the following definition about Borel sets.

Definition 6.7. Let X be a topological space. Countable unions of closed subsets of X are called *F_σ-sets*, and countable intersections of open sets of X are called *G_δ-sets*.

The following result is proven in Rudin [79] (Chapter 5, Page 102). The proof uses the Banach–Steinhaus theorem.

Proposition 6.5. *For every $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, there is a subset $E_\theta \subseteq \mathcal{C}(\mathbb{T};\mathbb{C})$ of continuous functions which is a dense G_δ set in $\mathcal{C}(\mathbb{T};\mathbb{C})$ such that $S^*(f,\theta) = \infty$ for all $f \in E_\theta$. Consequently, the Fourier series of every $f \in E_\theta$ diverges at θ.*

Using Baire's theorem a stronger result can be obtained, as shown in Rudin [79] (Chapter 5, Theorem 5.12).

Proposition 6.6. *There is a set $E \subseteq \mathcal{C}(\mathbb{T};\mathbb{C})$ of continuous functions which is a dense G_δ set in $\mathcal{C}(\mathbb{T};\mathbb{C})$ and which has the following property: For every $f \in E$, the set*

$$Q_f = \{\theta \in \mathbb{R}/2\pi\mathbb{Z} \mid S^*(f,\theta) = \infty\}$$

is a dense G_δ set in $\mathbb{R}/2\pi\mathbb{Z}$.

As a consequence, the Fourier series of every continuous function $f \in E$ diverges for infinitely many points. In fact, E and Q_f are uncountable; see Rudin [79] (Chapter 5, Theorem 5.13).

We just saw that in general, the partial sums $S_{m,f}$ do not behave well, so if we want to approximate a continuous function on $\mathbb{T}$, we should not count on the partial sums to do the job. We will see in Example 8.10 that the Cesàro means,

$$A_{n,f} = \frac{1}{n}(S_{0,f} + \cdots + S_{n-1,f}),$$

have a much better behavior, since they converge uniformly to f.

We are led to the conclusion that in order to obtain positive results for pointwise convergence of the partial sums $S_{m,f}$, we *must restrict* the class of functions that we are considering. Dirichlet was the first to obtain a significant result. In 1829 he proved that the partial sums $S_{m,f}$ converge pointwise to $(f(x+) + f(x-))/2$, for every piecewise continuous and piecewise monotone function f. Here $f(x+)$ is the limit when y tends to x from above, and $f(x-)$ is the limit when y tends to x from below (see Definition 2.19). His paper is not only significant because of its results but because it raised the standards of rigor in mathematical exposition to a new level. Dirichlet's full paper is reproduced in Kahane and Lemarié–Rieusset [52]. As Kahane comments, "Dirichlet's style is superb and incredibly modern."

Later on it was realized that what is really needed for this pointwise convergence result to hold is that the functions have bounded variation. Camille Jordan, whose mathematical

interests were group theory, algebra, and its relations to geometry, introduced the notion of a function of bounded variation in 1881 and published one paper generalizing Dirichlet's paper to this class of functions.

Let $f\colon [a,b] \to \mathbb{R}$ be a function. The intuition is that the variation of f over $[a,b]$ is the total distance travelled from time a to time b. If f' exits and is continuous, then the variation is $\int_a^b |f'(t)|\,dt$. Otherwise, we approximate the curve by a piecewise affine function. For this we subdivide the interval $[a,b]$ into smaller intervals, $[t_{j-1}, t_j]$ and approximate f on this subinterval by the line segment from $(t_{j-1}, f(t_{j-1}))$ to $(t_j, f(t_j))$; see Figure 6.15.

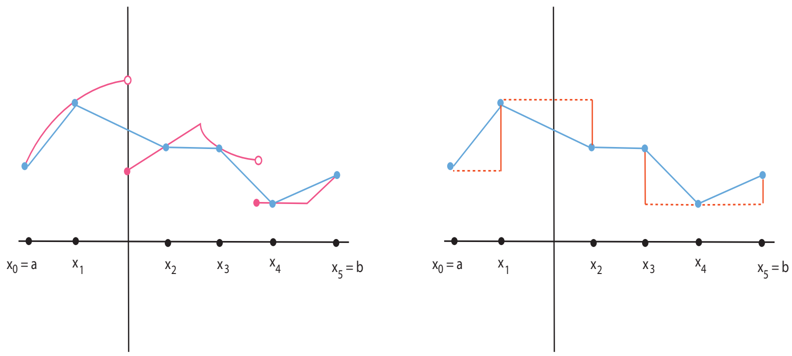


Figure 6.15: The graph of f is represented in pink. The linear approximation from $(t_{j-1}, f(t_{j-1}))$ to $(t_j, f(t_j))$, where $0 \le j \le 5$, is represented in blue. A variation of f over $[a,b]$ is the sum of lengths of the solid orange vertical lines found in the right figure.

More precisely, consider a function $f\colon \mathbb{R} \to \mathbb{C}$. For any $x \in \mathbb{R}$, we consider subdivisions of the interval $(-\infty, x]$ using finite sequences $x_0 < x_1 < \cdots < x_n = x$.

Definition 6.8. Let $f\colon \mathbb{R} \to \mathbb{C}$ be a function. The *total variation function* T_f of f is the function given by

$$T_f(x) = \sup \left\{ \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \mid -\infty < x_0 < x_1 < \cdots < x_n = x,\ n \in \mathbb{N} - \{0\} \right\},$$

where the supremum is taken over all finite subdivisions $x_0 < x_1 < \cdots < x_n = x$. If $[a,b]$ is a finite interval ($a \le b$), then the *total variation of f on* $[a,b]$ is the quantity

$$V(f,a,b) = \sup \left\{ \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \mid a = x_0 < x_1 < \cdots < x_n = b,\ n \in \mathbb{N} - \{0\} \right\}.$$

The set BV of functions of *bounded variation* is the set of functions $f\colon \mathbb{R} \to \mathbb{C}$ such that $\lim_{x \mapsto +\infty} T_f(x) < \infty$. The set $BV([a,b])$ of functions of *bounded variation over* $[a,b]$ is the set of functions $f\colon [a,b] \to \mathbb{C}$ such that $V(f,a,b)$ is finite.

If $f\colon \mathbb{R} \to \mathbb{C}$ is a function in BV, since a can be chosen as a subdivision point we see immediately that

$$V(f, a, b) = T_f(b) - T_f(a),$$

so $f \in BV([a, b])$. We also see that $T_f(x)$ is an increasing function. The limits $\lim_{x \mapsto -\infty} f(x)$ and $\lim_{x \mapsto +\infty} f(x)$ also exist, as a corollary of Proposition 6.9 below.

Example 6.3.

(1) If $f\colon \mathbb{R} \to \mathbb{R}$ is bounded and increasing, then $f \in BV$. In fact, $T_f(x) = f(x) - f(-\infty)$; see Figure 6.16.

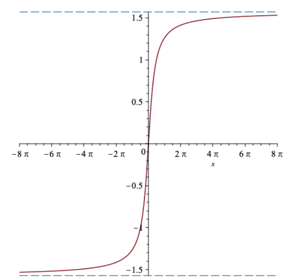


Figure 6.16: Let $f(x) = \tan^{-1}(x)$. Then $f(x)$ is a bounded, increasing funtion whose total variation $T_f(x) = f(x) - f(-\infty) = f(x) + \frac{\pi}{2}$.

(2) The space BV is a complex vector space.

(3) If f is differentiable on $\mathbb{R}$ and if f' is bounded, then $f \in BV([a, b])$ for every finite interval $[a, b]$ (by the mean value theorem); see Figure 6.17.

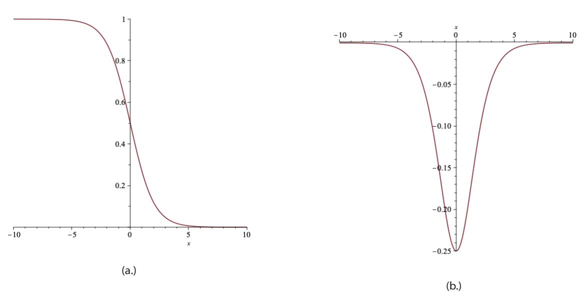


Figure 6.17: Figure (a) is the graph of $f(x) = \frac{1}{e^x+1}$, a bounded decreasing function. Its derivative $f'(x) = -\frac{e^x}{(e^x+1)^2}$, whose graph is shown in Figure (b), is also bounded. Hence $f \in BV([a, b])$ for every finite interval $[a, b]$.

(4) If $f(x) = \sin x$, then $f \in BV([a, b])$ for every finite interval $[a, b]$, but $f \notin BV$ (it oscillates forever); see Figure 6.18.

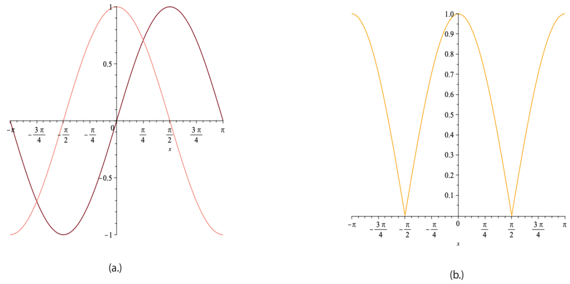


Figure 6.18: In Figure (a), the graph of $f(x) = \sin x$ is the dark red curve, while $f'(x) = \cos x$ is the lighter red curve. The total variation of $f(x)$ on $[\pi, \pi]$ is given by $\int_{-\pi}^{\pi} |f'(x)|\, dx = \int_{-\pi}^{\pi} |\cos(x)|\, dx = 4$ and is visualized as the area under the graph of $y = |\cos(x)|$, as illustrated by Figure (b).

(5) If $f(x) = x \sin(x^{-1})$ for $x \neq 0$ and $f(0) = 0$, then $f \notin BV([a, b])$ for $a \leq 0 < b$ or $a < 0 \leq b$; see Figure 6.19.

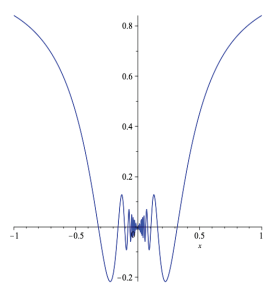


Figure 6.19: The graph of $f(x) = x \sin(x^{-1})$ for $x \neq 0$ and $f(0) = 0$. This function has too much oscillation around 0 for it to be of bounded variation.

Here are some of the main properties of functions of bounded variation; proofs can be found in Folland [34] (Chapter 3, Section 3.5).

Proposition 6.7. *Let $f \colon \mathbb{R} \to \mathbb{C}$ be a function. If $f \in BV$, then* $\lim_{x \mapsto -\infty} T_f(x) = 0$.

Proof. By definition of $T_f(x)$ as a least upper bound, for every $\epsilon > 0$, there is some subdivision $x_0 < x_1 < \cdots < x_n = x$ such that

$$\sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \geq T_f(x) - \epsilon.$$

By definition of $V(f, x_0, x)$, this implies that

$$T_f(x) - T_f(x_0) \geq T_f(x) - \epsilon,$$

so $T_f(x_0) \leq \epsilon$. Since T_f is increasing, we also have $T_f(y) \leq \epsilon$ for all $y \leq x_0$. Since ϵ is arbitrary, we must have $\lim_{x \mapsto -\infty} f(x) = 0$. ☐

Proposition 6.8. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function. If $f \in BV$, then $T_f + f$ and $T_f - f$ are increasing functions.*

Proposition 6.9. *Let $f\colon \mathbb{R} \to \mathbb{C}$ be a function.*

(1) We have $f \in BV$ iff both the real part and the imaginary part of f belong to BV.

(2) If $f\colon \mathbb{R} \to \mathbb{R}$, then $f \in BV$ iff f can be written as the difference of two bounded increasing functions; these can be chosen as $(T_f + f)/2$ and $(T_f - f)/2$. This is called a **Jordan decomposition**.

(3) If $f\colon \mathbb{R} \to \mathbb{C}$ and if $f \in BV$, then the left limit $f(x-)$ and the right limit $f(x+)$ exist for all $x \in \mathbb{R}$, including $x = -\infty$ and $x = +\infty$.

(4) If $f\colon \mathbb{R} \to \mathbb{C}$ and if $f \in BV$, then f has at most countably many discontinuities.

(5) If $f \in BV$ and if we let $g(x) = f(x+)$, then f' and g' exist almost everywhere and are equal almost everywhere.

Remark: The space NBV consists of the functions $f\colon \mathbb{R} \to \mathbb{C}$ in BV which are right continuous and such that $f(-\infty) = 0$. There is a relationship between the space NBV and the complex Borel measures on $\mathbb{R}$. If μ is a complex Borel measure, then the function $F(x) = \mu((-\infty, x])$ is in NBV, and conversely, given any function $f \in NBV$, there is a unique complex measure μ_f such that $f(x) = \mu_f((-\infty, x])$; see Folland [34] (Chapter 3, Section 3.5, Theorem 3.29).

We now return to Fourier series on $\mathbb{T}$ and state the following theorem essentially due to Jordan which generalizes an historically famous result of Dirichlet.

Theorem 6.10. *For any $f \in \mathrm{L}^1(\mathbb{T})$, if $f \in BV([-\pi, \pi])$, then*

$$\lim_{m \mapsto \infty} S_{m,f}(x) = \frac{f(x+) + f(x-)}{2}$$

for all $x \in [-\pi, \pi]$. In particular, $\lim_{m \mapsto \infty} S_{m,f}(x) = f(x)$ whenever f is continuous at x.

Theorem 6.10 is proven in Folland [34] (Chapter 8, Section 8.5, Theorem 8.43).

Other convergence theorems (some about pointwise convergence) are discussed in Folland [32] and Stein and Shakarchi [94].

The behavior of $S_{m,f}$ at a jump discontinuity $x = m$, with $m \in \mathbb{Z}$, (a point x where $f(x) \neq f(x-)$ or $f(x) \neq f(x+)$) is a little strange. It turns out that near an integer value of x, the function $S_{m,f}$ contains spikes that overshoot or undershoot the function f, and when m tends to infinity, the width of the spikes tends to zero but the height does not. This behavior is known as *Gibbs phenomenon*. For example, the function

$$\varphi(x) = 2\pi\left(\frac{1}{2} - (x - \lfloor x \rfloor)\right)$$

(where $\lfloor x \rfloor$ is the greatest integer $\leq x$) is periodic (of period 2π) and exhibits the Gibbs phenomenon. One easily computes the Fourier coefficients, which are

$$c_0 = 0, \quad c_m = \frac{1}{im}, \quad m \neq 0.$$

Then we get

$$S_{m,\varphi}(x) = \sum_{k=1}^{m} \frac{2 \sin kx}{k}.$$

See Figures 6.20 and 6.21.

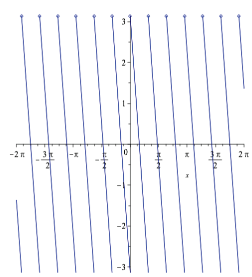


Figure 6.20: The graph of $\varphi(x) = 2\pi\left(\frac{1}{2} - (x - \lfloor x \rfloor)\right)$

For more details see Folland [34] (Chapter 8, Section 8.5).

6.4 The Fourier Transform and the Fourier Cotransform on $\mathbb{T}^n$ and $\mathbb{Z}^n$

In Section 6.1 we introduced the Fourier transform on $\mathbb{T}$ and the Fourier cotransform on $\mathbb{Z}$. In this section we briefly present the generalization to $\mathbb{T}^n = \underbrace{\mathbb{T} \times \cdots \times \mathbb{T}}_{n}$, called the *n-torus*, and to $\mathbb{Z}^n$. As in Section 6.1, a normalized Haar measure on $\mathbb{T}^n$ is $dx_n/(2\pi)^n$, where dx_n the Lebesgue measure on $\mathbb{R}^n$ (so that $\mathbb{T}^n$ has measure 1).

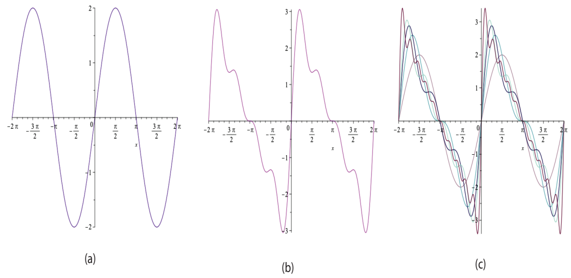


Figure 6.21: Figure (a.) is the graph of $S_{1,\varphi}(x)$, Figure (b.) is the graph of $S_{4,\varphi}(x)$, while Figure (c.) shows the supersition of the graphs of $S_{1,\varphi}(x)$, $S_{2,\varphi}(x)$, $S_{3,\varphi}(x)$, $S_{4,\varphi}(x)$, and $S_{10,\varphi}(x)$.

Recall that given any function $f \in \mathrm{L}^1(\mathbb{T})$, the function $\mathcal{F}(f)\colon \mathbb{Z} \to \mathbb{C}$ given by $\mathcal{F}(f)(m) = c_m$, where c_m is the *Fourier coefficient*

$$c_m = \int_{-\pi}^{\pi} f(t)e^{-imt}\,\frac{dx(t)}{2\pi},$$

is called the *Fourier transform* of f. We identify the sequence $\mathcal{F}(f)$ with the sequence $(c_m)_{m\in\mathbb{Z}}$, which is also denoted by $\widehat{f}$.

Given a sequence $c = (c_m)_{m\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$, we define the *Fourier cotransform* $\overline{\mathcal{F}}(c)$ of c as the function $\overline{\mathcal{F}}(c)\colon \mathbb{T} \to \mathbb{C}$ defined on $\mathbb{T}$ given by

$$\overline{\mathcal{F}}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta},$$

the Fourier series associated with c (with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$). Note that $e^{im\theta}$ is used instead of the term $e^{-im\theta}$ occurring in the Fourier transform.

For symmetry reasons, it seems natural to define a Fourier cotransform on $\mathbb{T}$ and a Fourier transform on $\mathbb{Z}$.

Definition 6.9. The *Fourier cotransform* $\overline{\mathcal{F}}(f)$ of a function $f \in \mathrm{L}^1(\mathbb{T})$ is the $\mathbb{Z}$-indexed sequence $\overline{\mathcal{F}}(f)\colon \mathbb{Z} \to \mathbb{C}$ given by

$$\overline{\mathcal{F}}(f)(m) = \int_{-\pi}^{\pi} f(t)e^{imt}\,\frac{dx(t)}{2\pi},$$

and the *Fourier transform* $\mathcal{F}(c)$ of a sequence $c = (c_m)_{m\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$ is the function $\mathcal{F}(c)\colon \mathbb{T} \to \mathbb{C}$ given by

$$\mathcal{F}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{-im\theta},$$

with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$.

Observe that if $f \in \mathrm{L}^1(\mathbb{T})$, then

$$\overline{\mathcal{F}}(f)(m) = \mathcal{F}(f)(-m) = \overline{\mathcal{F}(\overline{f})(m)},$$

with $m \in \mathbb{Z}$, and if $c \in \ell^1(\mathbb{Z})$, then

$$\overline{\mathcal{F}}(c)(\theta) = \mathcal{F}(c)(-\theta) = \overline{\mathcal{F}(\overline{c})(\theta)}.$$

where $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Thus only one of the two transforms is really needed, but it is convenient to use both (especially in stating Fourier inversion).

Remark: Note a certain asymmetry in the measure chosen on $\mathbb{T}$ and $\mathbb{Z}$. The measure on $\mathbb{T}$ is $dx/2\pi$, so that $\mathbb{T}$ has measure 1, and the measure on $\mathbb{Z}$ is the counting measure.

The main results are:

(1) The spectral synthesis, Theorem 6.2.

(2) The Fourier inversion formula, Theorem 6.3. This result can be expressed as follows. If $f \in \mathrm{L}^1(\mathbb{T})$ and if $\widehat{f} = \mathcal{F}(f) \in \ell^1(\mathbb{Z})$, then

$$f(\theta) = (\overline{\mathcal{F}} \circ \mathcal{F})(f)(\theta) = \sum_{m \in \mathbb{Z}} \widehat{f}_m e^{im\theta}.$$

(3) Plancherel's theorem, Theorem 6.4. This theorem asserts that $\mathcal{F}$ is an isometric isomorphisms between the Hilbert spaces $\mathrm{L}^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$.

All three results stated above generalize to $\mathbb{T}^n$ and $\mathbb{Z}^n$. First we need a bit of notation.

Definition 6.10. A *multi-index* is a sequence $\alpha = (\alpha_1, \ldots, \alpha_n)$ of natural numbers $\alpha_i \in \mathbb{N}$. Define $|\alpha|, \alpha!, \partial^\alpha$, and x^α by

$$|\alpha| = \sum_{i=1}^n \alpha_i, \qquad \alpha! = \alpha_1! \times \cdots \times \alpha_n!, \qquad \partial^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}, \qquad x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}.$$

Example 6.4. For a specific example of Definition 6.10, let $n = 3$ and $\alpha = (\alpha_1, \alpha_2, \alpha_3) = (1, 3, 4)$. Then $|\alpha| = \alpha_1 + \alpha_2 + \alpha_3 = 1 + 3 + 4 = 8$, $\alpha! = \alpha_1!\alpha_2!\alpha_3! = 1!3!4! = 144$, $\partial^\alpha = \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2}\right)^3\left(\frac{\partial}{\partial x_3}\right)^4$, and $x^\alpha = x_1^{\alpha_1}x_2^{\alpha_2}x_3^{\alpha_3} = x_1x_2^3x_3^4$.

We now generalize the Poisson kernel and the Fourier transform (and cotransform) to $\mathbb{T}^n$ and $\mathbb{Z}^n$.

Observe that a function $z\colon \mathbb{Z}^n \to \mathbb{C}$ can be viewed as a $\mathbb{Z}^n$-indexed sequence $z = (z_m)_{m \in \mathbb{Z}^n}$, with $z_m \in \mathbb{C}$.

Example 6.5. To gain some insight into a $\mathbb{Z}^n$-indexed sequence, set $n = 2$ and $z\colon \mathbb{Z}^2 \to \mathbb{C}$. The indices of z are the integer-indexed lattice points of $\mathbb{R}^2$. In particular, if we assume that the nonzero elements of z are entries whose lattice points lie in the closed unit square centered at the origin, z is the finite sequence

$$z = \left(z_{(-1,-1)}, z_{(-1,0)}, z_{(-1,1)}, z_{(0,-1)}, z_{(0,0)}, z_{(0,1)}, z_{(1,-1)}, z_{(1,0)}, z_{(1,1)}\right),$$

where we implicitly made use of the following total ordering for $\mathbb{Z}^2$: given $(i,j),(p,q) \in \mathbb{Z}^2$, $(i,j) < (p,q)$ if either $i < p$ or $i = p$ and $j < q$; see Figure 6.22.

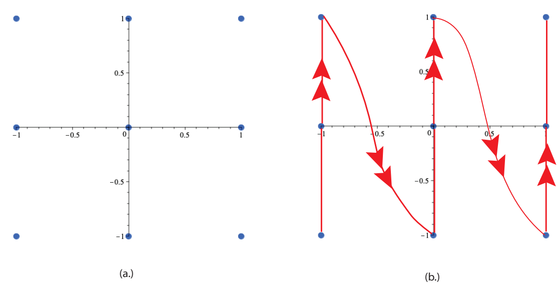


Figure 6.22: Figure (a) illustrates the lattice points in $\mathbb{R}^2$ associated with the $\mathbb{Z}^2$-indexed sequence of Example 6.5. The directed red curve of Figure (b) illustrates the total ordering of $\mathbb{Z}^2$ used in Example 6.5.

Let $\ell^p(\mathbb{Z}^n)$ $(p \geq 1)$ be the space

$$\ell^p(\mathbb{Z}^n) = \left\{ z = (z_m)_{m\in\mathbb{Z}^n},\ z_m \in \mathbb{C} \;\middle|\; \sum_{m\in\mathbb{Z}^n} |z_m|^p < \infty \right\}.$$

As in the case $n = 1$, if $1 \leq p < q$, then $\ell^p(\mathbb{Z}^n) \subseteq \ell^q(\mathbb{Z}^n)$.

We denote the product measure on $\mathbb{T}^n$ by $dx_n/(2\pi)^n = (1/(2\pi)^n)\underbrace{dx \otimes \cdots \otimes dx}_{n}$, where dx is the Lebesgue measure on $\mathbb{R}$. With this measure, $\mathbb{T}^n$ has measure 1.

Definition 6.11. The *Poisson kernel* $P_r(\theta)$ on $\mathbb{T}^n$ (with $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n/2\pi\mathbb{Z}^n$) is the family of functions $P_r(\theta)$, parametrized by $r \in [0,1)$, given by

$$P_r(\theta) = \prod_{k=1}^{n} P_r(\theta_k),$$

with

$$P_r(\theta_k) = \sum_{n=-\infty}^{n=\infty} r^{|n|}\, e^{in\theta_k}.$$

Definition 6.12. For any function $f \in \mathrm{L}^1(\mathbb{T}^n)$, the *Fourier transform* $\widehat{f} = \mathcal{F}(f)$ of f is the $\mathbb{Z}^n$-indexed sequence $\mathcal{F}(f)\colon \mathbb{Z}^n \to \mathbb{C}$ given by

$$\widehat{f}(m) = \mathcal{F}(f)(m) = \int_{\mathbb{T}^n} f(\theta) e^{-im\cdot\theta} \frac{dx_n(\theta)}{(2\pi)^n},$$

and the *Fourier cotransform* $\overline{\mathcal{F}}(f)$ of f is the $\mathbb{Z}^n$-indexed sequence $\overline{\mathcal{F}}(f)\colon \mathbb{Z}^n \to \mathbb{C}$ given by

$$\overline{\mathcal{F}}f(m) = \int_{\mathbb{T}^n} f(\theta) e^{im\cdot\theta} \frac{dx_n(\theta)}{(2\pi)^n},$$

with $m \in \mathbb{Z}^n$, $\theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n$, and with $m \cdot \theta = \sum_{k=1}^n m_k\theta_k$, the inner product of the vectors $m = (m_1, \ldots, m_n)$ and $\theta = (\theta_1, \ldots, \theta_n)$.

For any $c \in \ell^1(\mathbb{Z}^n)$, the *Fourier transform* $\mathcal{F}(c)$ of c is the function $\mathcal{F}(c)\colon \mathbb{T}^n \to \mathbb{C}$ defined on $\mathbb{T}^n$ given by

$$\mathcal{F}(c)(\theta) = \sum_{m\in\mathbb{Z}^n} c_m e^{-im\cdot\theta},$$

and the *Fourier cotransform* $\overline{\mathcal{F}}(c)$ of c is the function $\overline{\mathcal{F}}(c)\colon \mathbb{T}^n \to \mathbb{C}$ defined on $\mathbb{T}^n$ given by

$$\overline{\mathcal{F}}(c)(\theta) = \sum_{m\in\mathbb{Z}^n} c_m e^{im\cdot\theta},$$

with $\theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n$.

Remark: The Fourier cotransform is also called the *inverse Fourier transform* by some authors, including Hewitt and Ross.

It can be shown that $|\widehat{f}(m)|$ tends to zero when $|m|$ tends to infinity. This is a special case of Proposition 10.18.

Definition 6.13. The *convolution* $f * g$ of two functions $f, g \in \mathrm{L}^1(\mathbb{T}^n)$ is given by

$$(f * g)(\theta) = \int_{\mathbb{T}^n} f(\theta - \varphi) g(\varphi) \frac{dx_n(\varphi)}{(2\pi)^n} = \int_{\mathbb{T}^n} f(\varphi) g(\theta - \varphi) \frac{dx_n(\varphi)}{(2\pi)^n},$$

where dx_n is the Lebesgue measure on $\mathbb{R}^n$.

By Proposition 8.48, we have $f * g \in \mathrm{L}^1(\mathbb{T}^n)$.

One of the main reasons why the Fourier transform is useful is that it converts a convolution into a product.

Proposition 6.11. *For any two functions $f, g \in \mathrm{L}^1(\mathbb{T}^n)$, we have*

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g), \qquad \overline{\mathcal{F}}(f * g) = \overline{\mathcal{F}}(f)\overline{\mathcal{F}}(g).$$

The equation $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$ can also be written as $\widehat{f * g} = \widehat{f}\,\widehat{g}$.

Proposition 6.11 actually holds in the more general framework of locally compact abelian groups, and a proof is given in Proposition 10.5 (see also Proposition 10.18).

It is not hard to adapt the proof of Proposition 6.1 to prove that for any $f \in \mathrm{L}^1(\mathbb{T}^n)$, for all $\theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n$, we have

$$(f * P_r)(\theta) = \sum_{m\in\mathbb{Z}^n} \widehat{f}(m) r^{\|m\|_1}\, e^{im\cdot\theta},$$

where $\|m\|_1 = |m_1| + \cdots + |m_n|$. As a consequence we have the following result.

Theorem 6.12. *(Spectral Synthesis)*

(1) If $f \in \mathrm{L}^p(\mathbb{T}^n)$ for $p = 1, 2$ and if $r \in [0, 1)$, for any $\theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n$, write

$$f_r(\theta) = (P_r * f)(\theta) = \sum_{m\in\mathbb{Z}^n} \widehat{f}(m) r^{\|m\|_1}\, e^{im\cdot\theta}.$$

Then $\lim_{r\mapsto 1} \|f - f_r\|_p = 0$.

(2) If $f \in \mathcal{C}(\mathbb{T}^n)$, then $\lim_{r\mapsto 1} \|f - f_r\|_\infty = 0$.

For any $p \in \mathbb{N}$, let

$$S_p = \{m \in \mathbb{Z}^n \mid |m_k| \leq p,\ k = 1, \ldots, n\}.$$

Note that the sequence z of Example 6.5 is the case of S_1 (when $n = 2$).

Recall that the inner product of two functions $f, g \in \mathrm{L}^2(\mathbb{T}^n)$ is given by

$$\langle f, g\rangle = \int_{\mathbb{T}^n} f(\theta)\overline{g(\theta)}\, \frac{dx_n(\theta)}{(2\pi)^n},$$

and the inner product of two sequences $x, y \in \ell^2(\mathbb{Z}^n)$ is given by

$$\langle x, y\rangle = \sum_{m\in\mathbb{Z}^n} x_m\overline{y_m}.$$

Theorem 6.13. *Let $f \in \mathrm{L}^2(\mathbb{T}^n)$. Then we have the equality (Parseval)*

$$\|f\|_2^2 = \sum_{m\in\mathbb{Z}^n} |\widehat{f}(m)|^2.$$

Define $s_p(\theta)$ by

$$s_p(\theta) = \sum_{m\in S_p} \widehat{f}(m)e^{im\cdot\theta}.$$

Then we have

$$\lim_{p\mapsto\infty} \|f - s_p\|_2 = 0.$$

Plancherel's theorem holds.

Theorem 6.14. *(Plancherel) The map $f \mapsto \widehat{f}$ is an isometric isomorphism of the Hilbert spaces $\mathrm{L}^2(\mathbb{T}^n)$ and $\ell^2(\mathbb{Z}^n)$.*

The Fourier inversion formula is generalized as follows.

Theorem 6.15. *(Fourier inversion formula) Let $f \in \mathrm{L}^1(\mathbb{T}^n)$. If $\widehat{f} \in \ell^1(\mathbb{Z}^n)$ then*

$$f(\theta) = \sum_{m \in \mathbb{Z}^n} \widehat{f}_m e^{im\cdot\theta} = (\overline{\mathcal{F}}(\widehat{f}))(\theta),$$

for all almost all $\theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n$. Furthermore, if f is continuous, then equality holds everywhere.

Theorem 6.14 and Theorem 6.15 are proven in Malliavin [68]. They allow the extension of the Fourier cotransform $\overline{\mathcal{F}}$ on $\ell^1(\mathbb{Z}^n)$ to $\ell^2(\mathbb{Z}^n)$ in such a way that $\mathcal{F}$ and $\overline{\mathcal{F}}$ are mutual inverses.

We now turn to the Fourier transform on $\mathbb{R}$.

6.5 The Fourier Transform and the Fourier Cotransform on $\mathbb{R}$

In this section we discuss the Fourier transform of functions defined on the entire real line $\mathbb{R}$ that are not necessarily periodic. Because $\mathbb{R}$ is not compact, $\mathrm{L}^1(\mathbb{R})$ and $\mathrm{L}^2(\mathbb{R})$ are incomparable (with respect to inclusion), and the theory of the Fourier transform on $\mathbb{R}$ is more delicate than the Fourier theory on $\mathbb{T}$. In particular, although Plancherel's theorem holds (Theorem 6.14), its proof is more complicated.

Definition 6.14. For any function $f \in \mathrm{L}^1(\mathbb{R})$, the *Fourier transform* $\widehat{f} = \mathcal{F}(f)$ of f is the function $\mathcal{F}(f)\colon \mathbb{R} \to \mathbb{C}$ defined on $\mathbb{R}$ given by

$$\widehat{f}(x) = \mathcal{F}(f)(x) = \int_{\mathbb{R}} f(y) e^{-iyx}\, \frac{dx(y)}{\sqrt{2\pi}},$$

and the *Fourier cotransform* $\overline{\mathcal{F}}(f)$ of f is the function $\overline{\mathcal{F}}(f)\colon \mathbb{R} \to \mathbb{C}$ defined on $\mathbb{R}$ given by

$$\overline{\mathcal{F}} f(x) = \int_{\mathbb{R}} f(y) e^{iyx}\, \frac{dx(y)}{\sqrt{2\pi}},$$

where dx is the Lebesgue measure on $\mathbb{R}$.

Remark: The Fourier cotransform is also called the *inverse Fourier transform* by some authors, including Hewitt and Ross.

The formula for $\mathcal{F}(f)$ (and $\overline{\mathcal{F}}(f)$) is reminiscent of the formula

$$\mathcal{F}(c)(x) = \sum_{m\in\mathbb{Z}} c_m e^{-imx},$$

where $(c_m)_{m\in\mathbb{Z}}$ is a sequence, except that the infinite sum is replaced by an integral. The integer m is replaced by the real number y, the coefficient c_m is replaced by the value $f(y)$ of the function f at y, and the exponential $e^{-im\theta}$ is replaced by e^{-iyx}. Thus we can view $\mathcal{F}(f)(x)$ as a continuous superposition of the basic periodic functions $y \mapsto e^{-iyx}$. However, this time, $\mathcal{F}(f)(x)$ is not necessarily periodic. We can still think of y as a frequency. In fact, in signal analysis, the domain of the Fourier transform is called the frequency domain.

The reader might be puzzled by the presence of the scale factor $1/\sqrt{2\pi}$. The reason why it is included is that it makes certain formulae more symmetric, for example, the Fourier inversion formula and the Plancherel isomorphism. The deep reason for its need has to do with the fact that the domain of a Fourier transform $\widehat{f}$ is not actually $\mathbb{R}$, but an isomorphic copy $\widehat{\mathbb{R}}$ of $\mathbb{R}$, with a certain measure which is not necessarily identical to the measure on $\mathbb{R}$.

In order for certain results to hold, such as Fourier inversion, if $\mathbb{R}$ is given the Lebesgue measure dx, then $\widehat{\mathbb{R}}$ should be given the measure $dx/2\pi$. Some authors use this normalization. Following Rudin [79, 80], a more symmetric normalization is to use the same scale factor $1/\sqrt{2\pi}$ for both $\mathbb{R}$ and $\widehat{\mathbb{R}}$. Another approach is to incorporate the factor 2π in the exponential; that is, to use $e^{-2\pi iyx}$ instead of e^{-iyx}. In this case, the Lebesgue measure can be used for both $\mathbb{R}$ and $\widehat{\mathbb{R}}$; see Folland [34, 33]. All of this will be elucidated in Chapter 10.

A consequence of using the measure $dx/\sqrt{2\pi}$ is that the *convolution* of two functions $f, g \in \mathrm{L}^1(\mathbb{R})$ is

$$(f * g)(x) = \int_{\mathbb{R}} f(x-y)g(y)\,\frac{dx(y)}{\sqrt{2\pi}} = \int_{\mathbb{R}} f(y)g(x-y)\,\frac{dx(y)}{\sqrt{2\pi}},$$

and the inner product of two functions $f, g \in \mathrm{L}^2(\mathbb{R})$ is given by

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)\overline{g(x)}\,\frac{dx}{\sqrt{2\pi}}.$$

By Proposition 8.48, we have $f * g \in \mathrm{L}^1(\mathbb{R})$.

It is immediately verified that $\overline{\mathcal{F}}(f)(x) = \mathcal{F}(f)(-x) = \overline{\mathcal{F}(\overline{f})(x)}$.

We will now state the most important results about the Fourier theory for $\mathbb{R}$ without proof. Proofs of these results can be found in Folland [34], Rudin [79, 80], Stein and Shakarchi [94], and Malliavin [68]. The most important results will be proven in Chapter 10.

First, following Stein and Shakarchi [94] (Chapter 5 Section 1), observe that there is a nice class $\mathrm{Mod}(\mathbb{R})$ of continuous functions f such that $\mathrm{Mod}(\mathbb{R}) \subseteq \mathrm{L}^1(\mathbb{R})$, and such that the Fourier transform $\widehat{f} = \mathcal{F}(f)$ of f is well-defined.

Definition 6.15. A continuous function $f\colon \mathbb{R} \to \mathbb{C}$ is of *moderate decrease* is there is some $A > 0$ such that

$$|f(x)| \leq \frac{A}{1+x^2} \qquad \text{for all} \in \mathbb{R};$$

see Figure 6.23. The set of functions of moderate decrease is denoted by $\mathrm{Mod}(\mathbb{R})$.

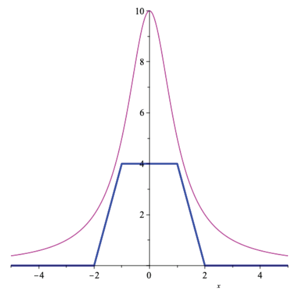


Figure 6.23: The blue real-valued "bump" function is of moderate decrease since it is under magenta graph of $g(x) = \frac{10}{1+x^2}$.

It is shown in Stein and Shakarchi [94] (Chapter 5 Section 1) that $\mathrm{Mod}(\mathbb{R})$ is a vector space contained in $\mathrm{L}^1(\mathbb{R})$, and that the Fourier transform $\widehat{f} = \mathcal{F}(f)$ of every function $f \in \mathrm{Mod}(\mathbb{R})$ is well-defined. However, the Fourier transform $\widehat{f}$ *may not* be of moderate decrease.

Let us give a few examples of Fourier transforms.

Example 6.6.

(1) Let f be the characteristic function $\chi_{[-a,a]}$ of the interval $[-a, a]$, with $a > 0$. Then we have

$$\mathcal{F}(f)(x) = \int_{\mathbb{R}} \chi_{[-a,a]}(y) e^{-iyx} \frac{dy}{\sqrt{2\pi}} = \int_{-a}^{a} e^{-iyx} \frac{dy}{\sqrt{2\pi}} = \frac{e^{-iax} - e^{iax}}{-ix\sqrt{2\pi}} = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x}.$$

Therefore,

$$\mathcal{F}(f)(x) = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x}.$$

We also have

$$\overline{\mathcal{F}}(f)(x) = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x},$$

because

$$\overline{\mathcal{F}}(f)(x) = \int_{\mathbb{R}} \chi_{[-a,a]}(y) e^{iyx} \frac{dy}{\sqrt{2\pi}} = \int_{-a}^{a} e^{iyx} \frac{dy}{\sqrt{2\pi}} = \frac{e^{iax} - e^{-iax}}{ix\sqrt{2\pi}} = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x};$$

see Figure 6.24.

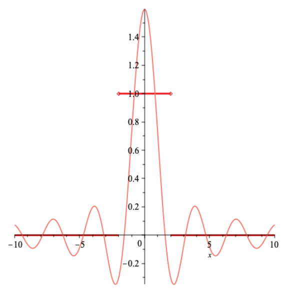


Figure 6.24: The red graph is the plot of $\chi_{[-2,2]}$, while the wavy salmon graph is the plot of $\mathcal{F}(f)(x) = \frac{2}{\sqrt{2\pi}}\frac{\sin 2x}{x} = \overline{\mathcal{F}}(f)(x)$.

Remark: The function sinc is defined by

$$\operatorname{sinc}(x) = \begin{cases} \frac{\sin \pi x}{\pi x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0. \end{cases}$$

Since clearly $\chi_{[-a,a]} \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, by Plancherel theorem (Theorem 6.22), $\widehat{\chi_{[-a,a]}} = \frac{2}{\sqrt{2\pi}}\frac{\sin ax}{x} \in \mathrm{L}^2(\mathbb{R})$, so by setting $a = \pi$ we see that sinc $\in \mathrm{L}^2(\mathbb{R})$. This can also be shown directly by showing that $(\sin \pi x/(\pi x))^2$ is continuous and bounded near zero. However, the function sinc is not in $\mathrm{L}^1(\mathbb{R})$, because

$$\int_{-\infty}^{\infty} \left| \frac{\sin \pi x}{\pi x} \right| dx = \infty;$$

see Figure 6.25. As a consequence, its Fourier transform is undefined. However, by

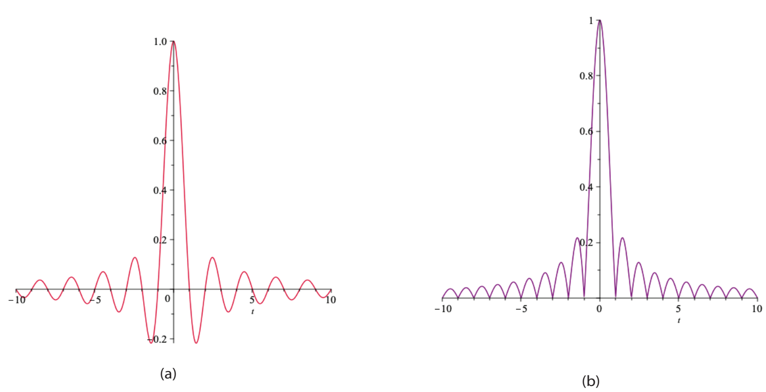


Figure 6.25: Figure (a) is the graph of $\operatorname{sinc}(x)$, while Figure (b) is the graph of $|\operatorname{sinc}(x)|$. Since sinc $\notin \mathrm{L}^1(\mathbb{R})$, the area under the graph of the purple curve is unbounded.

Plancherel's theorem (Theorem 6.22) the Fourier transform has a unique extension to

$L^2(\mathbb{R})$, and the Fourier inversion formula holds. This implies that

$$f(x) = (\overline{\mathcal{F}}(\widehat{f}))(x) = (\mathcal{F}(\mathcal{F}(f))(-x),$$

so

$$(\mathcal{F}(\mathcal{F}(f))(x) = f(-x).$$

Consequently, since

$$\widehat{\chi_{[-\pi,\pi]}} = \sqrt{2\pi}\,\frac{\sin \pi x}{\pi x},$$

the (extended) Fourier transform of sinc is $(1/\sqrt{2\pi})\chi_{[-\pi,\pi]}$.

The function sinc plays a crucial role in the *sampling theorem*, which gives a nice expression for a function $f \in L^2(\mathbb{R})$ which is band-limited, which means that $\widehat{f}(x) = 0$ for all x such that $|x| > a$.

(2) Let f be the function given by

$$f(x) = \frac{y}{x^2 + y^2},$$

with $y > 0$ fixed, and let g be the function given by

$$g(x) = \frac{\pi}{\sqrt{2\pi}} e^{-y|x|};$$

see Figure 6.26. Then we have

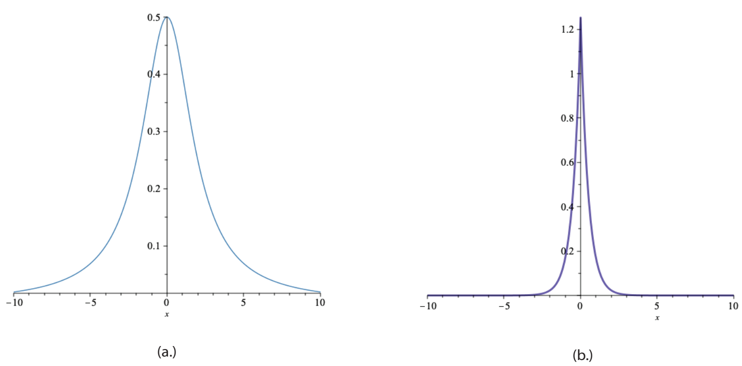


Figure 6.26: Let $y = 2$. Figure (a) is the graph of $f(x) = \frac{2}{x^2+4}$, while Figure (b) is the graph of $g(x) = \frac{\pi}{\sqrt{2\pi}} e^{-2|x|}$.

$$\mathcal{F}(f)(x) = g(x)$$
$$\overline{\mathcal{F}}(g)(x) = f(x).$$

The second formula is proven as follows. Using the fact that $y > 0$, we have

$$\begin{aligned}
\overline{\mathcal{F}}(g)(x) &= \int_{\mathbb{R}} g(t)e^{itx}\, \frac{dt}{\sqrt{2\pi}} \\
&= \int_{-\infty}^{0} \frac{\pi}{\sqrt{2\pi}} e^{yt} e^{itx}\, \frac{dt}{\sqrt{2\pi}} + \int_{0}^{\infty} \frac{\pi}{\sqrt{2\pi}} e^{-yt} e^{itx}\, \frac{dt}{\sqrt{2\pi}} \\
&= \int_{-\infty}^{0} \frac{\pi}{\sqrt{2\pi}} e^{i(x-iy)t}\, \frac{dt}{\sqrt{2\pi}} + \int_{0}^{\infty} \frac{\pi}{\sqrt{2\pi}} e^{i(x+iy)t}\, \frac{dt}{\sqrt{2\pi}} \\
&= \left[\frac{\pi}{2\pi} \frac{e^{i(x-iy)t}}{i(x-iy)}\right]_{-\infty}^{0} + \left[\frac{\pi}{2\pi} \frac{e^{i(x+iy)t}}{i(x+iy)}\right]_{0}^{\infty} \\
&= \frac{1}{2i(x-iy)} - \frac{1}{2i(x+iy)} \\
&= \frac{y}{x^2+y^2}.
\end{aligned}$$

The first formula is harder to prove directly, but it follows from Fourier inversion (see Theorem 6.20).

We now return to the general case of functions in $\mathrm{L}^1(\mathbb{R})$.

Proposition 6.16. *(Riemann–Lebesgue) For any function $f \in \mathrm{L}^1(\mathbb{R})$, the Fourier transform $\widehat{f}$ (and the Fourier cotransform $\overline{\mathcal{F}}(f)$) is continuous and tends to zero at infinity; that is, $\widehat{f} \in \mathcal{C}_0(\mathbb{R};\mathbb{C})$. Furthermore*

$$\|\widehat{f}\|_\infty \leq \|f\|_1 .$$

Proposition 6.16 is proven in Malliavin [68].

As for the Fourier transform on $\mathbb{T}$, the Fourier transform converts a convolution into a product. The following proposition is a special case of Proposition 10.18 and Proposition 10.19, Parts (3) and (4). First we need some notation. For any function $f\colon \mathbb{R}^n \to \mathbb{R}$, for any $y \in \mathbb{R}^n$, the function $\lambda_y(f)\colon \mathbb{R}^n \to \mathbb{R}$ is given by

$$(\lambda_y(f))(x) = f(x-y) \quad \text{for all } x \in \mathbb{R}^n.$$

The above is a special case of Definition 8.7 for the abelian group $\mathbb{R}^n$. The operator λ_y is often called a *translation operator*.

Proposition 6.17. *For any two functions $f, g \in \mathrm{L}^1(\mathbb{R})$, the following properties hold:*

(1) $\widehat{f * g} = \widehat{f}\,\widehat{g}$.

(2) $(\lambda_y(f))\widehat{\ }(x) = e^{-iyx}\widehat{f}(x)$.

(3) $(e^{iyx}f)\widehat{\ }(x) = \lambda_y(\widehat{f})(x)$.

(4) If $\alpha > 0$ and $h(x) = f(x/\alpha)$, then $\widehat{h}(x) = \alpha\widehat{f}(\alpha x)$.

For spectral synthesis in $\mathrm{L}^1(\mathbb{R})$, the Poisson kernel is replaced by a function G_μ defined using the following result.

Proposition 6.18. *For any $\mu > 0$, we have*

$$e^{-\frac{\mu x^2}{2}} = \frac{1}{\sqrt{\mu}} \int e^{-\frac{y^2}{2\mu}}\, e^{ixy}\, \frac{dx(y)}{\sqrt{2\pi}}.$$

Let φ be the function given by

$$\varphi(x) = e^{-\frac{x^2}{2}}.$$

Then $\widehat{\varphi} = \varphi$, and $\varphi(0) = \int \varphi(x)\, dx$.

For a proof of Proposition 6.18, see Rudin [80] (Chapter 7, Lemma 7.6) or Folland [34] (Chapter 8, Section 8.3, Proposition 8.24).

Definition 6.16. The function φ given by

$$\varphi(x) = e^{-\frac{x^2}{2}}$$

is called a *Gauss kernel* or *Weierstrass kernel*.

For any $\mu > 0$, let G_μ be the following function:

$$G_\mu(x) = \frac{1}{\sqrt{\mu}} e^{-\frac{x^2}{2\mu}}.$$

In view of Proposition 6.18 (replacing μ by $1/\mu$) we have

$$G_\mu(x) = \frac{1}{\sqrt{\mu}} e^{-\frac{x^2}{2\mu}} = \int e^{-\frac{\mu y^2}{2}}\, e^{ixy}\, \frac{dx(y)}{\sqrt{2\pi}}.$$

Here is our first result about spectral synthesis analogous to Theorem 6.2(1)-(2). First, we leave it as an exercise to prove that

$$(f * G_\mu)(x) = \int e^{iyx} \widehat{f}(y) e^{-\frac{\mu y^2}{2}}\, \frac{dx(y)}{\sqrt{2\pi}}.$$

Proposition 6.19. *(Spectral Synthesis) Let $f \in \mathrm{L}^1(\mathbb{R})$, let $\widehat{f}$ be its Fourier transform, and for any $\mu > 0$, let*

$$g_\mu(x) = (f * G_\mu)(x) = \int e^{iyx} \widehat{f}(y) e^{-\frac{\mu y^2}{2}}\, \frac{dx(y)}{\sqrt{2\pi}}.$$

If $f \in \mathrm{L}^1(\mathbb{R})$, then

$$\lim_{\mu \mapsto 0} \|f - g_\mu\|_1 = 0,$$

and if $f \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, then

$$\lim_{\mu \mapsto 0} \|f - g_\mu\|_2 = 0.$$

Proposition 6.19 is proven in Malliavin [68] (Chapter 3, Section 2.4, Theorem 2.4.5). The proof uses Fubini's theorem and some technical properties about g_μ that are proven in Malliavin [68].

In general, given a function $f \in \mathrm{L}^1(\mathbb{R})$, the integral

$$\int e^{iyx} \widehat{f}(y) \, \frac{dx(y)}{\sqrt{2\pi}} = (\overline{\mathcal{F}}(\widehat{f}))(x)$$

does not converge. However, for $\mu > 0$, the function $g_\mu(x) = (f * G_\mu)(x)$ is well-defined and when μ tends to 0, the function g_μ is an approximation of f that tends to f.

Now comes our first Fourier inversion theorem analogous to Theorem 6.3.

Theorem 6.20. *(Fourier inversion formula) Let $f \in \mathrm{L}^1(\mathbb{R})$. If $\widehat{f} \in \mathrm{L}^1(\mathbb{R})$, then*

$$f(x) = \int e^{iyx} \widehat{f}(y) \, \frac{dx(y)}{\sqrt{2\pi}} = (\overline{\mathcal{F}}(\widehat{f}))(x),$$

almost everywhere. If f is continuous, the equation holds for all $x \in \mathbb{R}$.

Theorem 6.20 is proven in Rudin [79] (Chapter 9, Theorem 9.11) Folland [34] (Chapter 8, Section 8.3, Theorem 8.26) and Malliavin [68] (Chapter 3, Section 2.4).

Proposition 6.21. *If $f \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, then*

$$\|f\|_2 = \|\widehat{f}\|_2.$$

Proposition 6.21 is proven in Malliavin [68] (Chapter 3, Section 4.2).

Here is the version of Plancherel's theorem for $\mathrm{L}^2(\mathbb{R})$.

Theorem 6.22. *(Plancherel) If $f \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, then $\widehat{f} \in \mathrm{L}^2(\mathbb{R})$. The Fourier transform defined on $\mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$ has a unique extension $\mathcal{F}$ to $\mathrm{L}^2(\mathbb{R})$ which is an isometric isomorphism of the Hilbert space $\mathrm{L}^2(\mathbb{R})$ whose inverse is the (extension of) Fourier cotransform $\overline{\mathcal{F}}$.*

Theorem 6.22 proven in Rudin [79] (Chapter 9, Theorem 9.13) Folland [34] (Chapter 8, Section 8.3, Theorem 8.29) and Malliavin [68] (Chapter 3, Section 2.4)

Although Theorem 6.22 says that the Fourier transform $\mathcal{F}$ extends to an isometric isomorphism of the Hilbert space $\mathrm{L}^2(\mathbb{R})$, this result is useless in practice because for a function $f \in \mathrm{L}^2(\mathbb{R})$ not in $\mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, the extension of $\mathcal{F}$ to f is given by a limit.

The Fourier inversion formula also holds in the following situation.

Proposition 6.23. *(Fourier inversion formula, II) Let $f \in \mathrm{L}^2(\mathbb{R})$. If $\widehat{f} \in \mathrm{L}^1(\mathbb{R})$, then*

$$f(x) = \int e^{iyx} \widehat{f}(y) \frac{dx(y)}{\sqrt{2\pi}} = (\overline{\mathcal{F}}(\widehat{f}))(x),$$

almost everywhere. If f is continuous, the equation holds for all $x \in \mathbb{R}$.

Proposition 6.23 is proven in Rudin [79] (Chapter 9, Theorem 9.14).

Definition 6.17. Let $B(\mathbb{R}) = \{f \in \mathrm{L}^1(\mathbb{R}) \mid \widehat{f} \in \mathrm{L}^1(\mathbb{R})\}$.

Proposition 6.24. *The space $B(\mathbb{R})$ is dense in $\mathrm{L}^1(\mathbb{R})$, $\mathrm{L}^2(\mathbb{R})$, and $\mathcal{C}_0(\mathbb{R};\mathbb{C})$.*

Proposition 6.24 is proven in Malliavin [68] (Chapter 3, Section 2.4).

6.6 The Sampling Theorem

In signal analysis a function $f\colon \mathbb{R} \to \mathbb{C}$ represents a physical signal, and a common problem is to try to reconstruct this signal by sampling it, which means to compute its values at some sequence $t_1 < t_2 < \ldots$ of times. A basic issue is to determine how much information can be gained this way.

It turns out that if the signal f is *band-limited*, which means that its Fourier transform $\widehat{f}$ vanishes outside some interval $[-\Omega, \Omega]$, then f can be completely reconstructed by sampling at the points $t_n = n\pi/\Omega$, for $n \in \mathbb{N}$.

Theorem 6.25. *(Sampling theorem) Suppose that $f \in \mathrm{L}^2(\mathbb{R})$ and that there is some $\Omega > 0$ such that $\widehat{f}(x) = 0$ for all $|x| \geq \Omega$. Then f is completely determined by its values at the points $t_n = n\pi/\Omega$, $n \in \mathbb{N}$. In fact, we have*

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi}.$$

Proof. We follow Folland [32] (Chapter 7, Section 7.3). We can extend $\widehat{f}$ to a periodic function of period 2Ω, and expand it as a Fourier series over the interval $[-\Omega, \Omega]$. For reason of later convenience, we use the index $-n$ instead of n, so we write

$$\widehat{f}(t) = \sum_{n \in \mathbb{Z}} c_{-n} e^{-in\pi t/\Omega}, \qquad (|t| \leq \Omega).$$

By Plancherel's theorem (extended to $\mathrm{L}^2(\mathbb{R})$), $\widehat{f} \in \mathrm{L}^2(\mathbb{R})$, and since $\widehat{f}(t) = 0$ for $|t| \geq \Omega$, by Proposition 5.43, we have $\mathrm{L}^2(\mathbb{R}) \subseteq \mathrm{L}^1(\mathbb{R})$ and so $\widehat{f} \in \mathrm{L}^1(\mathbb{R})$. By adjusting the computation below, we can show that the Fourier coefficients c_{-n} are given by

$$c_{-n} = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} \widehat{f}(t) e^{in\pi t/\Omega}\, dt = \frac{1}{2\Omega} \int_{-\infty}^{\infty} \widehat{f}(t) e^{in\pi t/\Omega}\, dt = \frac{\sqrt{2\pi}}{2\Omega} f\left(\frac{n\pi}{\Omega}\right),$$

where we used Fourier inversion (Theorem 6.23) and the fact that $\widehat{f}(t) = 0$ for $|t| \geq \Omega$. Again, using these two facts we have

$$\begin{aligned} f(t) = \int_{-\Omega}^{\Omega} \widehat{f}(\omega) e^{i\omega t} \frac{d\omega}{\sqrt{2\pi}} &= \int_{-\Omega}^{\Omega} \sum_{n=-\infty}^{\infty} \frac{\sqrt{2\pi}}{2\Omega} f\left(\frac{n\pi}{\Omega}\right) e^{-in\pi\omega/\Omega} e^{i\omega t} \frac{d\omega}{\sqrt{2\pi}} \\ &= \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) e^{i(\Omega t - n\pi)\omega/\Omega)} \, d\omega \\ &= \frac{1}{2\Omega} \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) \left[\frac{e^{i(\Omega t - n\pi)\omega/\Omega)}}{i(\Omega t - n\pi)/\Omega} \right]_{\omega=-\Omega}^{\omega=\Omega} \\ &= \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi}. \end{aligned}$$

Since $\widehat{f} \in \mathrm{L}^2(\mathbb{R})$ the above manipulations are legitimate. $\square$

Observe that variants of the function $\mathrm{sinc}(x) = \frac{\sin \pi x}{\pi x}$ show up.

Theorem 6.25 is due independently to E.T. Whittaker and Shannon (a similar result was published by Kotelnikov).

It worth noting that the functions

$$s_n(t) = \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi}$$

form an orthonormal Hilbert basis for the Hilbert space of functions f in $\mathrm{L}^2(\mathbb{R})$ such that $\widehat{f} = 0$ a.e. outside $(-\Omega, \Omega)$; see Figure 6.27. This is because the computations in the proof of the sampling theorem show that s_n is the Fourier cotransform (inverse Fourier transform) of the function

$$\widehat{s_n}(t) = \begin{cases} \frac{\pi}{\Omega} e^{-in\pi t/\Omega} & \text{if } |t| < \Omega \\ 0 & \text{otherwise.} \end{cases}$$

By Plancherel theorem and the fact that the functions $t \mapsto e^{-in\pi t/\Omega}$ constitute an orthonormal Hilbert basis for $\mathrm{L}^2(-\Omega, \Omega)$, we deduce that the s_n form a Hilbert basis.

From a practical point of view, the expansion of f given by the sampling theorem has the disadvantage that it generally does not converge very rapidly. By oversampling, that is, evaluating f at a more closely spaced sequence of points $n\pi/\lambda\Omega$, with $\lambda > 1$, we can replace the functions s_n by functions $g_\lambda(t - n\pi/\lambda\Omega)$ that vanish like $1/t^2$ when t goes to infinity. The function g_λ is given by

$$g_\lambda(t) = \frac{\cos \Omega t - \cos \lambda \Omega t}{\pi(\lambda - 1)\Omega t^2};$$

see Folland [32] (Chapter 7, Section 7.3, Exercise 8) and Stein and Shakarchi [94] (Chapter 5, Exercise 20). Also see Figure 6.28.

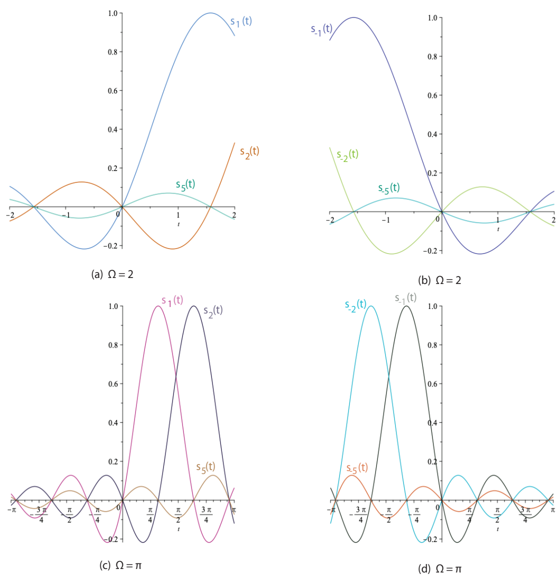


Figure 6.27: The graphs of various $s_n(t)$. Note that $n \to -n$ results in a reflection over the y-axis.

There is also a dual version of the sampling theorem for functions $f \in \mathrm{L}^2(\mathbb{R})$ that vanish outside an interval $[-L, L]$. Then the Fourier transform $\widehat{f}$ of f is determined by sampling at the points $\omega = n\pi/L$, and $\widehat{f}$ is given by the formula

$$\widehat{f}(t) = \sum_{n=-\infty}^{\infty} \widehat{f}\left(\frac{n\pi}{\Omega}\right) \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi},$$

which is obtained from the formula of Theorem 6.25 by replacing f with $\widehat{f}$.

6.7 The Fourier Transform and the Fourier Cotransform on $\mathbb{R}^n$

The generalization of the results of Section 6.5 to $\mathbb{R}^n$ is straightforward.

Definition 6.18. For any function $f \in \mathrm{L}^1(\mathbb{R}^n)$, the *Fourier transform* $\widehat{f} = \mathcal{F}(f)$ of f is the function $\mathcal{F}(f)\colon \mathbb{R}^n \to \mathbb{C}$ defined on $\mathbb{R}^n$ given by

$$\widehat{f}(x) = \mathcal{F}(f)(x) = \int_{\mathbb{R}^n} f(y)e^{-iy\cdot x}\, \frac{dx_n(y)}{(2\pi)^{n/2}},$$

and the *Fourier cotransform* $\overline{\mathcal{F}}(f)$ of f is the function $\overline{\mathcal{F}}(f)\colon \mathbb{R}^n \to \mathbb{C}$ defined on $\mathbb{R}^n$ given by

$$\overline{\mathcal{F}}f(x) = \int_{\mathbb{R}^n} f(y)e^{iy\cdot x}\, \frac{dx_n(y)}{(2\pi)^{n/2}},$$

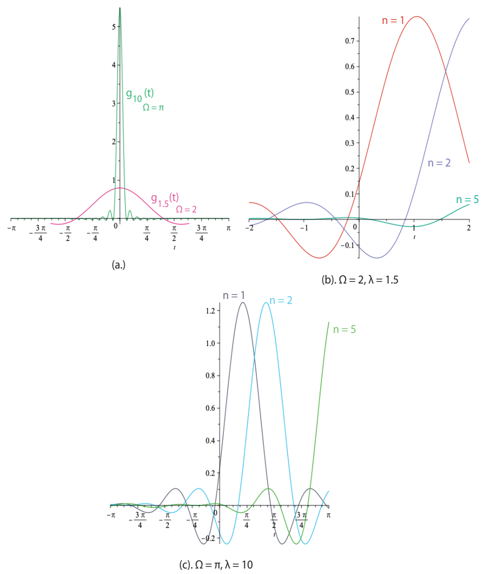


Figure 6.28: Figure (a) shows the graphs of various $g_\lambda(t)$. Figure (b) shows graphs of $g_{1.5}(t)$ when $t \to t - n\pi/\lambda\Omega$ and $\Omega = 2$. Figure (c) shows graphs of $g_{10}(t)$ when $t \to t - n\pi/\lambda\Omega$ and $\Omega = \pi$.

where dx_n is the Lebesgue measure on $\mathbb{R}^n$, and $x \cdot y = \sum_{k=1}^n x_k y_k$ is the inner product of $x, y \in \mathbb{R}^n$.

Remark: The Fourier cotransform is also called the *inverse Fourier transform* by some authors, including Hewitt and Ross.

Again, we are using Rudin's normalization scale factor $1/(2\pi)^{n/2}$, so we are really using the measure $dx_n/(2\pi)^{n/2}$. In particular, the *convolution* of two functions $f, g \in \mathrm{L}^1(\mathbb{R}^n)$ is

$$(f * g)(x) = \int_{\mathbb{R}} f(x-y)g(y)\, \frac{dx_n(y)}{(2\pi)^{n/2}} = \int_{\mathbb{R}} f(y)g(x-y)\, \frac{dx_n(y)}{(2\pi)^{n/2}},$$

and the inner product of two functions $f, g \in \mathrm{L}^2(\mathbb{R}^n)$ is given by

$$\langle f, g\rangle = \int_{\mathbb{R}^n} f(x)\overline{g(x)}\, \frac{dx_n}{(2\pi)^{n/2}}.$$

By Proposition 8.48, we have $f * g \in \mathrm{L}^1(\mathbb{R}^n)$.

It is immediately verified that $\overline{\mathcal{F}}(f)(x) = \mathcal{F}(f)(-x) = \overline{\mathcal{F}(\overline{f})(x)}$.

Proposition 6.26. *(Riemann–Lebesgue) For any function $f \in \mathrm{L}^1(\mathbb{R}^n)$, the Fourier transform $\widehat{f}$ (and the Fourier cotransform $\overline{\mathcal{F}}(f)$) is continuous and tends to zero at infinity; that is, $\widehat{f} \in \mathcal{C}_0(\mathbb{R}^n; \mathbb{C})$. Furthermore*

$$\|\widehat{f}\|_\infty \le \|f\|_1\,.$$

As for the Fourier transform on $\mathbb{T}^n$, the Fourier transform converts a convolution into a product.

Proposition 6.27. *For any two functions $f, g \in \mathrm{L}^1(\mathbb{R}^n)$, the following properties hold for all $x, y \in \mathbb{R}^n$:*

(1) $\widehat{f * g} = \widehat{f}\,\widehat{g}$.

(2) $(\lambda_y(f))\widehat{\ }(x) = e^{-iy\cdot x}\widehat{f}(x)$.

(3) $(e^{iy\cdot x}f)\widehat{\ }(x) = \lambda_y(\widehat{f})(x)$.

(4) If $\alpha > 0$ and $h(x) = f(x/\alpha)$, then $\widehat{h}(x) = \alpha^n \widehat{f}(\alpha x)$.

Proposition 6.27 is proven in Rudin [80] (Chapter 7, Theorem 7.2).

Another useful property of convolution is that under certain conditions it allows differentiation under the integral sign. This property is another regularization feature of convolution. By convolving a function with a "nice" function, we obtain a "nice" function.

Proposition 6.28. *If $f \in \mathrm{L}^1(\mathbb{R}^n)$, $g \in C^k(\mathbb{R}^n)$, and $\partial^\alpha g$ is bounded for all α such that $|\alpha| \leq k$, then $f * g \in C^k(\mathbb{R}^n)$ and*

$$\partial^\alpha(f * g) = f * (\partial^\alpha g), \quad |\alpha| \leq k.$$

See Folland [34] (Proposition 8.10)

For any $\mu > 0$, an for any $x \in \mathbb{R}^n$, let G_μ be the following function:

$$G_\mu(x) = \frac{1}{\mu^{n/2}} e^{-\frac{\|x\|^2}{2\mu}},$$

where $\|x\|^2 = x_1^2 + \cdots + x_n^2$; see Figure 6.29. Using Proposition 6.18, it is easy to see that

$$G_\mu(x) = \int e^{-\frac{\mu\|y\|^2}{2}} e^{ix\cdot y} \frac{dx_n(y)}{(2\pi)^{n/2}}.$$

We also easily verify that

$$(f * G_\mu)(x) = \int e^{iy\cdot x}\widehat{f}(y) e^{-\frac{\mu\|y\|^2}{2}} \frac{dx_n(y)}{(2\pi)^{n/2}}.$$

Proposition 6.29. *(Spectral Synthesis) Let $f \in \mathrm{L}^1(\mathbb{R}^n)$, let $\widehat{f}$ be its Fourier transform, and for any $\mu > 0$, let*

$$g_\mu(x) = (f * G_\mu)(x) = \int e^{iy\cdot x}\widehat{f}(y) e^{-\frac{\mu\|y\|^2}{2}} \frac{dx_n(y)}{(2\pi)^{n/2}}.$$

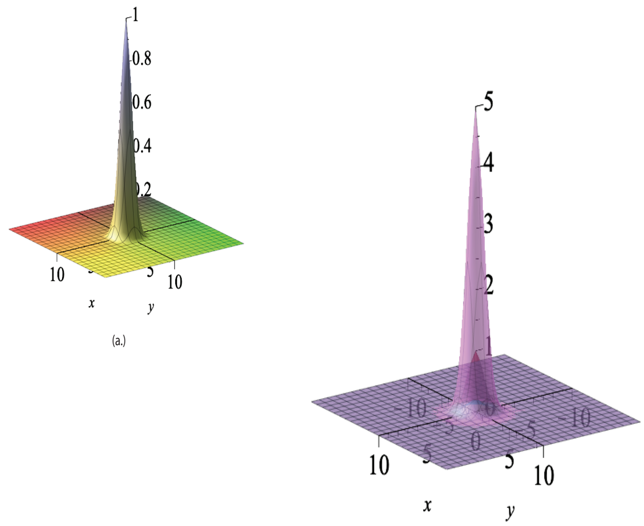


Figure 6.29: Let $n = 2$, Figure (a) is the graph of $G_1(x)$, while Figure (b) shows graph of $G_5(x)$ nested inside the graph of $G_1(x)$, which itself is nested inside $G_{\frac{1}{5}}(x)$. A smaller μ leads to a larger peak above the origin.

If $f \in \mathrm{L}^1(\mathbb{R}^n)$, then

$$\lim_{\mu \mapsto 0} \|f - g_\mu\|_1 = 0,$$

and if $f \in \mathrm{L}^1(\mathbb{R}^n) \cap \mathrm{L}^2(\mathbb{R}^n)$, then

$$\lim_{\mu \mapsto 0} \|f - g_\mu\|_2 = 0.$$

Proposition 6.19 is proven in Malliavin [68] (Chapter 3, Section 2.4).

Theorem 6.30. *(Fourier inversion formula) Let $f \in \mathrm{L}^1(\mathbb{R}^n)$. If $\widehat{f} \in \mathrm{L}^1(\mathbb{R}^n)$, then*

$$f(x) = \int e^{iy\cdot x} \widehat{f}(y) \, \frac{dx_n(y)}{(2\pi)^{n/2}} = (\overline{\mathcal{F}}(\widehat{f}))(x),$$

almost everywhere. If f is continuous, the equation holds for all $x \in \mathbb{R}^n$.

Theorem 6.20 is proven in Rudin [80] (Chapter 7, Theorem 7.7) Folland [34] (Chapter 8, Section 8.3, Theorem 8.26) and Malliavin [68] (Chapter 3, Section 2.4).

Definition 6.19. Let $B(\mathbb{R}^n) = \{f \in \mathrm{L}^1(\mathbb{R}^n) \mid \widehat{f} \in \mathrm{L}^1(\mathbb{R}^n)\}$.

Proposition 6.31. *The space $B(\mathbb{R}^n)$ is dense in $\mathrm{L}^1(\mathbb{R}^n)$, $\mathrm{L}^2(\mathbb{R}^n)$, and $\mathcal{C}_0(\mathbb{R}^n; \mathbb{C})$.*

Proposition 6.24 is proven in Malliavin [68] (Chapter 3, Section 2.4)

Here is the version of Plancherel's theorem for $\mathrm{L}^2(\mathbb{R}^n)$.

Theorem 6.32. *(Plancherel) If $f \in \mathrm{L}^1(\mathbb{R}^n) \cap \mathrm{L}^2(\mathbb{R}^n)$, then $\widehat{f} \in \mathrm{L}^2(\mathbb{R}^n)$. The Fourier transform defined on $\mathrm{L}^1(\mathbb{R}^n) \cap \mathrm{L}^2(\mathbb{R}^n)$ has a unique extension $\mathcal{F}$ to $\mathrm{L}^2(\mathbb{R}^n)$ which is an isometric isomorphism of the Hilbert space $\mathrm{L}^2(\mathbb{R}^n)$ whose inverse is the (extension of) Fourier cotransform $\overline{\mathcal{F}}$.*

Theorem 6.22 proven in Rudin [80] (Chapter 7, Theorem 7.9), Folland [34] (Chapter 8, Section 8.3, Theorem 8.29) and Malliavin [68] (Chapter 3, Section 2.4)

6.8 The Schwartz Space

It turns out that $\mathrm{L}^1(\mathbb{R}^n)$ contains an important subspace $\mathcal{S}(\mathbb{R}^n)$ of rapidly decreasing smooth functions and that the Fourier transform is an isomorphism of this space, whose inverse is the Fourier cotransform. Functions in the space $\mathcal{S}(\mathbb{R}^n)$ and all their derivatives vanish at infinity faster than any power of $\|x\|$, where $\|x\|$ is the Euclidean norm on $\mathbb{R}^n$. Technically, we introduce the following family of norms.

Definition 6.20. A continuous function $f \in \mathcal{C}(\mathbb{R}^n, \mathbb{C})$ is *rapidly decreasing* if for every integer $m \geq 0$, there is some $C > 0$ such that $(1 + \|x\|^2)^m |f(x)|$ remains bounded for all x such that $\|x\| \geq C$. Let $\mathcal{C}_{0,0}(\mathbb{R}^n)$ be the set of rapidly decreasing functions. For every $m \in \mathbb{N}$, define the norm $\|f\|_{m,0}$ by

$$\|f\|_{m,0} = \sup_{x \in \mathbb{R}^n} (1 + \|x\|^2)^m |f(x)|.$$

Observe that Definition 6.20 immediately implies that $\mathcal{C}_{0,0}(\mathbb{R}^n)$ is a subspace of $\mathcal{C}_0(\mathbb{R}^n; \mathbb{C})$. Also, since $(1 + \|x\|^2)^m |f(x)| \leq (1 + \|x\|^2)^{m+1} |f(x)|/(1 + \|x\|^2)$, if $(1 + \|x\|^2)^{m+1} |f(x)|$ is bounded for all x such that $\|x\|$ is large enough, we see that

$$\lim_{\|x\| \mapsto \infty} (1 + \|x\|^2)^m |f(x)| = 0, \qquad \text{for all } m \in \mathbb{N}. \tag{$*$}$$

Conversely, Condition (*) implies that $(1 + \|x\|^2)^{m+1} |f(x)|$ is bounded for all x such that $\|x\|$ is large enough. Therefore, $(*)$ is equivalent to the condition used in Definition 6.20. In view of all this, we have

$$\mathcal{C}_{0,0}(\mathbb{R}^n) = \{f \in \mathcal{C}_0(\mathbb{R}^n; \mathbb{C}) \mid \|f\|_{m,0} < \infty, \text{ for all } m \geq 0\}.$$

Definition 6.21. The *Schwartz space* $\mathcal{S}(\mathbb{R}^n)$ consists of all smooth functions (that is, differentiable at all orders) given by

$$\mathcal{S}(\mathbb{R}^n) = \{f \in \mathcal{C}^\infty(\mathbb{R}^n) \cap \mathcal{C}_{0,0}(\mathbb{R}^n) \mid \partial^\alpha f \in \mathcal{C}_{0,0}(\mathbb{R}^n) \quad \text{for every multi-index } \alpha\}.$$

For all $m, p \in \mathbb{N}$, define the norm $\|f\|_{m,p}$ by

$$\|f\|_{m,p} = \sup_{|\alpha| \leq p} \|\partial^\alpha f\|_{m,0} = \sup_{x \in \mathbb{R}^n, |\alpha| \leq p} (1 + \|x\|^2)^m |\partial^\alpha f(x)|.$$

Definition 6.21 is due to Laurent Schwartz. Observe that when $p = 0$, the norm $\|f\|_{m,p}$ is just the norm $\|f\|_{m,0}$ introduced in Definition 6.20, and that by definition,

$$\mathcal{S}(\mathbb{R}^n) = \{f \in \mathcal{C}_0(\mathbb{R}^n; \mathbb{C}) \mid \|f\|_{m,p} < \infty, \quad \text{for all } m, p \in \mathbb{N}\}.$$

Functions such as $x^k e^{-x^2}$ where $k \in \mathbb{N}$ belong to $\mathcal{S}(\mathbb{R})$; see Figure 6.30. The functions $e^{-c\|x\|^{2m}}$ where m is a positive integer and $e^{-c(1+\|x\|^2)^\alpha}$ with $c > 0$ and $\alpha > 0$ belong to $\mathcal{S}(\mathbb{R}^n)$; see Figures 6.31 and 6.32.

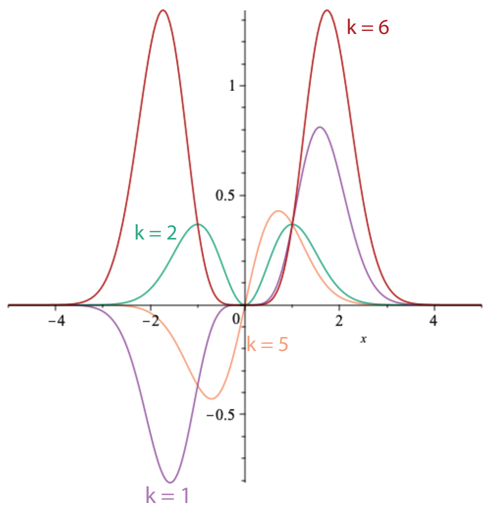


Figure 6.30: Various graphs of $x^k e^{-x^2}$, where k is a nonnegative integer.

Remark: Although it decreases very fast at infinity, the function $x \mapsto e^{-y|x|}$ (with $y > 0$ fixed) does not belong to $\mathcal{S}(\mathbb{R})$, because it is not differentiable at $x = 0$; see the sharp peak at $x = 0$ in Figure 6.26 (b.).

The space $\mathcal{D}(\mathbb{R}^n)$ of smooth functions with compact support is obviously a subspace of $\mathcal{S}(\mathbb{R}^n)$. In Section 5.13 the space $\mathcal{D}(\mathbb{R}^n)$ was denoted $\mathcal{K}^\infty_{\mathbb{C}}(\mathbb{R}^n)$, but the notation $\mathcal{D}(\mathbb{R}^n)$ is more common.

Since the Schwartz space is a subspace of $\mathcal{C}_0(\mathbb{R}^n; \mathbb{C})$, we can make it a normed vector space by giving it the norm $\| \|_\infty$. Unfortunately, with this norm it is *not* complete. We can give it a topology induced by the family of norms $\| \|_{m,p}$, according to the standard process for defining a topology in terms of a family of semi-norms described in Section 2.7. Moreover, because this topology is Hausdorff and the family of norms is countable, $\mathcal{S}(\mathbb{R}^n)$ is actually a metric space; better, a complete metric space.

Let us now use the family of semi-norms $\| \|_{m,p}$ to define a topology on the Schwartz space $\mathcal{S}(\mathbb{R}^n)$. The semi-norms $\| \|_{m,p}$ are actually norms, so by Proposition 2.18, the space $\mathcal{S}(\mathbb{R}^n)$ is Hausdorff.

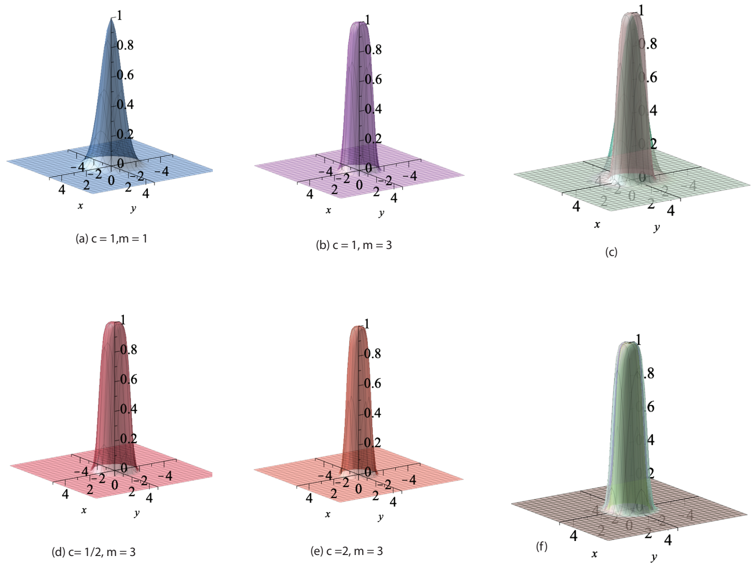


Figure 6.31: Let $n = 2$. Figure (a) is the graph of $e^{-\|x\|^2}$, while Figure (b) is the graph of $e^{-\|x\|^6}$. Figure (c) is the juxtaposition of these two graphs and shows for fixed c, as m increases, the peak becomes wider. Figure (d) is $e^{-\frac{1}{2}\|x\|^6}$, while Figure (e) is $e^{-2\|x\|^6}$. Figure (f) is the juxtaposition of Figures (b), (e), and (d), and shows that for fixed m, as c increases, the peak becomes thinner.

Definition 6.22. The vector space $\mathcal{S}(\mathbb{R}^n)$ endowed with the topology induced by the countable family of norms $\| \, \|_{m,p}$ is a Hausdorff space called the *topological Schwartz space*.

We usually omit the word topological in topological Schwartz space. The value of the topology defined above is that $\mathcal{S}(\mathbb{R}^n)$ is complete.

Theorem 6.33. *The topological Schwartz space $\mathcal{S}(\mathbb{R}^n)$ is a Fréchet space; that is, it is complete for the metric given by*

$$d(x, y) = \sum_{m=0,p=0}^{\infty} \frac{1}{2^{m+p}} \frac{\|y - x\|_{m,p}}{1 + \|y - x\|_{m,p}}.$$

The space $\mathcal{D}(\mathbb{R}^n)$ of smooth functions with compact support is dense in $\mathcal{S}(\mathbb{R}^n)$.

Note that the above metric is the metric used in Proposition 2.20. Theorem 6.33 is proven in Rudin [80] (Chapter 7, Theorem 7.4 and Theorem 7.10).

The following result is proven in Malliavin [68] using the technique of regularization by some suitable convolution (Chapter 3, Section 3.2, Proposition 3.2.4).

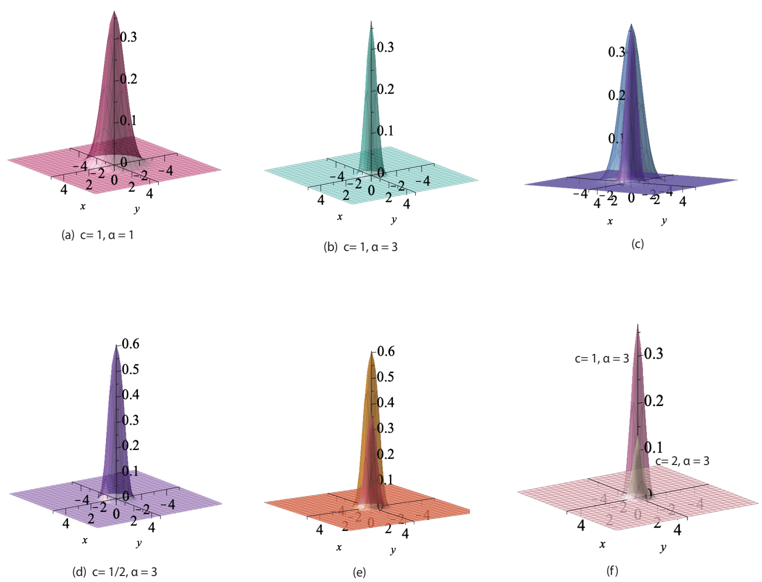


Figure 6.32: Let $n = 2$. Figure (a) is the graph of $e^{-(1+\|x\|^2)}$, while Figure (b) is the graph of $e^{-(1+\|x\|^2)^3}$. Figure (c) is the juxtaposition of these two graphs and shows for fixed c, as α increases, the peak becomes narrower. Figure (d) is $e^{-\frac{1}{2}(1+\|x\|^2)^3}$, while Figure (e) is the juxtaposition of Figures (b) and (d), and shows that for fixed α, as c increases, the peak becomes shorter and narrower. This phenomenon is also see in Figure (f), which is the juxtaposition of the graphs of $e^{-(1+\|x\|^2)^3}$ and $e^{-2(1+\|x\|^2)^3}$.

Proposition 6.34. *The space $\mathcal{D}(\mathbb{R}^n)$ of smooth functions with compact support is dense in $\mathrm{L}^1(\mathbb{R}^n)$ and $\mathrm{L}^2(\mathbb{R}^n)$ (with the Lebesgue measure). As a corollary, the Schwartz space $\mathcal{S}(\mathbb{R}^n)$ is dense in $\mathrm{L}^1(\mathbb{R}^n)$ and $\mathrm{L}^2(\mathbb{R}^n)$.*

The Fourier theory of $\mathcal{S}(\mathbb{R}^n)$ is particularly nice because the Fourier transform is a map from $\mathcal{S}(\mathbb{R}^n)$ to itself. The following results can be shown.

Theorem 6.35. *If f is any function in $\mathcal{S}(\mathbb{R}^n)$, then the following properties hold:*

(1) We have $\widehat{f} = \mathcal{F}(f) \in \mathrm{L}^1(\mathbb{R}^n)$ and Fourier inversion holds:

$$f(x) = \int \widehat{f}(y)e^{iy\cdot x}\,\frac{dx_n(y)}{(2\pi)^{n/2}} = (\overline{\mathcal{F}}(\widehat{f}))(x).$$

(2) Actually, $\widehat{f} \in \mathcal{S}(\mathbb{R}^n)$ and there exist constants $c_{r,s}$ such that

$$\|\widehat{f}\|_{r,s} \leq c_{r,s}\, \|f\|_{m+s,r}\,, \quad m > n.$$

(3) *The map $f \mapsto \widehat{f} = \mathcal{F}(f)$ is an algebra isomorphism and a homeomorphism from $\mathcal{S}(\mathbb{R}^n)$ to itself whose inverse is $\overline{\mathcal{F}}$, under both algebra structures given by pointwise multiplication and convolution.*

(4)
$$\widehat{x_k f(x)} = i\frac{\partial}{\partial x_k}\widehat{f}(x).$$

(5)
$$\widehat{\left(\frac{\partial}{\partial x_k} f\right)}(x) = ix_k\widehat{f}(x).$$

(6) *If $f, g \in \mathcal{S}(\mathbb{R}^n)$, then $fg \in \mathcal{S}(\mathbb{R}^n)$ and $\widehat{fg} = \widehat{f} * \widehat{g}$.*

(7) *If $f, g \in \mathcal{S}(\mathbb{R}^n)$, then $f * g \in \mathcal{S}(\mathbb{R}^n)$ and $\widehat{f * g} = \widehat{f}\,\widehat{g}$.*

Theorem 6.35 is proven Malliavin [68] (Chapter 3, Section 4, Theorem 4.2) and Rudin [80] (Chapter 7, Sections 7.3 pages 184-189). Parts of it are also proven in Folland [34] (Chapter 8, Section 8.3).

Equation (5) is a small miracle since it says that the Fourier transform of a derivative acts as multiplication of the Fourier transform by ix_k, and it can be used to solve certain partial differential equations. Several examples of this technique are presented in Folland [32] and Stein and Shakarchi [94]. We give an example involving the heat equation.

Consider a region of the plane. Given an initial heat distribution, we are interested in finding the temperature $u(x, y, t)$ of the point (x, y) at time t. Using Newton's law of cooling, it can be shown that u satisfies the *partial differential equation* called the *time-dependent heat equation*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\sigma}{\kappa}\frac{\partial u}{\partial t};$$

see Stein and Shakarchi [94] (Chapter 1) or Folland [34] (Section 8.7). After a long period of time, there is no more heat exchange, so that the system reaches a thermal equilibrium, and then $\frac{\partial u}{\partial t} = 0$. In this case, u depends only on x and y, and the time-dependent equation reduces to the *steady-state heat equation*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{1}$$

The expression Δu on the left-hand side of (1) is the *Laplacian* of u. Suppose our domain is the upper half plane

$$\mathbb{R}^2_+ = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}.$$

We would like to find a (the?) solution $u(x, y)$ of the above equation, given that

$$u(x, 0) = f(x) \tag{2}$$

on the boundary, where f is some given function (in $\mathcal{S}(\mathbb{R})$).

The method for finding the solution u proceeds in two steps.

Step 1. The first trick is to apply the Fourier transform with respect to x to both (1) and (2). We assume that $u \in \mathcal{S}(\mathbb{R}^2)$ even though the solution is only defined on the closure of the upper half plane and may not be extendable to a function in $\mathcal{S}(\mathbb{R}^2)$. The goal of this step is to show that u must be given in terms of a convolution defined on the upper-half plane. After guessing a solution using this step and the next, it is still necessary to prove that it works.

In view of Equation (5) of Theorem 6.35, we get the two equations

$$\begin{aligned} -x^2\widehat{u}(x,y) + \frac{\partial^2\widehat{u}}{\partial y^2}(x,y) &= 0 \\ \widehat{u}(x,0) &= \widehat{f}(x). \end{aligned}$$

Observe that we now have a much simpler problem, namely an *ordinary differential equation* with respect to y in the unknown $\widehat{u}(x,y)$. The solution of the first equation is well-known:

$$\widehat{u}(x,y) = C_1(x)e^{|x|y} + C_2(x)e^{-|x|y},$$

with

$$C_1(x) + C_2(x) = \widehat{f}(x).$$

Since the first term has exponential increase, it has to be discarded (because we are seeking solutions in $\mathcal{S}(\mathbb{R})$), so we must have $C_1 = 0$, and we get

$$\widehat{u}(x,y) = \widehat{f}(x)e^{-|x|y}. \tag{$\dagger$}$$

Step 2. The second trick is that if we can find the Fourier cotransform (inverse Fourier transform) $x \mapsto \mathcal{P}_y(x)$ of $x \mapsto e^{-|x|y}$, since $\mathcal{F}(\mathcal{P}_y)(x) = e^{-|x|y}$, we have (by Proposition 6.17(1)),

$$\widehat{u}(x,y) = \widehat{f}(x)e^{-|x|y} = \mathcal{F}(f)(x)\mathcal{F}(\mathcal{P}_y)(x) = \mathcal{F}(f * \mathcal{P}_y)(x), \quad y > 0.$$

Note that for y fixed, $\mathcal{P}_y \notin \mathcal{S}(\mathbb{R})$, but $\mathcal{P}_y \in \mathrm{L}^1(\mathbb{R})$, so $f * \mathcal{P}_y \in \mathrm{L}^1(\mathbb{R})$ for y fixed. Since $\widehat{u}(x,y) \in \mathcal{S}(\mathbb{R}) \subseteq \mathrm{L}^1(\mathbb{R})$ for $y > 0$ fixed, by Fourier inversion (Theorem 6.20) we deduce that

$$u(x,y) = (f * \mathcal{P}_y)(x) \quad \text{for all } x \in \mathbb{R} \text{ and all } y > 0.$$

But we showed in Example 6.6(2) that the Fourier cotransform (inverse Fourier transform) of

$$g(x) = \frac{\pi}{\sqrt{2\pi}}e^{-y|x|}$$

(with $y > 0$) is

$$f(x) = \frac{y}{x^2 + y^2},$$

so the Fourier cotransform (inverse Fourier transform) of $x \mapsto e^{-|x|y}$ is

$$\mathcal{P}_y(x) = \frac{\sqrt{2\pi}}{\pi} \frac{y}{x^2 + y^2},$$

Therefore we obtain the solution

$$u(x, y) = (f * \mathcal{P}_y)(x).$$

Explicitly, we have

$$u(x, y) = \int_{\mathbb{R}} \frac{y f(x - t)}{\pi(x^2 + y^2)} \, dt,$$

which is called the *Poisson integral formula*, and the function

$$\mathcal{P}_y(x) = \frac{\sqrt{2\pi}}{\pi} \frac{y}{x^2 + y^2}$$

is called the *Poisson kernel* for the upper half plane (there are variants of $\mathcal{P}_y(x)$ with different constants).

To be honest, we still need to check carefully that $u(x, y) = (f * \mathcal{P}_y)(x)$ is indeed a solution of the problem. For this we use Proposition 6.28. It can be shown that $\Delta u = 0$ on $\mathbb{R}^2_+$, but $u(x, 0)$ may not be equal to $f(x)$ on the boundary. What we can claim is that $u(x, y)$ tends to $f(x)$ uniformly as y tends to 0. For details see Folland [34] (Theorem 8.53) and Stein and Shakarchi [94] (Chapter 5, Theorem 2.6).

Various other problems involving the wave equation or the heat equation can be solved using the above method; see Stein and Shakarchi [94] and Folland [32].

It turns out that if we use certain kinds of generalized functions, called *distributions*, then we can apply a more general version of Theorem 6.35 and obtain more general solutions for various partial differential equations.

6.9 The Poisson Summation Formula

Given a function $f \in \mathcal{S}(\mathbb{R})$ it is sometimes desirable to make a periodic function from f. One way to do this is to define the function F_1 as follows.

Definition 6.23. Given a function $f \in \mathcal{S}(\mathbb{R})$, the function $F_1 \colon \mathbb{R} \to \mathbb{C}$ is given by

$$F_1(x) = \sum_{n \in \mathbb{Z}} f(x + 2\pi n).$$

Since $f \in \mathcal{S}(\mathbb{R})$, the series converges absolutely and uniformly on every compact subset of $\mathbb{R}$, so F_1 is continuous. It is also clear that

$$F_1(x) = F_1(x + 2\pi n), \quad n \in \mathbb{Z},$$

so F_1 is indeed periodic. We call F_1 the *periodization* of f.

There is another way to make f periodic, which is to use the sequence of numbers $(\widehat{f}(n))_{n\in\mathbb{Z}}$ and to make the Fourier series from it,

$$F_2(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{inx}.$$

Again, since $f \in \mathcal{S}(\mathbb{R})$, the sum converges absolutely and uniformly since $\widehat{f} \in \mathcal{S}(\mathbb{R})$, so F_2 is continuous. The remarkable fact is that $F_1 = F_2$.

Theorem 6.36. *(Poisson summation formula) For any function $f \in \mathcal{S}(\mathbb{R})$, we have*

$$\sum_{n\in\mathbb{Z}} f(x + 2\pi n) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{inx}.$$

In other words, the Fourier coefficients of $F_1(x) = \sum_{n\in\mathbb{Z}} f(x + 2\pi n)$ are the numbers $\widehat{f}(n)$. In particular,

$$\sum_{n\in\mathbb{Z}} f(2\pi n) = \sum_{n\in\mathbb{Z}} \widehat{f}(n).$$

The proof of Theorem 6.36 can be found in Stein and Shakarchi [94] (Chapter 5, Theorem 3.1). It consists in computing the Fourier coefficients of F_1.

Theorem 6.36 also holds when both f and $\widehat{f}$ are of moderate decrease (recall Definition 6.15).

Remark: There is a relationship between the Poisson kernel on the unit disk,

$$P_r(\theta) = \sum_{n=-\infty}^{n=\infty} r^{|n|}e^{in\theta} = \frac{1 - r^2}{1 - 2r\cos\theta + r^2},$$

and the Poisson kernel on the upper-half plane,

$$\mathcal{P}_y(x) = \frac{\sqrt{2\pi}}{\pi}\frac{y}{x^2 + y^2},$$

obtained by applying the Poisson summation formula to $f(x) = \mathcal{P}_y(x)$ and $\widehat{f}(\theta) = e^{-|\theta|y}$. We get

$$\sum_{n\in\mathbb{Z}} \mathcal{P}_y(x + 2\pi n) = \sum_{n\in\mathbb{Z}} e^{-|n|y}e^{inx} = \sum_{n\in\mathbb{Z}} (e^{-y})^{|n|}e^{inx} = P_{e^{-y}}(x).$$

In summary, (for $y > 0$), we have

$$P_{e^{-y}}(x) = \sum_{n\in\mathbb{Z}} \mathcal{P}_y(x + 2\pi n).$$

6.10 The Heisenberg Uncertainly Principle

A fundamental fact about Fourier series is that it is *impossible* for a nonzero function $f \in \mathrm{L}^2(\mathbb{R})$ that both f and its Fourier transform $\widehat{f}$ vanish outside of some finite interval. This can be shown easily using some elementary complex analysis; see Folland [32] (Chapter 7).

There is an even stronger limitation. Roughly, f and $\widehat{f}$ can't be both highly localized. A precise way to state this fact is to define the notion of dispersion.

Definition 6.24. For any function $f \in \mathrm{L}^2(\mathbb{R})$, the *dispersion* of f about the point a is given by

$$\Delta_a f = \int (x-a)^2 |f(x)|^2\, dx \Big/ \int |f(x)|^2\, dx.$$

Then we have the following theorem.

Theorem 6.37. *(Heisenberg inequality) Let f be a function in $\mathrm{L}^2(\mathbb{R})$. Then for all $a, b \in \mathbb{R}$, we have*

$$(\Delta_a f)(\Delta_b \widehat{f}) \geq \frac{1}{4}.$$

Theorem 6.37 is proven in Stein and Shakarchi [94] in the case where $f \in \mathcal{S}(\mathbb{R})$ (Chapter 5, Section 4), and in Folland [32] (Chapter 7) in a more general situation.

Theorem 6.37 has an interpretation in quantum mechanics (we apologize to those who are familiar with quantum mechanics for the vagueness of our comments). In quantum mechanics, among other things, one studies the motion of particles. For this, we need to know the position and the momentum of the particle, but these are not known exactly, but instead described in terms of probabilities. For simplicity, assume that we are dealing with an electron that travels along the real line. There is a function ψ, called a state function (or wave function), which we assume to be in $\mathcal{S}(\mathbb{R})$, normalized so that

$$\int |\psi(t)|^2\, dt = 1,$$

such that the probability that the electron is located in the interval $[a, b]$ is

$$\int_a^b |\psi(t)|^2\, dt.$$

The *expectation* of where the particle might be is the best guess of the position of the particle, and it is given by

$$\overline{x} = \int_{\mathbb{R}} t|\psi(t)|^2\, dt.$$

The *uncertainty* attached to our expectation, or *variance*, is given by the quantity

$$\Delta_{\overline{x}} \psi = \int_{\mathbb{R}} (t - \overline{x})^2 |\psi(t)|^2\, dt.$$

By differentiating under the integral sign with respect to a, we can show that the expectation $\overline{x}$ is the choice of a that minimizes the variance $\int_{\mathbb{R}}(t-a)^2|\psi(t)|^2\,dt$.

Now in quantum mechanics the momentum ξ of the particle is determined by the Fourier transform $\widehat{\psi}$ of ψ, in the sense that the probability that the electron has momentum ξ in the interval $[a,b]$ is

$$\int_a^b |\widehat{\psi}(t)|^2\,dt.$$

As above, we also have the expectation

$$\overline{\xi} = \int_{\mathbb{R}} t|\widehat{\psi}(t)|^2\,dt,$$

and the variance

$$\Delta_{\overline{\xi}}\widehat{\psi} = \int_{\mathbb{R}} (t-\overline{\xi})^2|\widehat{\psi}(t)|^2\,dt.$$

Theorem 6.37 states that

$$(\Delta_{\overline{x}}\psi)(\Delta_{\overline{\xi}}\widehat{\psi}) \geq \frac{1}{4},$$

which is the *Heisenberg uncertainty principle*. Intuitively, it says that the more certain we are about the position of the particle, the less certain we are about its momentum, and vice versa. Actually, we have ignored units of measurements, and in fact Planck's constant $\hbar$ should be inserted, so the physically correct statement of Heisenberg uncertainty principle is that

$$(\Delta_{\overline{x}}\psi)(\Delta_{\overline{\xi}}\widehat{\psi}) \geq \frac{\hbar}{4}.$$

For more details, see Stein and Shakarchi [94] (Chapter 5, Section 4), and Folland [32] (Chapter 7), and for even more, any text on quantum mechanics.

6.11 Fourier's Life; a Brief Summary

Joseph Fourier was born on March 21, 1768, in Auxerre, a town in northern Burgundy, France, and died in 1830. Because he was 21 during the French revolution (1789), he had a particularly exciting life. In this section we give a very condensed summary of his life, based on the wonderful account in Chapter 1 of Kahane [52].

Fourier's family was poor. At age 10 Fourier had already lost his mother and his father. The organist of the cathedral had noticed that Fourier was exceptionally gifted, so he arranged to have him attend the military college in Auxerre. Teaching was provided by Benedictine monks. Fourier fell in love with mathematics through the writings of Bézout and Clairaut. He worked very hard and completed his studies early at age 14. It was arranged that he stayed in the college, in preparation for starting teaching there at age 16. He already sent some papers on locating the roots of algebraic equations to the Institut, which were

noticed by Legendre. Legendre requested that Fourier join the army in the artillery (the most scientific branch), but his request was denied because Fourier was not a "noble" (he was of humble extraction). So in 1787 he entered the benedictine abbey of Saint-Benoit in Fleury in preparation for becoming a monk.

Fourier stayed in Fleury until 1789, where he taught mathematics. He was going to become a monk on November 5, 1798, but the revolution had taken place and put a hold on new religious positions on November 2. Fourier never became a monk!

Between 1789 and 1793 Fourier continues working on mathematics, but also gets involved in the revolution. He is involved in the supply of food and weapons to Orléans, and being a good politician, does a very good job at that.

In 1794 he is sent to Paris where a new school called "École Normale" has been created. There he meets other mathematicians such as Laplace and Monge. The "École Polytechnique" is created in 1794, and Fourier teaches there between 1795 and 1798.

Apparently, Fourier is noticed by Napoleon, and he follows Napoleon for the expedition to Egypt. There, Napoleon creates a replica of the "Institut de France," headed by Monge, and with Fourier as "perpetual secretary." So Fourier becomes an archeologist.

Napoleon goes back to France where he proclaims himself emperor. Still in Egypt, Fourier negotiates the retreat of the French defeated by the British. Fourier returns to France in 1801. At his return Napoleon charges Fourier with the important administrative position of "préfet" (sort of superintendent) of the department of Isère. France was divided in 90 departments (districts), and the main city in Isère is Grenoble. One might think that this would signal the end of Fourier's mathematical life, but not at all. Fourier was also an astute politician, and a good administrator, so he excelled at everything he did. He started working on his theory of heat propagation.

In 1807 he submitted a paper on this subject to the Institut. Lagrange, Laplace, Lacroix, and Monge were the referees. Lagrange felt that the paper was not rigorous enough, and the paper was rejected. The topic of heat propagation was then proposed for a competition. Fourier reworked his paper which was submitted in 1811, and this time the same referees awarded him the price. However, the commentaries, although they praised the originality of the work, especially the heat equation, pointed out some lack of rigor.

Fourier continued to work on a major manuscript on the analytic theory of heat, but this manuscript was not published until 1822.

In the meantime, Napoleon abdicated in 1815. Life is hectic. Fourier is opposed to Napoleon III. Although he is promoted as préfet of the Rhone, he resigned from this position and returns to Paris. There, with the help of a former student, he finds the position of director of the bureau of statistics! He is elected at the Academy of Sciences in 1817.

Fourier continues working on his book on the analytic theory of heat, but also does some work in statistics. In 1822 he finally publishes his book, *Théorie anaytique de la chaleur*,

which includes all of his work on the subject, starting with his work between 1807 and 1811, and then 1816, 1821.

Laplace, Monge, Liouville, Dirichlet, Navier, Sturm had great respect for Fourier. However, Poisson and Cauchy, who were his rivals, were not his friends. The obituaries by Arago and Cousin did not do justice to Fourier's work. It is sad that the collected works of Fourier were never gathered and published. Darboux collected some of Fourier's papers, but ignored all his work on what is now called linear programming, saying that Fourier attributed an exaggerated importance to this type of work. After all, Fourier adapted Gaussian elimination to linear inequalities (Fourier–Motzkin elimination).

However, Fourier's work had a tremendous influence in mathematics, physics, and engineering, so even if he did not get the recognition that he deserved from his peers, the public voted with their feet.

We must conclude with a famous note of Jacobi to Legendre, sent on July 2, 1831, after Fourier' death.

Fourier deeply believed that the main goal of mathematics was to provide a clear explanation of natural phenomena. In his is book he writes:

"L'étude approfondie de la nature est la source la plus féconde des découvertes mathématiques."

Jacobi (1804-1851) complains to Legendre that Poisson included in a report that Fourier made the reproach to Abel and Jacobi that they did not work enough on the theory of heat, but instead on number theory. Jacobi says:

"... mais un philosophe comme lui aurait dû savoir que le but unique de la science, c'est l'honneur de l'esprit humain, et que, sous ce titre, une question de nombres vaut autant qu'une question du système du monde."

Roughly translated: But such a philosopher should have known that the unique goal of science is the honor of the human spirit, and that, as such, a question about numbers is as worthy as a question about the system of the world.

A very complete account of the mathematical history of Fourier series and its influence on mathematics can be found in the captivating book by Kahane and Lemarié–Rieusset [52].

Chapter 7

Radon Functionals and Radon Measures on Locally Compact Spaces

After having considered a very general theory of integration of functions defined on an arbitrary measure space and taking their values in any Banach space, we turn to the special case of complex-valued or real-valued functions defined on a locally compact space X. This corresponds to measure spaces $(X, \mathcal{B}, \mu)$, where X is a locally compact space, $\mathcal{B}$ is the σ-algebra of Borel sets (which is the smallest σ-algebra containing the open subsets of X), and μ is any (positive) measure on $\mathcal{B}$, which we call a *Borel measure*.

The theme of this chapter is that *a Borel measure μ can be used to define linear forms on various function spaces*. For example, pick the space $\mathcal{K}_{\mathbb{C}}(X)$ of continuous functions on X with compact support. For every function $f \in \mathcal{K}_{\mathbb{C}}(X)$ we can compute the integral

$$\varphi_\mu(f) = \int f d\mu.$$

We have to check that functions in $f \in \mathcal{K}_{\mathbb{C}}(X)$ are integrable, which is indeed true if $\mu(K)$ is finite for every compact subset. We obtain a map $\varphi_\mu \colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$, and since the integral is a linear operator, the map φ_μ is linear. In general it is not continuous, but it satisfies some weaker continuity properties. It is also a positive map, which means that $\varphi_\mu(f) \geq 0$ for every positive function $f \geq 0$.

What F. Riesz and J. Radon and discovered is that, in some sense to be made precise, a special class of Borel measures is in one-to-one correspondence with the positive linear forms on the space $\mathcal{K}_{\mathbb{C}}(X)$. This means that for every positive linear form Φ on $\mathcal{K}_{\mathbb{C}}(X)$, there is a (unique) Borel measure m_Φ with some special properties such that Φ is represented by m_Φ, in the sense that

$$\Phi(f) = \int f dm_\Phi \quad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

There are two versions of this correspondence theorem known as the Radon–Riesz theorem, depending on the conditions imposed on the Borel measures.

These results are similar in flavor to the fact known from linear algebra that, in a finite-dimensional vector space E with an inner product $\langle -, - \rangle$, every linear form $\varphi \in E^*$ is represented by a unique vector $u \in E$, in the sense that

$$\varphi(v) = \langle v, u \rangle \quad \text{for all } v \in E.$$

If $(E, \langle -, - \rangle)$ is an infinite-dimensional vector space which is a Hilbert space (it is complete for the norm $u \mapsto \sqrt{\langle u, u \rangle}$), then by the *Riesz representation theorem*, every *continuous* linear form $\varphi \in E'$ is represented by a unique vector $u \in E$, in the sense that

$$\varphi(v) = \langle v, u \rangle \quad \text{for all } v \in E.$$

The Radon–Riesz theorems show that certain kinds of (possibly discontinuous) linear forms on $\mathcal{K}_{\mathbb{C}}(X)$ can be represented *using integration instead of an inner product.*

The main limitation of this approach is that the linear forms Φ induced by a positive measure are positive, which means that $\Phi(f) \geq 0$ if $f \geq 0$. In particular, it is impossible to represent an arbitrary continuous linear form on $\mathcal{K}_{\mathbb{C}}(X)$ using integration. The solution to overcome this limitation is to generalize the notion of measure so that a measure can take negative, or even complex values! We will show how to do this. We will also see that, in the end, complex measures can be expressed in terms of four positive measures, but these positive measures only take finite values in $\mathbb{R}_+$. Then we will obtain a third Radon–Riesz correspondence between the continuous linear forms on $\mathcal{K}_{\mathbb{C}}(X)$ and certain kinds of complex Borel measures. This correspondence plays a crucial role in defining the notion of convolution on a locally compact group.

In this chapter every topological space X is assumed to be locally compact (and Hausdorff).

7.1 Positive Radon Functionals Induced by Borel Measures

For the record a Borel measure is defined as follows.

Definition 7.1. A *Borel measure* is any (positive) measure on a measurable space $(X, \mathcal{B})$ where X is a locally compact space and $\mathcal{B}$ is the σ-algebra of Borel sets (which is the smallest σ-algebra containing the open subsets of X).

One direction of the correspondence (Borel measures $\Longrightarrow$ linear forms) is easy to describe. It is the observation that the linear forms induced by Borel measures are positive.

Definition 7.2. For any function $f\colon X \to \mathbb{C}$, we write $f \geq 0$ if $f(X) \subseteq [0, \infty)$. If $f, g\colon X \to \mathbb{R}$, we write $f \leq g$ iff $g - f \geq 0$. A linear form $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is *positive* if for every $f \in \mathcal{K}_{\mathbb{C}}(X)$, if $f \geq 0$, then $\Phi(f) \in \mathbb{R}$ and $\Phi(f) \geq 0$.

A positive linear form has the following properties.

Proposition 7.1. *If $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is a positive linear form, then the following properties hold:*

(1) For any real-valued function $f \in \mathcal{K}_{\mathbb{R}}(X)$ we must have $\Phi(f) \in \mathbb{R}$.

(2) For any two real-valued functions $f, g \in \mathcal{K}_{\mathbb{R}}(X)$, if $f \leq g$, then $\Phi(f) \leq \Phi(g)$.

Proof. Indeed, a real-valued function f can be written uniquely as $f = f^+ - f^-$, with $f^+, f^- \in \mathcal{K}_{\mathbb{R}}(X)$, $f^+ \geq 0$ and $f^- \geq 0$. Since Φ is linear,

$$\Phi(f) = \Phi(f^+) - \Phi(f^-) \in \mathbb{R},$$

since $\Phi(f^+) \geq 0$ and $\Phi(f^-) \geq 0$ as Φ is positive.

We have $f \leq g$ iff $g - f \geq 0$, and since Φ is positive, $\Phi(g - f) \geq 0$, but since Φ is linear and positive, $\Phi(g) - \Phi(f) \geq 0$ with $\Phi(f), \Phi(g) \in \mathbb{R}$, that is, $\Phi(f) \leq \Phi(g)$. $\square$

The following proposition yields the mapping from Borel measures to positive linear forms.

Proposition 7.2. *Assume that the Borel measure μ has the property that $\mu(K)$ is finite for every compact subset of X (since X is Hausdorff, a compact set is closed, and thus a Borel set). Every function $f \in \mathcal{K}_{\mathbb{C}}(X)$ is integrable. Furthermore, the map $\varphi_\mu\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ given by*

$$\varphi_\mu(f) = \int f d\mu$$

is a positive linear form.

Proof. Since f has compact support, say K, and since it is continuous, it is bounded, say $|f| \leq M\chi_K$. Since f is continuous, it is measurable, and the function $M\chi_K$ is a step function which is integrable since $\mu(K)$ is finite. By Theorem 5.35, the function f is integrable. By Proposition 5.24, the map φ_μ is linear and positive. $\square$

Remark: As a point of terminology, the map $\varphi_\mu\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ just is just a linear form, but since its domain is a function space ($\mathcal{K}_{\mathbb{C}}(X)$), it is customary to call it a *linear functional*.

The remarkable fact is that any positive linear functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ determines a Borel measure m_Φ (with some special properties) such that

$$\Phi(f) = \int f dm_\Phi \quad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

Knowing how to integrate functions in $\mathcal{K}_{\mathbb{C}}(X)$ is sufficient to determine the measure m_Φ completely. In some sense, continuous functions with compact support play the role of μ-step functions.

Recall that for any compact subset K of X, we denote by $\mathcal{K}(K;\mathbb{C})$ the set of complex-valued continuous functions whose support is contained in K (and similarly $\mathcal{K}(K;\mathbb{R})$ for real-valued functions). Interestingly, every positive linear functional on $\mathcal{K}_{\mathbb{C}}(X)$ is continuous on $\mathcal{K}(K;\mathbb{C})$ for every compact subset K of X.

Proposition 7.3. *If $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is a positive linear functional on $\mathcal{K}_{\mathbb{C}}(X)$, then for every compact subset K of X, there is some real number $c_K \geq 0$ such that $|\Phi(f)| \leq c_K \|f\|_\infty$ for all $f \in \mathcal{K}(K;\mathbb{C})$.*

Proof. Every function f in $\mathcal{K}(K;\mathbb{C})$ can be written uniquely as $f = f_1 + if_2$ with $f_1, f_2 \in \mathcal{K}_{\mathbb{R}}(X)$. Since Φ is a positive linear functional, we have $\Phi(f_1) \in \mathbb{R}$, $\Phi(f_2) \in \mathbb{R}$ and $\Phi(f) = \Phi(f_1 + if_2) = \Phi(f_1) + i\Phi(f_2)$, so

$$|\Phi(f)| = \sqrt{\Phi(f_1)^2 + \Phi(f_2)^2}.$$

Since

$$\|f\|_\infty = \sup_{x\in K} |f(x)| = \sup_{x\in K} \sqrt{(f_1(x))^2 + (f_2(x))^2} = \sqrt{\sup_{x\in K}(f_1(x))^2 + \sup_{x\in K}(f_2(x))^2},$$

we obtain the inequalities

$$\|f_1\|_\infty = \sup_{x\in K} |f_1(x)| \leq \|f\|_\infty\,,$$

and

$$\|f_2\|_\infty = \sup_{x\in K} |f_2(x)| \leq \|f\|_\infty\,.$$

Using the above inequalities, if we can show that $|\Phi(f_1)| \leq c_1 \|f_1\|_\infty$ and $|\Phi(f_2)| \leq c_2 \|f_2\|_\infty$, then we get

$$|\Phi(f)| \leq \sqrt{c_1^2 \|f_1\|_\infty^2 + c_2^2 \|f_2\|_\infty^2} \leq \sqrt{c_1^2 + c_2^2}\, \|f\|_\infty\,.$$

Therefore we may assume that $f \in \mathcal{K}(K;\mathbb{R})$. By Proposition A.39, there is a continuous function with compact support $g \in \mathcal{K}_{\mathbb{C}}(X)$ (a bump function) such that $g(x) = 1$ for all $x \in K$. For any $f \in \mathcal{K}_{\mathbb{R}}(X)$, we have

$$-g\,\|f\|_\infty \leq f \leq g\,\|f\|_\infty$$

and since Φ is a positive linear functional, by Proposition 7.1(2), we get

$$-\Phi(g)\,\|f\|_\infty \leq \Phi(f) \leq \Phi(g)\,\|f\|_\infty$$

that is

$$|\Phi(f)| \leq \Phi(g)\,\|f\|_\infty\,,$$

as desired ∎

Proposition 7.3 suggests that the linear functionals $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ satisfying the conclusion of the proposition are of particular interest, and they are. In fact the measure theory and the integration theory for complex-valued functions on a locally compact space can be developed entirely in terms of these functionals. This approach is presented in Dieudonné [24], Bourbaki [6, 8, 11], and Schwartz [86]. Dieudonné and Bourbaki even go as far as calling such functionals *measures*, which we feel is unfortunate because this term already has a well established meaning. Unlike these two previous sources, Schwartz actually develops in parallel both the theory of integration using measure theory, and the theory of integration using certain linear functionals that he calls *Radon measures*. Again, we find this terminology unfortunate because these are functionals and not measures in the traditional sense. We propose to use the term *Radon functional*.

Definition 7.3. A linear functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is a *Radon functional* if for every compact subset K of X, there is some real number $c_K \geq 0$ such that $|\Phi(f)| \leq c_K \|f\|_\infty$ for all $f \in \mathcal{K}(K; \mathbb{C})$. The set of Radon functionals is denoted $\mathrm{M}_{\mathbb{C}}(X)$, or simply, $\mathrm{M}(X)$. The set of *positive Radon functionals* is denoted $\mathrm{M}^+(X)$, and the set of *continuous (or bounded) Radon functionals* is denoted $\mathrm{M}^1(X)$. See Figure 7.1.

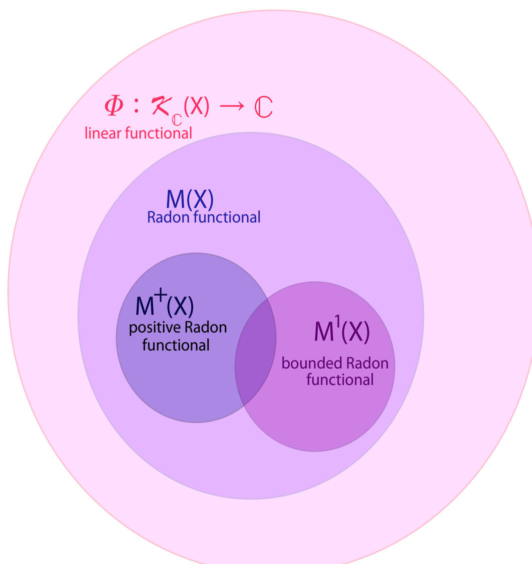


Figure 7.1: A Venn diagram classification of Radon functionals.

Equivalently, a linear functional is a Radon functional if it is continuous when restricted to $\mathcal{K}(K; \mathbb{C})$, for every compact subset K of X.

In general, a Radon functional is *not* continuous on $\mathcal{K}_{\mathbb{C}}(X)$ for the sup norm $\| \|_\infty$. For a continuous Radon functional, there is a *uniform constant* $c \geq 0$ such that

$$|\Phi(f)| \leq c \|f\|_\infty \quad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

Continuous Radon functionals are often called *bounded Radon functionals*.

Proposition 7.3 immediately implies the following result.

Proposition 7.4. *Any positive linear functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is a positive Radon functional.*

Observe that a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is completely determined by its restriction $\Phi_{\mathbb{R}}\colon \mathcal{K}_{\mathbb{R}}(X) \to \mathbb{C}$ to the space of real-valued functions in $\mathcal{K}_{\mathbb{R}}(X)$. Indeed, every function $f \in \mathcal{K}_{\mathbb{C}}(X)$ can be written uniquely as $f = f_1 + i f_2$ with $f_1, f_2 \in \mathcal{K}_{\mathbb{R}}(X)$, and by $\mathbb{C}$-linearity,

$$\Phi(f) = \Phi(f_1 + i f_2) = \Phi_{\mathbb{R}}(f_1) + i\Phi_{\mathbb{R}}(f_2).$$

Furthermore, if Φ is a positive Radon functional, then by Proposition 7.1 we have $\Phi(f) \in \mathbb{R}$ for all $f \in \mathcal{K}_{\mathbb{R}}(X)$, so $\Phi_{\mathbb{R}}\colon \mathcal{K}_{\mathbb{R}}(X) \to \mathbb{R}$. Therefore, there is a bijection between the space $\mathrm{M}^+(X)$ of positive linear functionals $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ and the space $\mathrm{M}^+_{\mathbb{R}}(X)$ of positive linear functionals $\Psi\colon \mathcal{K}_{\mathbb{R}}(X) \to \mathbb{R}$ as illustrated by Figure 7.2

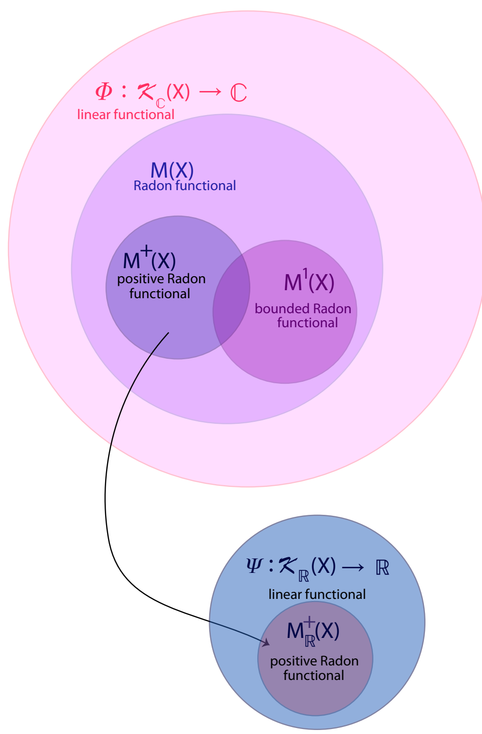


Figure 7.2: The correspondence between $\mathrm{M}^+(X)$ and $\mathrm{M}^+_{\mathbb{R}}(X)$.

Also observe that $\mathrm{M}(X)$ and $\mathrm{M}^1(X)$ are vector spaces. The operator norm $\|\ \|$ is well defined on the vector space $\mathrm{M}^1(X)$. For any bounded linear functional Φ, by definition

$$\|\Phi\| = \sup\{|\Phi(f)| \mid f \in \mathcal{K}_{\mathbb{C}}(X),\ \|f\|_\infty = 1\}.$$

Using Proposition 2.17 it is easy to show that $\mathrm{M}^1(X)$ is isomorphic to the dual $\mathcal{C}_0(X;\mathbb{C})'$ of the space $\mathcal{C}_0(X;\mathbb{C})$, that is, the space of all continuous linear forms on $\mathcal{C}_0(X;\mathbb{C})$. Recall that $\mathcal{C}_0(X;\mathbb{C})$ is the space of continuous functions which tend to 0 at infinity; see Definition 2.16.

Proposition 7.5. *Let X be a locally compact space. The space $\mathrm{M}^1(X)$ of bounded Radon functionals is isomorphic to the dual $\mathcal{C}_0(X;\mathbb{C})'$ of $\mathcal{C}_0(X;\mathbb{C})$, that is, the space of all continuous linear forms on $\mathcal{C}_0(X;\mathbb{C})$. Consequently $\mathrm{M}^1(X)$ is a Banach space (w.r.t. the sup norm).*

Proof. By Proposition 2.17, the space $\mathcal{C}_0(X;\mathbb{C})$ is the closure of $\mathcal{K}_{\mathbb{C}}(X)$. By definition, $\mathrm{M}^1(X)$ is the space of continuous linear forms on $\mathcal{K}_{\mathbb{C}}(X)$. By Theorem A.73, every continuous linear form has a unique continuous extension to $\mathcal{C}_0(X;\mathbb{C})$. Therefore $\mathrm{M}^1(X)$ is isomorphic to the dual of $\mathcal{C}_0(X;\mathbb{C})$. Since $\mathbb{C}$ is complete, it is known that the set of continuous linear maps from any vector space into $\mathbb{C}$ is complete. ☐

Here are some example of Radon functionals.

Example 7.1.

1. Pick any $a \in X$. The map δ_a given by

$$\delta_a(f) = f(a)$$

   for all $f \in \mathcal{K}_{\mathbb{C}}(X)$ is a Radon functional called (with an abuse of terminlogy) the *Dirac measure*. Since $|f(a)| \leq \|f\|_\infty$, it is a bounded Radon functional.

2. Consider the space $\mathcal{K}_{\mathbb{C}}(\mathbb{R})$ of continuous functions $f\colon \mathbb{R} \to \mathbb{C}$ with compact support. For each function $f \in \mathcal{K}_{\mathbb{C}}(\mathbb{R})$, there is a compact interval $[a, b]$ such that f vanishes outside of $[a, b]$, and from Section 3.1, the Riemann integral 5

$$I(f) = \int_a^b f(t)dt$$

   is defined. We obtain a map $I\colon \mathcal{K}_{\mathbb{C}}(\mathbb{R}) \to \mathbb{C}$ which is obviously linear. Since

$$\left|\int_a^b f(t)dt\right| \leq (b-a)\,\|f\|_\infty\,,$$

   this map is a Radon functional. Actually, this functional is positive. We will see later that this Radon functional corresponds to the Lebesgue measure.

3. Let Φ be any Radon functional and pick any continuous function $g \in \mathcal{C}(X;\mathbb{C})$. It is clear that if $f \in \mathcal{K}_{\mathbb{C}}(X)$, then $gf \in \mathcal{K}_{\mathbb{C}}(X)$, and we have a map Ψ given by

$$\Psi(f) = \Phi(gf) \quad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

Clearly, this is a linear functional. For any compact subset K of X, if $f \in \mathcal{K}_{\mathbb{C}}(X)$, then we have

$$\|gf\|_\infty \leq \|f\|_\infty \sup_{x \in K} |g(x)|.$$

Since Φ is a Radon functional, there is some real $c_K \geq 0$ such that

$$|\Phi(gf)| \leq c_K \|gf\|_\infty ,$$

so we obtain

$$|\Phi(gf)| \leq c_K \sup_{x \in K} |g(x)| \, \|f\|_\infty ,$$

which shows that Ψ is a Radon functional. The Radon functional Ψ is called the *Radon functional with density g relative to* Φ, and it is denoted $g \cdot \Phi$. Such Radon functionals play an important role in the definition of the notion of convolution in the theory of integration based on Radon functionals developed in Dieudonné [24] and Bourbaki [6, 8, 11, 7].

In the next section we state the most important theorem of the theory of Radon functionals, which is that every positive Radon functional arises from a unique Borel measure with some regularity properties.

7.2 The Radon–Riesz Theorem and Positive Radon Functionals

In this section we deal with the direction of the correspondence positive Radon functionals $\Longrightarrow$ Borel measures. Our first goal is to show that for every positive Radon functional Φ, there is a σ-algebra $\mathfrak{M}$ and a unique positive measure m_Φ on $\mathfrak{M}$ (with certain properties) representing Φ as an integral, which means that

$$\Phi(f) = \int f dm_\Phi \quad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

For instance, the positive Radon functional of Example 7.1(2) yields the Lebesgue measure. In a second stage, by imposing some reasonable conditions on the measure, we obtain a bijective correspondence.

Complete proofs of these results are quite long and intricate. Such proofs can be found in Rudin [79] (Chapter 2), Lang [62] (Chapter IX), Folland [34] (Chapter 7, Theorem 7.2), and Schwartz [86] (Chapters 5 and 7). Going back and forth between Rudin, Folland, and Lang is a possible strategy to understanding the proof.

Theorem 7.6 is often referred to as the *Riesz representation theorem*. A version of this theorem for $X = [0, 1]$ was first proven by Frigyes Riesz[1] in 1909. In 1913, Radon extended

[1]Not to be confused with his younger brother Marcel Riesz.

Riesz' result to a compact subset of $\mathbb{R}^n$ in terms of regular measures rather than a Stieltjes integral. Following Malliavin [68], it seems appropriate to call it the Radon–Riesz theorem, but it should be noted that other versions of this theorem were obtained by Banach, Saks, Markov, and Kakutani, which gives the most general version stated in Theorem 7.30; see Dunford and Schwartz [30].

Theorem 7.6. *(Radon–Riesz) Let X be a locally compact (Hausdorff) space. For every positive linear functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$, there is a σ-algebra $\mathfrak{M}$ containing the Borel σ-algebra, and there is a unique positive measure m_Φ on $\mathfrak{M}$ with the following properties:*

(1) *The linear functional Φ is represented by m_Φ, that is,*

$$\Phi(f) = \int f dm_\Phi \quad \textit{for all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

(2) *The measure $m_\Phi(K)$ is finite for every compact subset K of X.*

(3) *We have*

$$m_\Phi(E) = \inf\{m_\Phi(V) \mid E \subseteq V,\ V \text{ open}\}$$

for every $E \in \mathfrak{M}$.

(4) *We have*

$$m_\Phi(E) = \sup\{m_\Phi(K) \mid K \subseteq E,\ K \text{ compact}\}$$

for every open subset E, and for every $E \in \mathfrak{M}$ with $m_\Phi(E) < +\infty$

(5) *For any $E \in \mathfrak{M}$ and any $A \subseteq E$, if $m_\Phi(E) = 0$, then $m_\Phi(A) = 0$, in other words, m_Φ is a complete measure.*

Let us make a few comments about the proof. The uniqueness of m_Φ is not so bad. Observe that by (3) and (4), the measure m_Φ is determined by its values on compact subsets. Hence it suffices to prove that if two measures μ_1 and μ_2 satisfy the theorem, then they agree on all compact subsets.

Pick any compact K and any $\epsilon > 0$. By (3) and (4), there is some open subset V such that $K \subseteq V$ and $\mu_2(V) < \mu_2(K) + \epsilon$. By Proposition A.39, there is a continuous function $f\colon X \to [0, 1]$ such that $f(x) = 1$ for all $x \in K$, and such that $\text{supp}(f)$ is compact and $\text{supp}(f) \subseteq V$; this implies that

$$\begin{aligned}
\mu_1(K) &= \int \chi_K \, d\mu_1 \\
&\leq \int f d\mu_1 = \Phi(f) = \int f d\mu_2 \\
&\leq \int \chi_V \, d\mu_2 = \mu_2(V) \\
&< \mu_2(K) + \epsilon.
\end{aligned}$$

Therefore, $\mu_1(K) \leq \mu_2(K)$. By swapping μ_1 and μ_2, we obtain $\mu_2(K) \leq \mu_1(K)$, and thus $\mu_1(K) = \mu_2(K)$. Observe that the above derivation also shows that $\mu_1(K)$ is finite for every compact subset K.

To construct m_Φ we proceed as follows; for simplicity of notation, write μ instead of m_Φ.

(a) For every open set V in X, for every continuous function $g\colon X \to \mathbb{R}$, write $g \prec V$ if $g\colon X \to [0,1]$, $\mathrm{supp}(g)$ is compact, and $\mathrm{supp}(g) \subseteq V$. Let

$$\mu(V) = \sup\{\Phi(g) \mid g \prec V\}.$$

This will force Condition (4).

(b) Next, to force Condition (3), we extend μ to arbitrary subsets. For every $E \subseteq X$, let

$$\mu(E) = \inf\{\mu(V) \mid E \subseteq V,\ V \text{ open}\}.$$

It can be checked that μ is an outer measure.

(c) In order to obtain a σ-algebra and a measure, we need to cut down the family of subsets, still forcing Conditions (3) and (4). Let $\mathcal{A}$ be the family of all subsets A of X such that $\mu(A) < +\infty$ and

$$\mu(A) = \sup\{\mu(K) \mid K \subseteq A,\ K \text{ compact}\}.$$

Then $\mathcal{A}$ is an algebra containing all compact sets and all open sets of finite measure. The map μ is a measure on $\mathcal{A}$, and if $\mu(A) < +\infty$, then $A \in \mathcal{A}$.

(d) Let $\mathfrak{M}$ be the family of all subsets Y of X such that $Y \cap K$ lies in $\mathcal{A}$ for all compact subsets K. Then $\mathfrak{M}$ is the desired σ-algebra containing the Borel sets, and μ is a positive measure on $\mathfrak{M}$. The algebra $\mathcal{A}$ consists of the sets of finite measure in $\mathfrak{M}$.

Having done all this, one still needs to check that Conditions (1), (3), and (4) hold. Proposition A.40 (existence of finite partitions of unity) is used for some of these checks.

Theorem 7.6 shows that the measure that arises from a positive linear functional has special regularity properties that we already encountered when we met the Lebesgue measure in Section 4.5.

7.3 σ-Regular Borel Measures

Definition 7.4. A Borel measure μ on the Borel σ-algebra $\mathcal{B}$ of a locally compact space X is *σ-regular* if the following two conditions hold:

For every $E \in \mathcal{B}$,

$$\mu(E) = \inf\{\mu(V) \mid E \subseteq V,\ V \text{ open}\}. \tag{$*$}$$

For every open subset E, and for every $E \in \mathcal{B}$ with $\mu(E) < +\infty$,

$$\mu(E) = \sup\{\mu(K) \mid K \subseteq E,\ K \text{ compact}\}. \tag{$**_\sigma$}$$

Condition $(*)$ is called *outer regularity*, and Condition $(**_\sigma)$ is called *σ-inner regularity*.

We say that μ is *locally finite* if $\mu(K)$ is finite for every compact subset K.

The following proposition justifies the terminology σ-inner regularity.

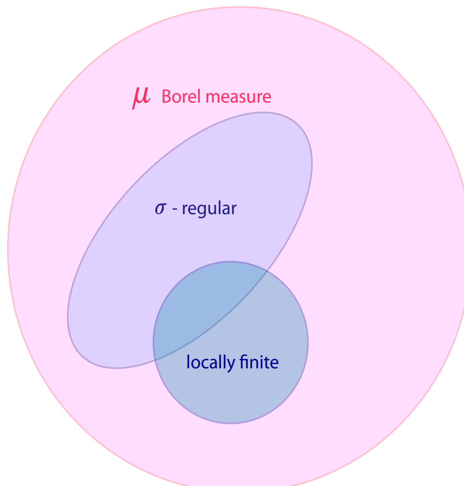


Figure 7.3: A Venn diagram classification of Borel measures.

Proposition 7.7. *Let X be a locally compact (Hausdorff) space. If a Borel measure μ is σ-inner regular, then*

$$\mu(E) = \sup\{\mu(K) \mid K \subseteq E,\ K \text{ compact}\} \tag{$**_\sigma$}$$

holds for every σ-finite subset $E \in \mathcal{B}$.

Proof. Say $E = \bigcup_{i=1}^\infty E_i$ with $E_i \in \mathcal{B}$ and $\mu(E_i) < +\infty$. We may assume that $\mu(E) = +\infty$, since if $\mu(E) < +\infty$ then we already have σ-inner regularity by definition. For every $M > 0$, there is some $n \geq 1$ such that $\mu\left(\bigcup_{i=1}^n E_i\right) > M$. Since $\bigcup_{i=1}^n E_i$ has finite measure, σ-inner regularity applies, so there is some compact subset K such that $\mu(K) > M$. This shows that

$$\sup\{\mu(K) \mid K \subseteq E,\ K \text{ compact}\} = +\infty = \mu(E),$$

which shows σ-inner regularity for E. $\square$

Definition 7.5. Let X be a locally compact (Hausdorff) space. A Borel measure μ is called a *(positive) σ-Radon measure* if it is σ-regular and locally finite. The space of σ-Radon measures is denoted by $\mathcal{M}_\sigma^+(X)$. See Figure 7.3.

Theorem 7.6 immediately implies the following correspondence which we illustrate in Figure 7.4.

Theorem 7.8. *(Radon–Riesz Correspondence, I) Let X be a locally compact (Hausdorff) space. The maps $m\colon \mathrm{M}^+(X) \to \mathcal{M}_\sigma^+(X)$ and $\varphi\colon \mathcal{M}_\sigma^+(X) \to \mathrm{M}^+(X)$ given by*

$$\begin{aligned} m(\Phi) &= m_\Phi \quad \textit{for all } \Phi \in \mathrm{M}^+(X) \\ \varphi(\mu) &= \varphi_\mu \quad \textit{for all } \mu \in \mathcal{M}_\sigma^+(X) \end{aligned}$$

are mutual inverses that define a bijection between the space $\mathrm{M}^+(X)$ of positive Radon functionals and the space $\mathcal{M}_\sigma^+(X)$ of (positive) σ-Radon measures (recall from Proposition 7.2 that

$$\varphi_\mu(f) = \int f d\mu$$

for any $f \in \mathcal{K}_{\mathbb{C}}(X)$.)

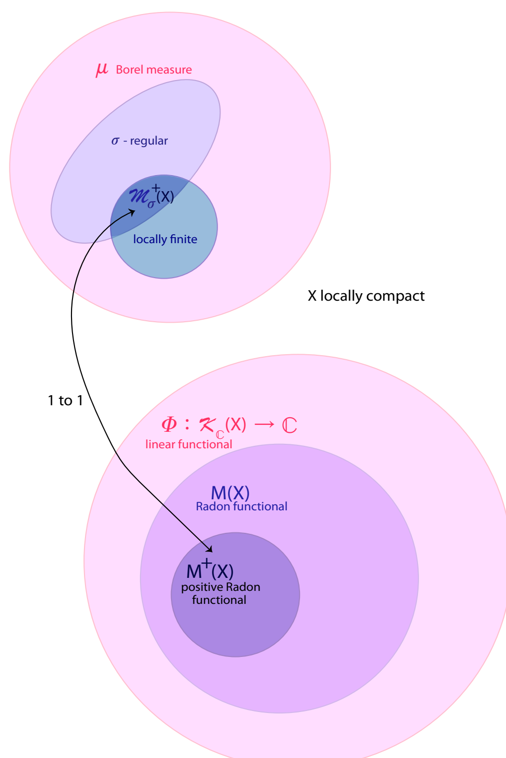


Figure 7.4: Radon–Riesz Correspondence, Version I.

Measurable functions on a locally compact space with a σ-regular, locally finite, Borel measure are very close to being continuous as stated in the following theorem of Lusin.

Theorem 7.9. *(Lusin's Theorem) Let X be a locally compact space equipped with a σ-regular, locally finite, Borel measure μ, and let f be any measurable function on X. If f vanishes outside of a set A of finite measure, for any $\epsilon > 0$, there is some function $g \in \mathcal{K}_{\mathbb{C}}(X)$ and a measurable set Z with $\mu(Z) < \epsilon$, such that $f(x) = g(x)$ for all $x \in X - Z$, and $\|g\|_\infty \leq \|f\|_\infty$.*

Theorem 7.9 is proven in Rudin [79] (Chapter 2, Theorem 2.24) and Lang [62] (Chapter IX, Theorem 3.3).

The Vitali–Carathéodory theorem states that every function in $\mathrm{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$ can be approximated from below and from above by certain kinds of functions called upper semicontinuous and lower semicontinuous, see Rudin [79] (Chapter 2, Theorem 2.25).

We have the following density result which uses Lusin's theorem (Theorem 7.9).

Theorem 7.10. *Let X be a locally compact space equipped with a σ-regular, locally finite, Borel measure μ. The space $\mathcal{K}_{\mathbb{C}}(X)$ is dense in $\mathrm{L}^p_\mu(X, \mathcal{B}, \mathbb{C})$ for $p = 1, 2$.*[2]

Theorem 7.10 is proven in Rudin [79] (Chapter 3, Theorem 3.14) and Lang [62] (Chapter IX, Theorem 3.1).

The following corollary of Theorem 7.10 will be used in Chapter 12.

Theorem 7.11. *Let X be a locally compact, metrizable, separable space equipped with a σ-regular, locally finite, Borel measure μ. Then $\mathrm{L}^p_\mu(X, \mathcal{B}, \mathbb{C})$ is separable for $p = 1, 2$.*

Theorem 7.11 follows immediately from Theorem 7.10 and Theorem 2.16.

The following proposition is needed for proving the uniqueness of the Haar measure up to a constant.

Proposition 7.12. *Let X be a locally compact space equipped with a σ-regular, locally finite, Borel measure μ. For any function $f \in \mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$, if*

$$\int f g \, d\mu = 0 \quad \text{for all } g \in \mathcal{K}_{\mathbb{C}}(X),$$

then $f = 0$ almost everywhere.

Proof. We use Proposition 5.39, recalling the fact that $\int_A f d\mu = \int f \chi_A \, d\mu$. Let A be any subset of finite measure. By Theorem 7.10, χ_A is the L^1-limit of a sequence (g_n) of functions $g_n \in \mathcal{K}_{\mathbb{C}}(X)$ with $g_n(X) \subseteq [0, 1]$. By Proposition 5.26, there is a subsequence $(g_{n_k})_{k \geq 1}$ that converges pointwise to χ_A a.e., and thus $(f g_{n_k})$ converges pointwise to $f \chi_A$ a.e. By Proposition 7.2, the functions g_{n_k} are integrable, so the functions $(f g_{n_k})$ are also integrable, and since $g_{n_k}(X) \subseteq [0, 1]$, by the dominated convergence theorem, we conclude that $\int_A f d\mu = \int f \chi_A \, d\mu = 0$ for all subsets A of finite measure, and by Proposition 5.39, we have $f = 0$ a.e. ∎

[2]Even for all p with $1 \leq p < +\infty$.

In the next section we show that by requiring the locally compact space X to be also σ-compact, then we obtain Borel measures that are not only σ-regular, but regular as well, which means that inner regularity holds for all $E \in \mathfrak{B}$.

7.4 Regular Borel Measures

In Theorem 7.6 outer regularity holds, but σ-inner regularity holds only for open subsets and measurable sets of finite measure. It is often desirable for inner regularity to hold for *arbitrary* subsets $E \in \mathcal{B}$, possibly not σ-finite. It turns out that making some mild restrictions on X, we obtain a bijection between positive linear functionals and these regular measures. On this subject, Rudin's exposition seems clearer than Lang's exposition.

Definition 7.6. A Borel measure μ on the Borel σ-algebra $\mathcal{B}$ of a locally compact space X is *regular* if the following two conditions hold for every $E \in \mathcal{B}$:

$$\mu(E) = \inf\{\mu(V) \mid E \subseteq V,\ V \text{ open}\} \tag{$*$}$$

and

$$\mu(E) = \sup\{\mu(K) \mid K \subseteq E,\ K \text{ compact}\}. \tag{$**$}$$

Condition $(*)$ is called *outer regularity*, and Condition $(**)$ is called *inner regularity*. See Figure 7.5.

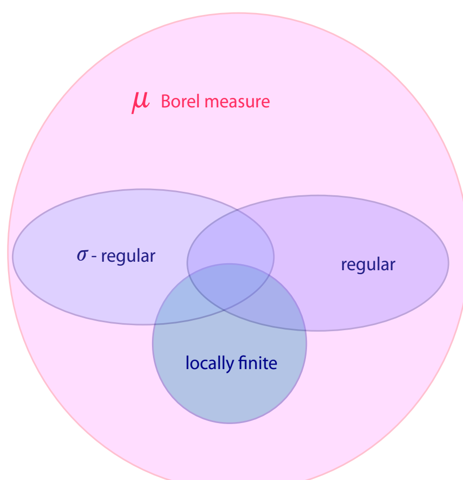


Figure 7.5: Another Venn diagram classification of Borel measures.

Observe that if a Borel measure μ is σ-finite (on X) and if it is σ-regular, then it is actually regular. Another sufficient condition is given in the next proposition.

Proposition 7.13. *Let X be a locally compact (Hausdorff) space in which every open subset is σ-compact. If μ is a locally finite Borel measure, then μ is a regular measure.*

Proposition 7.13 is proven in Rudin [79] (Chapter 2, Theorem 2.18).

Observe that $X = \mathbb{R}^n$ satisfies the condition of Proposition 7.13. Thus a locally finite Borel measure on $\mathbb{R}^n$ is a regular measure.

A way to obtain the Radon–Riesz correspondence between positive Radon functionals and regular locally finite Borel measures is to require X to be σ-compact, which means that X is the countable union of compact subsets (see Definition A.43).

Theorem 7.14. *Let X be a locally compact (Hausdorff), σ-compact space. For every positive linear functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$, if $\mathfrak{M}$ and m_Φ are the σ-algebra and the measure obtained in Theorem 7.6, then the following properties holds:*

(1) For any $E \in \mathfrak{M}$ and any $\epsilon > 0$, there is a closed set F and an open set O such that $F \subseteq E \subseteq O$ and $\mu(O - F) < \epsilon$.

(2) The measure m_Φ is a regular, locally finite Borel measure on the Borel σ-algebra $\mathcal{B}$.

Theorem 7.14 is proven in Rudin [79] (Chapter 2, Theorem 2.17). The following theorem allows us to get a bijective correspondence between positive linear functional and regular locally finite Borel measures, and to state this theorem it is convenient to introduce the following definition.

Definition 7.7. Let X be a locally compact (Hausdorff) space. A Borel measure μ is called a *(positive) Radon measure* if it is regular and locally finite. The space of Radon measures is denoted by $\mathcal{M}^+_{\mathrm{rad}}(X)$, or simply $\mathcal{M}^+(X)$. See Figure 7.5.

Theorem 7.15. *(Radon–Riesz Correspondence, II) Let X be a locally compact (Hausdorff), σ-compact space. The maps $m\colon \mathrm{M}^+(X) \to \mathcal{M}^+(X)$ and $\varphi\colon \mathcal{M}^+(X) \to \mathrm{M}^+(X)$ given by*

$$\begin{aligned} m(\Phi) &= m_\Phi \quad \textit{for all } \Phi \in \mathrm{M}^+(X) \\ \varphi(\mu) &= \varphi_\mu \quad \textit{for all } \mu \in \mathcal{M}^+(X) \end{aligned}$$

are mutual inverses that define a bijection between the space $\mathrm{M}^+(X)$ of positive Radon functionals and the space $\mathcal{M}^+(X)$ of (positive) Radon measures. See Figure 7.6.

An interesting application of Theorem 7.15 is obtained by choosing $X = \mathbb{R}$ and Φ to be the Radon functional I induced by the Riemann integral defined in Example 7.1(2). The Radon measure m_I given by Theorem 7.15 turns out to be the Lebesgue measure μ_L. For details, see Rudin [79] (Chapter 2).

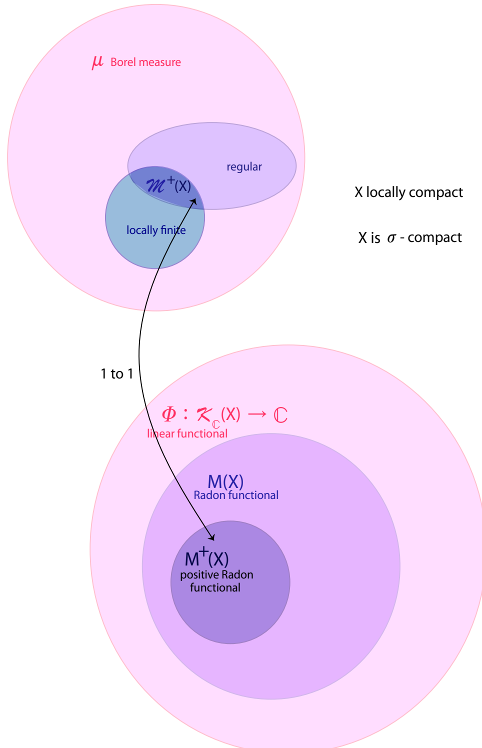


Figure 7.6: Radon–Riesz Correspondence, Version 2.

7.5 Complex and Real Measures

By Proposition 7.2, the functionals induced by Borel measures are positive, but there are Radon functionals that are not positive, so it is natural to ask if such functionals arise from some generalized measures allowed to take negative values, or even complex values. The answer is yes. It is even possible to define measures with values in any Banach space. Such measures are discussed in Lang [62], Schwartz [86] and Marle [69], but for simplicity we will only consider real and complex measures. In this section we take a small detour to define complex measures. Then we will show how they relate to functionals on $\mathcal{K}_{\mathbb{C}}(X)$ that are not necessarily positive, but continuous.

Going back to Definition 4.9, a (positive) measure on a measurable set $(X, \mathcal{A})$ is a map μ satisfying the following properties:

(μ1) $\mu\colon \mathcal{A} \to [0, +\infty]$, where $\mathcal{A}$ is a σ-algebra of subsets of X.

(μ2) $\mu(\emptyset) = 0$.

(μ3) For any countable sequence $(A_i)_{i\geq 1}$ of subsets A_i of $\mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for all

$i \neq j$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Such a function may have the value $+\infty$, but in (μ3), if $A = \bigcup_{i=1}^{\infty} A_i$ and if $\mu(A)$ is finite, then the series $\sum_{i=1}^{\infty} \mu(A_i)$ converges, and since it consists of nonnegative numbers, it converges absolutely, and thus *commutatively*, which means that for any permutation σ of $\mathbb{N}_+$, we have

$$\mu(A) = \mu\left(\bigcup_{i=1}^{\infty} A_\sigma(i)\right) = \sum_{i=1}^{\infty} \mu(A_{\sigma(i)}).$$

If we replace $[0, +\infty]$ by $\mathbb{R}$ or $\mathbb{C}$, then a new problem arises, namely that the convergence of the sum $\sum_{i=1}^{\infty} \mu(A_i)$ generally depends on the order of the A_i. The solution is to *require* commutative convergence of the series arising in (μ3). It is known from analysis that for $\mathbb{R}$ or $\mathbb{C}$, a series is commutatively convergent iff it is absolutely convergent, so we require the latter. We also require $\mu(A)$ be an element of $\mathbb{R}$ or $\mathbb{C}$, that is, $\mu(A)$ must be "finite." There is a way to define measures with values in $\mathbb{R} \cup \{+\infty\}$, and even in $\mathbb{R} \cup \{-\infty, +\infty\}$, but we have no need for such generality (see Schwartz [86], Chapter V, §9).

Definition 7.8. Let $(X, \mathcal{A})$ be a measurable space. A *complex measure* on $(X, \mathcal{A})$ is a map μ satisfying the following properties:

(μ1) $\mu\colon \mathcal{A} \to \mathbb{C}$.

(μ2) $\mu(\emptyset) = 0$.

(μ3) For any countable family $(A_i)_{i \geq 1}$ of subsets A_i of $\mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i),$$

where the series on the right-hand side is absolutely convergent.

A *real measure* (or *signed measure*) is a complex measure such that $\mu(\mathcal{A}) \subseteq \mathbb{R}$.

Observe that a real measure which is also positive is a positive measure according to Definition 4.9, but since a positive measure may take the value $+\infty$, there are positive measures that are not real measures in the sense of Definition 7.8. When we use the term *positive real measure*, we mean that this measure only takes finite values. By *positive measure*, we mean a measure that may take the value $+\infty$.

One might wonder if interesting real or complex measures exist. Indeed, for any *arbitrary measure space* $(X, \mathcal{A}, \mu)$, every function $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$ gives rise to such a measure.

Proposition 7.16. *Let $(X, \mathcal{A}, \mu)$ be a measure space (here, μ is a positive measure). For every integrable map $f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C})$, the function $\mu_f \colon \mathcal{A} \to \mathbb{C}$ given by*

$$\mu_f(A) = \int_A f d\mu = \int f\chi_A d\mu \quad \textit{for all } A \in \mathcal{A}$$

is a complex measure.

What is not obvious is that ($\mu 3$) holds. This follows from Proposition 5.37 (a consequence of the Lebesgue dominated convergence theorem). A detailed proof is given in Marle [69] (Chapter 2, Proposition 2.5.2).

The new twist here is that given a measure μ, rather than defining a functional by *varying the function being integrated*, we fix a function but we integrate by *varying the subset over which we integrate*.

It is trivial to check that the complex measures (and the real measures) form a vector space.

Remarkably, every complex measure μ arises as a measure of the form $|\mu|_h$ for some suitable positive measure $|\mu|$ and some well chosen function $h \in \mathcal{L}_{|\mu|}(X, \mathcal{A}, \mathbb{C})$; see Theorem 7.21. The measure $|\mu|$ is defined as follows.

Definition 7.9. Let $(X, \mathcal{A})$ be a measurable space, and let μ be a complex measure on $(X, \mathcal{A})$. Define the map $|\mu| \colon \mathcal{A} \to [0, +\infty]$ by

$$|\mu|(A) = \sup \sum_{i=1}^{\infty} |\mu(A_i)|,$$

for all $A \in \mathcal{A}$ and for all countable partitions $(A_i)_{i \geq 1}$ of A with $A_i \in \mathcal{A}$. The map $|\mu|$ is called the *total variation measure* (for short *total variation*) of μ.

Obviously, if μ is a real positive measure, then $|\mu| = \mu$. It is easy to see that by definition,

$$|\mu(A)| \leq |\mu|(A) \quad \text{for all } A \in \mathcal{A}.$$

In fact, it is minimal with this property. We have the following remarkable theorems.

Theorem 7.17. *Let $(X, \mathcal{A})$ be a measurable space, and let μ be a complex measure on $(X, \mathcal{A})$. The map $|\mu| \colon \mathcal{A} \to [0, +\infty]$ is a positive measure. The positive measure $|\mu|$ is the minimal measure such that*

$$|\mu(A)| \leq |\mu|(A) \quad \textit{for all } A \in \mathcal{A},$$

in the sense that if λ is any positive measure such that

$$|\mu(A)| \leq \lambda(A) \quad \textit{for all } A \in \mathcal{A},$$

then $|\mu| \leq \lambda$ (which means that $|\mu|(A) \leq \lambda(A)$ for all $A \in \mathcal{A}$).

A proof of Theorem 7.17 is given in Rudin [79] (Chapter 6, Theorem 6.2) and Lang [62] (Chapter VII, Theorem 3.1).

The next theorem is even more surprising.

Theorem 7.18. *Let $(X, \mathcal{A})$ be a measurable space, and let μ be a complex measure on $(X, \mathcal{A})$. The map $|\mu|\colon \mathcal{A} \to [0, +\infty]$ is a finite positive measure; that is, $|\mu|(X) < +\infty$.*

A proof of Theorem 7.18 is given in Rudin [79] (Chapter 6, Theorem 6.4) and Lang [62] (Chapter VII, Theorem 3.2). Theorem 7.18 implies that $\mu(X)$ is bounded: it is contained in a closed disk of finite radius. This fact shows that the convergence requirement of Condition ($\mu 3$) is quite strong.

Theorem 7.18 allows us to make the space of complex measures into a normed vector space.

Definition 7.10. Let $(X, \mathcal{A})$ be measurable space. For any complex measure μ, define $\|\mu\|$ as $\|\mu\| = |\mu|(X)$. The vector space of complex measures equipped with the norm defined above is denoted $\mathbb{C}\mathcal{M}^1(X, \mathcal{A})$.

It is not hard to show that $\mathbb{C}\mathcal{M}^1(X, \mathcal{A})$ is a Banach space.

Proposition 7.19. *Let $(X, \mathcal{A})$ be a measurable space. The normed vector space $\mathbb{C}\mathcal{M}^1(X, \mathcal{A})$ is a Banach space (it is complete).*

Another interesting fact is that if μ is a positive measure (possibly taking the value $+\infty$) then $\mathcal{L}^1_\mu(X, \mathcal{A}, \mathbb{C})$ can be embedded in $\mathbb{C}\mathcal{M}^1(X, \mathcal{A})$.

Proposition 7.20. *Let $(X, \mathcal{A}, \mu)$ be a measure space. The map $f \mapsto \mu_f$ is a linear embedding of $\mathcal{L}^1_\mu(X, \mathcal{A}, \mathbb{C})$ into $\mathbb{C}\mathcal{M}^1(X, \mathcal{A})$, and*

$$\|\mu_f\| = \|f\|_1 \quad \text{for all } f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{C}).$$

Proposition 7.20 is proven in Lang [62] (Chapter VII, §3, Theorem 3.3). The proof uses Proposition 5.18.

The next theorem shows an important fact that we mentioned earlier, namely that every complex measure μ arises as a measure of the form $|\mu|_h$ for some well chosen function $h \in \mathcal{L}_{|\mu|}(X, \mathcal{A}, \mathbb{C})$. This result is a special case of the Radon–Nikodym theorem, but for now we prefer not discussing this theorem.

Theorem 7.21. *For every complex measure μ on a measurable space $(X, \mathcal{A})$, there is a function $h \in \mathcal{L}^1_{|\mu|}(X, \mathcal{A}, \mathbb{C})$ such that $|h| = 1$ and*

$$\mu(A) = \int_A h \, d|\mu| \quad \text{for all } A \in \mathcal{A}.$$

In other words, $\mu = |\mu|_h$ (recall that $|\mu|$ is a positive measure). Furthermore, any two functions $h_1, h_2 \in \mathcal{L}^1_{|\mu|}(X, \mathcal{A}, \mathbb{C})$ satisfying the conditions of the theorem are equal $|\mu|$-a.e.

For a proof of Theorem 7.21, see Rudin [79] (Chapter 6, Theorem 6.12) and Lang [62] (Chapter VII, §2 and §4).

Let us now turn our attention to real measures. We will see that any real measure can be expressed in terms of two positive real measures. This implies that any complex measure can be expressed in terms of four positive real measures. This will allow us to explain how to integrate with respect to a complex measure.

7.6 Real Measures and the Hahn–Jordan Decomposition

We begin by showing that a real measure can be expressed as the difference of two finite positive measures. If μ is a real measure, since $|\mu|$ is a finite measure, we can define two finite positive measures μ^+ and μ^- such that $\mu = \mu^+ - \mu^-$.

Definition 7.11. If μ is a real measure, the real measures μ^+ and μ^- are defined by

$$\mu^+ = \frac{1}{2}(|\mu| + \mu), \quad \mu^- = \frac{1}{2}(|\mu| - \mu).$$

It is immediately checked that μ^+ and μ^- are *finite positive* measures, and we have

$$\mu = \mu^+ - \mu^-, \quad |\mu| = \mu^+ + \mu^-.$$

Definition 7.12. Given a real measure μ, the positive real measures μ^+ and μ^- are called the *positive variation* and *negative variation* of μ. The expression of μ as $\mu = \mu^+ - \mu^-$ is called the *Jordan decomposition* of μ.

The Jordan decomposition has certain minimality properties that we are going to describe.

Definition 7.13. Let $(X, \mathcal{A})$ be a measurable space. A complex measure μ is *concentrated* on (or *carried by*) a measurable subset A if $\mu(E) = 0$ for all $E \in \mathcal{A}$ such that $E \cap A = \emptyset$. Two complex measures μ_1 and μ_2 are *mutually singular* if there exist two disjoint measurable subsets A_1 and A_2 such that μ_1 is concentrated on A_1 and μ_2 is concentrated on A_2. We sometimes write $\mu_1 \perp \mu_2$.

Every real measure has a Hahn–Jordan decomposition as described by the following theorem.

Theorem 7.22. *(Hahn–Jordan Decomposition) Let $(X, \mathcal{A})$ be a measurable space. For any real measure μ, there is a partition (X^+, X^-) of X into two disjoint subsets of X such that if*

$$\mu = \mu^+ - \mu^-$$

is the Jordan decomposition of μ*, then* μ^+ *is concentrated on* X^+*, and* μ^- *is concentrated on* X^-*. Furthermore, for any* $E \in \mathcal{A}$*, we have*

$$\mu^+(E) = \sup\{\mu(A) \mid A \subseteq E,\ A \in \mathcal{A}\}, \quad \mu^-(E) = \sup\{-\mu(A) \mid A \subseteq E,\ A \in \mathcal{A}\}.$$

For any other partition (Y^+, Y^-) *of* X *such that* μ^+ *is concentrated on* Y^+ *and* μ^- *is concentrated on* Y^-*,*

$$\mu^+(E \cap X^+) = \mu^+(E \cap Y^+), \quad \mu^-(E \cap X^-) = \mu^-(E \cap Y^-),$$

for all $E \in \mathcal{A}$*.*

Let us now consider a complex measure $\mu\colon \mathcal{A} \to \mathbb{C}$.

Definition 7.14. Given a complex measure $\mu\colon \mathcal{A} \to \mathbb{C}$, the function $\overline{\mu}\colon \mathcal{A} \to \mathbb{C}$ called *the conjugate* of μ is defined by $\overline{\mu}(A) = \overline{\mu(A)}$ for all $A \in \mathcal{A}$. We also define $\mu_1\colon \mathcal{A} \to \mathbb{R}$ and $\mu_2\colon \mathcal{A} \to \mathbb{R}$ by

$$\mu_1(A) = \frac{1}{2}(\mu(A) + \overline{\mu(A)}), \quad \mu_2(A) = \frac{1}{2i}(\mu(A) - \overline{\mu(A)})$$

for all $A \in \mathcal{A}$. We call μ_1 the *real part* of μ and μ_2 the *imaginary part* of μ.

It is immediately checked that $\overline{\mu}$ is a complex measure, and that μ_1 and μ_2 are *real* measures such that

$$\begin{aligned} \mu &= \mu_1 + i\mu_2 \\ \overline{\mu} &= \mu_1 - i\mu_2. \end{aligned}$$

Using the Hahn–Jordan decomposition of μ_1 an μ_2, we see that we can write μ uniquely in terms of four positive real measure $\mu_1^+, \mu_1^-, \mu_2^+, \mu_2^-$, as

$$\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-).$$

Definition 7.15. For any complex measure $\mu\colon \mathcal{A} \to \mathbb{C}$, the expression

$$\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-).$$

is called the *Jordan decomposition* of μ.

Proposition 7.23. *For any complex measure* $\mu\colon \mathcal{A} \to \mathbb{C}$*, we have* $|\mu_1| \leq |\mu|$*,* $|\mu_2| \leq |\mu|$*, and that* $|\mu| \leq |\mu_1| + |\mu_2|$*. A function* f *is* $|\mu|$*-integrable iff it is integrable for all four positive real measures* $\mu_1^+, \mu_1^-, \mu_2^+$*, and* μ_2^-*.*

Proof. It is easy to check that $|\mu_1| \leq |\mu|$, $|\mu_2| \leq |\mu|$, and that $|\mu| \leq |\mu_1| + |\mu_2|$. It follows easily that f is $|\mu|$-integrable if f is $|\mu_1|$-integrable and $|\mu_2|$-integrable. Since $|\mu_1| = \mu_1^+ + \mu_1^-$ and $|\mu_2| = \mu_2^+ + \mu_2^-$, it is also easy to see that f is $|\mu_1|$-integrable iff f is μ_1^+-integrable and μ_1^--integrable, and similarly f is $|\mu_2|$-integrable iff f is μ_2^+-integrable and μ_2^--integrable. Therefore, f is $|\mu|$-integrable iff it is integrable for all four positive measures $\mu_1^+, \mu_1^-, \mu_2^+$, and μ_2^-. $\square$

The Jordan decomposition of the complex measure μ suggests defining the integral $\int f d\mu$ for any function $f \in \mathcal{L}^1_{|\mu|}(X, \mathcal{A}, \mathbb{C})$; see Dieudonné [24] (Chapter XIII, Section 16, no. 13.16.2), or Folland [34] (end of Section 3.1 and Section 3.3).

Definition 7.16. Given any complex measure $\mu \colon \mathcal{A} \to \mathbb{C}$, for any function $f \in \mathcal{L}^1_{|\mu|}(X, \mathcal{A}, \mathbb{C})$, we define the integral $\int f d\mu$ as

$$\int f d\mu = \int f d\mu_1 + i \int f d\mu_2 = \int f d\mu_1^+ - \int f d\mu_1^- + i \int f d\mu_2^+ - i \int f d\mu_2^-.$$

By Proposition 7.23, the above expression is well defined since f is $|\mu|$-integrable iff it is integrable for all four positive real measures $\mu_1^+, \mu_1^-, \mu_2^+$, and μ_2^-.

Remark: Alternatively, if μ is a complex measure, $\int f d\mu$ can be defined using Theorem 7.21 as $\int fh\, d|\mu|$, as in Rudin [79] (Chapter 6, Section 6.18).

The following fact will be needed later.

Proposition 7.24. *Given a complex measure μ, if $\overline{\mu}$ is the conjugate measure of μ, for any function $f \in \mathcal{L}^1_{|\mu|}(X, \mathcal{A}, \mathbb{C})$ we have*

$$\int f d\overline{\mu} = \overline{\int \overline{f}\, d\mu}, \quad \textit{or equivalently} \quad \int \overline{f}\, d\mu = \overline{\int f d\overline{\mu}}.$$

As a consequence, $\overline{\mu}$ is the unique complex measure such that

$$\int f d\overline{\mu} = \overline{\int \overline{f}\, d\mu}, \qquad \textit{for all } f \in \mathcal{C}_0(X; \mathbb{C}).$$

Proof. Write $\mu = \mu_1 + i\mu_2$ as above, where μ_1 and μ_2 are real measures. We have $\overline{\mu} = \mu_1 - i\mu_2$, and the measures μ_1 and μ_2 are written as $\mu_1 = \mu_1^+ - \mu_1^-$ and $\mu_2 = \mu_2^+ - \mu_2^-$, where $\mu_1^+, \mu_1^-, \mu_2^+$, and μ_2^-, are real positive measures. Now for any function f integrable for all four positive measures above it is obvious that

$$\int \overline{f} d\mu_i^+ = \overline{\int f d\mu_i^+}, \qquad \int \overline{f} d\mu_i^- = \overline{\int f d\mu_i^-},$$

so

$$\int \overline{f} d\mu_1 = \overline{\int f d\mu_1}, \qquad \int \overline{f} d\mu_2 = \overline{\int f d\mu_2},$$

thus

$$\begin{aligned}\int \overline{f} d\mu &= \int \overline{f} d\mu_1 + i \int \overline{f} d\mu_2 \\ &= \overline{\int f d\mu_1} + i \overline{\int f d\mu_2} \\ &= \overline{\int f d\mu_1 - i \int f d\mu_2} \\ &= \overline{\int f d\overline{\mu}},\end{aligned}$$

as claimed. Since $\mathcal{C}_0(X;\mathbb{C})$ is obviously contained in $\mathcal{L}^1_{|\mu|}(X,\mathcal{A},\mathbb{C})$, the last statement follows from Theorem 7.30 (Radon–Riesz III), which will be proven in Section 7.8. $\square$

Since the measures $\mu_1^+, \mu_1^-, \mu_2^+$, and μ_2^- are positive *real* measures, they are finite. This immediately implies that the Radon functional φ_μ induced by a complex measure μ *is bounded*. Therefore, complex measures represent only bounded Radon functionals. Actually they represent all of them, which is the object of Section 7.8.

To show the above fact, we need to decompose a bounded Radon functional in terms of (four) positive bounded Radon functionals, and for this we introduce the notion of total variation of a Radon functional.

7.7 Total Variation of a Radon Functional

The notion of total variation of a Radon functional allows the decomposition of a bounded Radon functional into four positive bounded functionals in a way that is similar to the Jordan decomposition of a complex measure. This fact is the key to the representation of a bounded Radon functional by a complex measure.

Recall that for any function $g\colon X \to \mathbb{C}$, we denote by $|g|$ the function $|g|\colon X \to \mathbb{R}$ given by $|g|(x) = |g(x)|$ for all $x \in X$.

The following result is shown in Dieudonné [24] (Chapter XIII, Section 3).

Theorem 7.25. *For any Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ on a locally compact space X, there is a smallest positive Radon functional $|\Phi|\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ such that*

$$|\Phi(f)| \leq |\Phi|(|f|) \quad \textit{for all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

The functional $|\Phi|$ is completely defined by its restriction to positive functions $f \geq 0$ in $\mathcal{K}_{\mathbb{R}}(X)$ by

$$|\Phi|(f) = \sup\{|\Phi(g)| \mid g \in \mathcal{K}_{\mathbb{C}}(X),\ |g| \leq f\}.$$

Proof sketch. We know from the remark just after Proposition 7.4 that a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is completely determined by its restriction $\Phi_{\mathbb{R}}\colon \mathcal{K}_{\mathbb{R}}(X) \to \mathbb{C}$ to the space of real-valued functions in $\mathcal{K}_{\mathbb{R}}(X)$. The first step in the proof of Theorem 7.25 is to show that the formula

$$|\Phi|(f) = \sup\{|\Phi(g)| \mid g \in \mathcal{K}_{\mathbb{C}}(X),\ |g| \leq f\}$$

defined on positive functions $f \geq 0$ in $\mathcal{K}_{\mathbb{R}}(X)$ is yields a finite number. Let K be the support of f, which is compact. Since $|g| \leq f$, the support of g is contained in K, so

$$|\Phi(g)| \leq c_K \left\|g\right\|_\infty \leq c_K \left\|f\right\|_\infty,$$

which shows that $|\Phi|(f)$ is finite. Next we show that $|\Phi|$ is additive, which is left as an exercise.

The second step is to extend $|\Phi|$ to arbitrary functions $f \in \mathcal{K}_{\mathbb{R}}(X)$ by writing $f = f' - f''$, where $f', f'' \in \mathcal{K}_{\mathbb{R}}(X)$ and $f', f'' \geq 0$, by setting

$$|\Phi|(f) = |\Phi|(f') - |\Phi|(f'').$$

This expression does not depend on the decomposition of f because if $f = f_1' - f_1'' = f_2' - f_2''$, then $f_1' + f_2'' = f_1'' + f_2'$, so $|\Phi|(f_1') + |\Phi|(f_2'') = |\Phi|(f_1'') + |\Phi|(f_2')$, which implies $|\Phi|(f_1') - |\Phi|(f_2') = |\Phi|(f_1'') - |\Phi|(f_2'')$.

The last step is to prove that $|\Phi|(\lambda f) = \lambda|\Phi|(f)$, which is clear $\lambda \geq 0$. For $\lambda < 0$, we write $f = f' - f''$ with $f', f'' \geq 0$, and then

$$\begin{aligned}
|\Phi|(\lambda f) &= |\Phi|(\lambda f' - \lambda f'') \\
&= |\Phi|(\lambda f') + |\Phi|(-\lambda f'') \\
&= -|\Phi|(-\lambda f') - \lambda|\Phi|(f'') \\
&= -(-\lambda)|\Phi|(f') - \lambda|\Phi|(f'') \\
&= \lambda(|\Phi|(f') - |\Phi|(f'')) \\
&= \lambda|\Phi|(f).
\end{aligned}$$

In summary, $|\Phi|$ is a positive linear functional. By Proposition 7.4, the functional $|\Phi|$ is a positive Radon functional. $\square$

Definition 7.17. Given any Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$, the positive Radon functional $|\Phi|$ is called *total variation* (or *absolute value*) of Φ.

If Φ is a positive Radon functional, then

$$|\Phi| = \Phi.$$

Definition 7.18. Given a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$, we define the *conjugate* $\overline{\Phi}$ of Φ by

$$\overline{\Phi}(f) = \overline{\Phi(\overline{f})}, \quad f \in \mathcal{K}_{\mathbb{C}}(X).$$

If we write $f = f_1 + if_2$ with $f_1, f_2 \in \mathcal{K}_{\mathbb{R}}(X)$, then we have

$$\overline{\Phi}(f) = \overline{\Phi}(f_1 + if_2) = \overline{\Phi(\overline{(f_1 + if_2)})} = \overline{\Phi(f_1 - if_2)} = \overline{(\Phi(f_1) - i\Phi(f_2))} = \overline{\Phi(f_1)} + i\overline{\Phi(f_2)}.$$

Definition 7.19. We say that a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is *real* if $\overline{\Phi} = \Phi$.

Proposition 7.26. *A Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is real iff its restriction $\Phi_{\mathbb{R}}$ to $\mathcal{K}_{\mathbb{R}}(X)$ is a real-valued function $\Phi_{\mathbb{R}}\colon \mathcal{K}_{\mathbb{R}}(X) \to \mathbb{R}$.*

Proof. In view of the above computation, a Radon functional Φ is real iff

$$\Phi(f_1) + i\Phi(f_2) = \overline{\Phi(f_1)} + i\overline{\Phi(f_2)}$$

for all $f_1, f_2 \in \mathcal{K}_{\mathbb{R}}(X)$, which by setting $f_2 = 0$ or $f_1 = 0$ means that $\Phi(f_i) \in \mathbb{R}$ for all $f_i \in \mathcal{K}_{\mathbb{R}}(X)$, for $i = 1, 2$. Equivalently, a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ is real iff its restriction $\Phi_{\mathbb{R}}$ to $\mathcal{K}_{\mathbb{R}}(X)$ is a real-valued function $\Phi_{\mathbb{R}}\colon \mathcal{K}_{\mathbb{R}}(X) \to \mathbb{R}$. $\square$

Since a Radon functional is Φ is completely determined by its restriction $\Phi_{\mathbb{R}}$ to $\mathcal{K}_{\mathbb{R}}(X)$, we often think of a real Radon functional as a linear map $\Phi\colon \mathcal{K}_{\mathbb{R}}(X) \to \mathbb{R}$.

Definition 7.20. Given a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$, we define Φ_r and Φ_i by

$$\Phi_r = \frac{1}{2}(\Phi + \overline{\Phi}), \quad \Phi_i = \frac{1}{2i}(\Phi - \overline{\Phi}).$$

It is immediately verified that Φ_r and Φ_i are *real* Radon functionals such that

$$\Phi = \Phi_r + i\Phi_i, \quad \overline{\Phi} = \Phi_r - i\Phi_i.$$

We also have

$$|\Phi_r| \leq |\Phi|, \quad |\Phi_i| \leq |\Phi|, \quad |\Phi| \leq |\Phi_r| + |\Phi_i|.$$

Definition 7.21. If $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{R}$ is a real Radon functional, then as in the case of real measures we can define Φ^+ and Φ^- by

$$\Phi^+ = \frac{1}{2}(|\Phi| + \Phi), \quad \Phi^- = \frac{1}{2}(|\Phi| - \Phi).$$

It is immediately checked that Φ^+ and Φ^- are *positive* Radon functionals, and we have

$$\Phi = \Phi^+ - \Phi^-, \quad |\Phi| = \Phi^+ + \Phi^-.$$

In the end, we have the following decomposition result analogous to the Jordan decomposition for complex measures.

Proposition 7.27. *Every Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ can be expressed in terms of four positive Radon functionals:*

$$\Phi = \Phi_r^+ - \Phi_r^- + i(\Phi_i^+ - \Phi_i^-).$$

By the Radon–Riesz I theorem (Theorem 7.8), there exist four positive σ-Radon measures m_1, m_2, m_3, m_4 such that

$$\Phi(f) = \int f dm_1 - \int f dm_2 + i\left(\int f dm_3 - \int f dm_4\right) \quad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

It is tempting to define the complex measure m by

$$m = m_1 - m_2 + i(m_3 - m_4),$$

but there is a problem, which is that the positive measures m_i may take the value $+\infty$, so expressions of the form $+\infty - (+\infty)$ may arise, but they do not make any sense!

We are not aware of a way around this problem in general. If X is compact, then the Radon–Riesz II theorem yields positive Radon measures m_i such that $m_i(X)$ is finite for $i = 1, \ldots, 4$, in which case the expression m is indeed a measure. It is even possible to define a bijective correspondence by adding disjointness conditions on the subsets over which the m_i are concentrated. Such results are given in Malliavin [68] (Chapter II, Section 5).

Another situation where m is a complex measure is the case where the Radon functional Φ is bounded (continuous). This is the object of the next section.

7.8 The Radon–Riesz Theorem and Bounded Radon Functionals

Let $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ be a bounded Radon functional. In this case the operator norm $\|\Phi\|$ is finite. Recall that

$$\|\Phi\| = \sup\{|\Phi(f)| \mid f \in \mathcal{K}_{\mathbb{C}}(X),\ \|f\|_\infty \leq 1\} = \sup\{|\Phi(f)| \mid f \in \mathcal{K}_{\mathbb{C}}(X),\ \|f\|_\infty = 1\}.$$

The following result is shown in Dieudonné [24] (Chapter VII, Section 20).

Proposition 7.28. *Given a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$, the norm $\|\Phi\|$ is finite, that is, Φ is bounded, iff $|\Phi|$ is is bounded. In this case, $\|\Phi\| = \||\Phi|\|$.*

We deduce that $\Phi = \Phi_r + i\Phi_i$ is bounded iff Φ_r and Φ_i are bounded (see Definition 7.20). But we also see that a real bounded Radon functional $\Psi = \Psi^+ - \Psi^-$ is bounded iff the positive Radon functionals Ψ^+ and Ψ^- are bounded (see Definition 7.21).

Proposition 7.29. *A Radon functional Φ is bounded iff the positive Radon functional $\Phi_r^+, \Phi_r^-, \Phi_i^+, \Phi_i^-$ are bounded.*

If m_1, m_2, m_3, m_4 are the positive σ-Radon measures representing $\Phi_r^+, \Phi_r^-, \Phi_i^+, \Phi_i^-$ given by the Radon–Riesz I theorem (Theorem 7.8), it turns out that they are all finite measures, so $m = m_1 - m_2 + i(m_3 - m_4)$ is a complex measure, and it represents Φ on functions in $\mathcal{C}_0(X)$. In order to state a suitable version of the Radon–Riesz correspondence, we need the following definition.

Definition 7.22. Let X be a locally compact (Hausdorff) space. A complex measure μ on the σ-algebra $\mathcal{B}$ of Borel sets of X is a *regular complex Borel measure* if the positive measure $|\mu|$ is a finite Radon measure, that is, a positive Borel measure that is regular and finite ($|\mu|(X)$ is finite). We denote the vector space of regular complex Borel measures $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$. See Figure 7.7.

Since $|\mu|(X)$ is finite, the measure $|\mu|(K)$ of every compact subset K of X is also finite (since X is Hausdorff, every compact subset K of X is closed and thus measurable, and since $K \subseteq X$, we have $|\mu|(K) \leq |\mu|(X)$). Thus the positive Borel measure $|\mu|$ is locally finite.

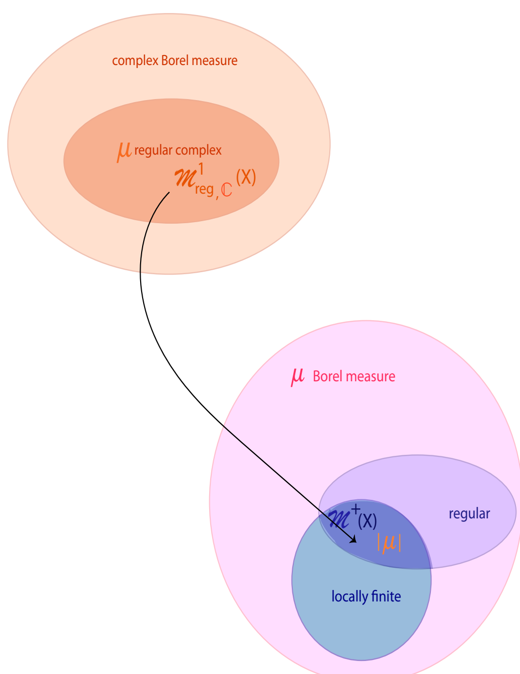


Figure 7.7: A Venn diagram representation of $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$.

We have the following beautiful theorem. Theorem 7.30 is also often referred to as the *Riesz representation theorem*, which is somewhat confusing.

Theorem 7.30. *(Radon–Riesz Correspondence, III) Let X be a locally compact (Hausdorff) space. There are bijections $m\colon \mathrm{M}^1(X) \to \mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$ and $\varphi\colon \mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X) \to \mathrm{M}^1(X)$ between the Banach space $\mathrm{M}^1(X) = \mathcal{C}_0(X,\mathbb{C})'$ of bounded Radon functionals, the dual of the space $\mathcal{C}_0(X,\mathbb{C})$ of continuous functions that tend to zero at infinity, and the Banach space $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$ of regular complex Borel measures. For every regular complex Borel measure $m \in \mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$, the bounded Radon functionals $\varphi(m) = \varphi_m$ is given by*

$$\varphi_m(f) = \int f dm, \quad \text{for all } f \in \mathcal{C}_0(X,\mathbb{C}).$$

For every bounded Radon functional $\Phi \in \mathrm{M}^1(X) = \mathcal{C}_0(X,\mathbb{C})'$, the regular complex Borel measure m_Φ represents Φ in the sense that

$$\begin{aligned}\Phi(f) &= \int f dm_\Phi \\ &= \int f d(m_\Phi)_r^+ - \int f d(m_\Phi)_r^- + i\left(\int f d(m_\Phi)_i^+ - \int f d(m_\Phi)_i^-\right) \quad \text{for all } f \in \mathcal{C}_0(X,\mathbb{C}).\end{aligned}$$

Furthermore, these bijections are norm preserving, that is, $\|\Phi\| = \|m_\Phi\| = |m_\Phi|(X)$. See Figure 7.8.

Theorem 7.30 is proven in Lang [62] (Chapter IX, §4, Theorem 4.2), Rudin [79] (Chapter 6 Theorem 6.19), Folland [34] (Chapter 7, Theorem 7.17), and Marle [69] (Chapter 9, Section 7, Proposition 9.7.3). The proof is quite involved. Among other things it uses Lusin's theorem (Theorem 7.9). It also uses the corollary of the Radon–Nikodym theorem (Theorem 7.21) and the fact that $\mathcal{K}_\mathbb{C}(X)$ is dense in $\mathcal{L}_{|\mu|}(X,\mathcal{B},\mathbb{C})$ to prove injectivity.

To prove surjectivity, by Proposition 7.27 we express the bounded Radon functional $\Phi\colon \mathcal{K}_\mathbb{C}(X) \to \mathbb{C}$ in terms of four positive Radon functionals:

$$\Phi = \Phi_r^+ - \Phi_r^- + i(\Phi_i^+ - \Phi_i^-).$$

By Proposition 7.29, these positive Radon functionals are bounded. By the Radon–Riesz theorem I (Theorem 7.8), there exist four positive σ-Radon measures m_1, m_2, m_3, m_4 such that

$$\Phi(f) = \int f dm_1 - \int f dm_2 + i\left(\int f dm_3 - \int f dm_4\right) \quad \text{for all } f \in \mathcal{K}_\mathbb{C}(X).$$

The reason why the σ-Radon measure m corresponding to a positive bounded Radon functional Φ is finite is that this measure is inner regular, that is,

$$m_\Phi(E) = \sup\{\mu(K) \mid K \subseteq E,\ K \text{ compact}\}$$

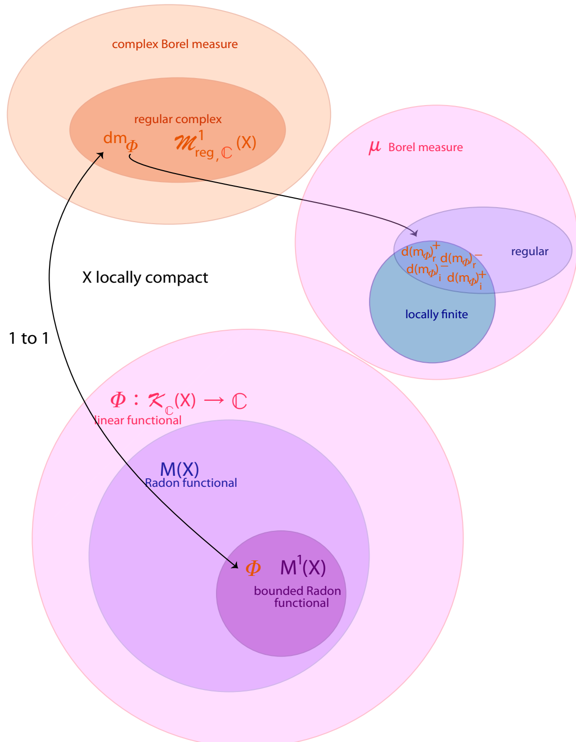


Figure 7.8: Radon-Riesz Correspondence, Version 3.

for every $E \in \mathcal{B}$. We use this to compute $m_\Phi(X)$. For every compact subset K, by Proposition A.39, there is a continuous function $f\colon X \to [0,1]$ of compact support such that $f(x) = 1$ for all $x \in K$. Then since Φ is bounded we have

$$m_\Phi(K) \leq \int f dm_\Phi = \Phi(f) \leq \|\Phi\| \, \|f\|_\infty = \|\Phi\|$$

since f has maximum value 1. Therefore,

$$m_\Phi(X) = \sup\{\mu(K) \mid K \subseteq X, \ K \text{ compact}\} \leq \|\Phi\|$$

is indeed finite. Since $m_\Phi(X)$ is finite, every measurable subset has finite measure and so the σ-regular measure m_Φ is actually regular.

We also need to check that $\varphi_m(f) = \int f dm$ is finite for every function $f \in \mathcal{C}_0(X;\mathbb{C})$ and every positive finite Borel measure m. Since $\mathcal{C}_0(X;\mathbb{C})$ is the closure of $\mathcal{K}_{\mathbb{C}}(X)$, there is a sequence (f_n) of functions $f_n \in \mathcal{K}_{\mathbb{C}}(X)$ that converges to f according to the sup norm, and thus converges pointwise to f. Also f is a bounded function, so there is some $M > 0$ such that $|f_n| \leq M$ for all $n \geq 1$. Since $m(X)$ is finite, the constant function M is integrable, and the continuous functions f_n are integrable. By the dominated convergence theorem (Theorem 5.34), f is integrable.

Theorem 7.30 plays a crucial role in defining the notion of convolutions of two measures in $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$. We will need the following simple fact.

Proposition 7.31. *Let X be any locally compact space, and let μ be any positive Borel measure on $\mathcal{B}$. For any function $f \in \mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$, the functional $\Phi_{f,\mu} \colon \mathcal{C}_0(X; \mathbb{C}) \to \mathbb{C}$ given by*

$$\Phi_{f,\mu}(g) = \int f g \, d\mu \quad \textit{for all } g \in \mathcal{C}_0(X; \mathbb{C})$$

is a bounded Radon functional.

Proof. Since $f \in \mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$ and g is continuous, g is measurable, and $|g|$ is bounded by some $M > 0$, so by Proposition 5.36(1) $fg \in \mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$. We have

$$|\Phi_{f,\mu}(g)| = \left| \int f g \, d\mu \right| \le \int |fg| \, d\mu = \int |f||g| \, d\mu \le \|g\|_\infty \int |f| d\mu,$$

which shows that $\Phi_{f,\mu}$ is bounded. ☐

By Theorem 7.30, the bounded Radon functional $\Phi_{f,\mu}$ corresponds to a unique regular complex Borel measure m such that

$$\int f g \, d\mu = \int g \, dm \quad \text{for all } g \in \mathcal{C}_0(X; \mathbb{C}).$$

The measure m is usually denoted by $f d\mu$. Proposition 7.31 gives us an embedding of $\mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$ into $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$ as stated in the next proposition.

Proposition 7.32. *Let X be a locally compact space. For every positive Borel measure μ on $\mathcal{B}$, the map $f \mapsto f d\mu$ is a norm-preserving embedding of $\mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$ into the space $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$ of regular complex Borel measures on X, with the property that*

$$\int f g \, d\mu = \int g \, f d\mu \quad \textit{for all } g \in \mathcal{C}_0(X; \mathbb{C}).$$

The reason why the embedding is norm-preserving is quite subtle. By Theorem 7.30, $\|f d\mu\| = \|\Phi_{f,\mu}\|$, where $\Phi_{f,\mu} \colon \mathcal{C}_0(X; \mathbb{C}) \to \mathbb{C}$ is the bounded Radon functional given by

$$\Phi_{f,\mu}(g) = \int f g \, d\mu \quad \text{for all } g \in \mathcal{C}_0(X; \mathbb{C}).$$

By an exercise in Folland [34] (Chapter 7, Section 7.2, Exercise 9), the measure $f d\mu$ associated with the functional $\Phi_{f,\mu}$ is equal to the measure μ_f of Proposition 7.16, with

$$\mu_f(A) = \int_A f \, d\mu, \quad A \in \mathcal{B},$$

when f is a positive continuous function. In this case, $\mu_f = fd\mu$ is a positive Radon measure. By Proposition 7.20,

$$\|\mu_f\| = \|f\|_1 \,,$$

so

$$\|\Phi_{f,\mu}\| = \|fd\mu\| = \|\mu_f\| = \|f\|_1 \,.$$

This fact is extended to continuous functions $f\colon X \to \mathbb{C}$ by writing $f = f_1 - f_2 + i(f_3 - f_4)$, where f_1, f_2, f_3, f_4 are four positive continuous functions. Finally, since $\mathcal{K}_{\mathbb{C}}(X)$ is dense in $\mathcal{L}^1(X, \mathcal{B}, \mathbb{C})$, the fact that $\|fd\mu\| = \|\Phi_{f,\mu}\| = \|f\|_1$ is extended to functions in $\mathcal{L}^1(X, \mathcal{B}, \mathbb{C})$.

This embedding is technically important because if X is a locally compact group and if μ is a Haar measure, convolution can be defined on both $\mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$ and $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$, but there is no identity element for convolution on $\mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$ while there is one for convolution on $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$. Technically $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$ is a unital normed Banach algebra but $\mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$ is a nonunital normed Banach algebra. This point will be significant in Chapter 9 and in Chapter 10.

Chapter 8

The Haar Measure and Convolution

Let G be a locally compact group. Haar proved (1933) the remarkable fact that there is a positive σ-regular locally finite Borel measure μ on G such that $\mu(U) > 0$ for every nonempty open subset U, and such that μ is left-invariant, which means that

$$\mu(A) = \mu(sA) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B},$$

where $\mathcal{B}$ is the σ-algebra of Borel sets on G. Furthermore, such a left-invariant measure is unique up to a positive scalar.

Actually, Haar proved the existence of a left-invariant measure in a special case. This result was established in full generality later by André Weil [105]. All proofs we are aware of (Weil [105], Halmos [44], Bourbaki [7], Dieudonné [24], Lang [62], Folland [33]) make use of Haar's original clever idea (1933). Except for Halmos who constructs directly a measure (as Haar did), all the other proofs are essentially André Weil's proof (which constructs a Haar functional) from his famous little book [105] first published in 1940.

In this chapter we sketch the existence of the (left) Haar measure, providing most details, and we also prove its uniqueness up to a scalar; see Sections 8.2, 8.3, 8.4. Some Examples are given in Section 8.5.

For any $s \in G$ and any measure μ on G, let $\rho_s(\mu)$ be the measure given by

$$(\rho_s(\mu))(A) = \mu(As) \quad \text{for all } A \in \mathcal{B}.$$

If μ is a left Haar measure, then it is easy to see that $\rho_s(\mu)$ is a left Haar measure, so by uniqueness up to a scalar, there is a unique positive number $\Delta(s)$ such that

$$\rho_s(\mu) = \Delta(s)\mu \tag{$*$}$$

The function $\Delta\colon G \to \mathbb{R}_+^*$ (given by $\Delta(s)$ for every $s \in G$) is called the *modular function* of G. We investigate properties of the modular function in Section 8.6. We say that the group G is *unimodular* if $\Delta(s) = 1$ for all $s \in G$, equivalently, if and only if a left Haar measure is

also a right Haar measure. If G is abelian, compact, or a connected semisimple Lie group, then G is unimodular. More examples of Haar measures are given in Section 8.7.

Let G be a locally compact group, and let $u\colon G \to G$ be an automorphism of G. For every left Haar measure μ, define the measure $u^{-1}(\mu)$ by

$$(u^{-1}(\mu))(A) = \mu(u(A)), \quad \text{for all } A \in \mathcal{B}.$$

It can be shown that there is a unique positive number $\mathrm{mod}(u)$ such that

$$u^{-1}(\mu) = \mathrm{mod}(u)\mu$$

for all left Haar measures μ. The number $\mathrm{mod}(u)$ is called the *modulus of the automorphism u*. Properties of the modulus of an automorphism are discussed Section 8.8. As an application, we obtain formulae for the measure (volume) of a parallelotope and of a simplex.

Some applications of the Haar measure are discussed in Section 8.9. In particular, we prove Theorem 8.36, a basic tool in representation theory.

Let G be a locally compact group, let X be a locally compact space, and let $\cdot\colon G \times X \to X$ be a continuous left action of G on X. A Borel measure μ on X is *G-invariant* if

$$\mu(s^{-1} \cdot A) = \mu(A) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B}.$$

Our goal is to find sufficient conditions to ensure that X has some G-invariant measure. We will consider the case where $X = G/H$, with the left action of G on G/H given by

$$a \cdot (bH) = abH, \qquad a, b \in G.$$

In this case, by Proposition 8.6, the space X is also locally compact (and Hausdorff).

A G-invariant measure on G/H does not always exist. It turns out that there is a necessary and sufficient condition for a G-invariant σ-Radon measure to exist on G/H in terms of Δ_G and Δ_H: Δ_H must the equal to the restriction of Δ_G on H. This topic is discussed in Section 8.10.

One of the main applications of the Haar measure is the definition of the notion of convolution on a locally compact group. Recall that $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(G)$ denotes the Banach space of complex regular Borel measures on G (see Definition 7.22), and that $\mathrm{L}^1_\lambda(G, \mathcal{B}, \mathbb{C})$ denotes the space of integrable functions on the measure space $(G, \mathcal{B}, \lambda)$, where $\mathcal{B}$ is the σ-algebra of Borel sets of G. To simplify notation, we write $\mathcal{M}^1(G)$ for $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(G)$, and $\mathrm{L}^1(G)$ for $\mathrm{L}^1_\lambda(G, \mathcal{B}, \mathbb{C})$. The vector space $\mathcal{M}^1(G)$ is a Banach space with the norm $\|\mu\| = |\mu|(G)$, and $\mathrm{L}^1(G)$ is a Banach space with the L^1-norm. There are three flavors of convolutions but we will use mostly the first two of them:

1. Convolutions of two measures $\mu, \nu \in \mathcal{M}^1(G)$. This makes $\mathcal{M}^1(G)$ into a Banach algebra with identity and with an involution.

2. Convolution of two functions $f, g \in \mathrm{L}^1(G)$, which makes $\mathrm{L}^1(G)$ into a Banach algebra with involution, but without a multiplicative unit element, unless G is discrete.

3. There is also a notion of convolution of a measure $\mu \in \mathcal{M}^1(G)$ and of a function $f \in \mathrm{L}^1(G)$, and of a function $f \in \mathrm{L}^1(G)$ and a measure $\mu \in \mathcal{M}^1(G)$.

These notions of convolution are discussed in Sections 8.11, 8.12, 8.13.

Convolution applied to functions and measures can be used as a regularization (or filtering) process; see Section 8.14.

8.1 Topological Groups

Since locally compact groups (and Lie groups) are topological groups, it is useful to gather a few basic facts about topological groups.

Definition 8.1. A set G is a *topological group* iff

(a) G is a Hausdorff topological space;

(b) G is a group (with identity 1);

(c) Multiplication $\cdot\colon G \times G \to G$, and the inverse operation $G \longrightarrow G\colon g \mapsto g^{-1}$, are continuous, where $G \times G$ has the product topology.

It is easy to see that the two requirements of Condition (c) are equivalent to

(c′) The map $G \times G \longrightarrow G\colon (g, h) \mapsto gh^{-1}$ is continuous.

Proposition 8.1. *If G is a topological group and H is any subgroup of G, then the closure $\overline{H}$ of H is a subgroup of G. If H is a normal subgroup of G, then $\overline{H}$ is also a normal subgroup of G.*

Proof. We use the fact that if $f\colon X \to Y$ is a continuous map between two topological spaces X and Y, then $f(\overline{A}) \subseteq \overline{f(A)}$ for any subset A of X. For any $a \in \overline{A}$, we need to show that for any open subset $W \subseteq Y$ containing $f(a)$, we have $W \cap f(A) \neq \emptyset$. Since f is continuous, $V = f^{-1}(W)$ is an open subset containing a, and since $a \in \overline{A}$, we have $f^{-1}(W) \cap A \neq \emptyset$, so there is some $x \in f^{-1}(W) \cap A$, which implies that $f(x) \in W \cap f(A)$, so $W \cap f(A) \neq \emptyset$, as desired. The map $f\colon G \times G \to G$ given by $f(x, y) = xy^{-1}$ is continuous, and since H is a subgroup of G, $f(H \times H) \subseteq H$. By the above property, if $a \in \overline{H}$ and if $b \in \overline{H}$, that is, $(a, b) \in \overline{H \times H}$, then $f(a, b) = ab^{-1} \in \overline{H}$, which shows that $\overline{H}$ is a subgroup of G.

For every $g \in G$, the map $C_g\colon G \to G$ given by $C_g(x) = gxg^{-1}$ for all $x \in G$ is continuous, and if H is a normal subgroup of G, then $C_g(H) \subseteq H$. It follows that $C_g(\overline{H}) \subseteq \overline{H}$ for all $g \in G$, which means that $\overline{H}$ is a normal subgroup of G. $\square$

Given a topological group G, for every $a \in G$ we define the *left translation* L_a as the map $L_a\colon G \to G$ such that $L_a(b) = ab$, for all $b \in G$, and the *right translation* R_a as the map $R_a\colon G \to G$ such that $R_a(b) = ba$, for all $b \in G$. Observe that $L_{a^{-1}}$ is the inverse of L_a and similarly, $R_{a^{-1}}$ is the inverse of R_a. As multiplication is continuous, we see that L_a and R_a are continuous. Moreover, since they have a continuous inverse, they are homeomorphisms. As a consequence, if U is an open subset of G, then so is $gU = L_g(U)$ (resp. $Ug = R_gU$), for all $g \in G$. Therefore, the topology of a topological group is *determined* by the knowledge of the open subsets containing the identity 1.

Given any subset $S \subseteq G$, let $S^{-1} = \{s^{-1} \mid s \in S\}$; let $S^0 = \{1\}$, and $S^{n+1} = S^nS$, for all $n \geq 0$. Property (c) of Definition 8.1 has the following useful consequences, which shows there exists an open set containing 1 which has a special symmetrical structure.

Proposition 8.2. *If G is a topological group and U is any open subset containing 1, then there is some open subset $V \subseteq U$, with $1 \in V$, so that $V = V^{-1}$ and $V^2 \subseteq U$. Furthermore, $\overline{V} \subseteq U$.*

Proof. Since multiplication $G \times G \longrightarrow G$ is continuous and $G \times G$ is given the product topology, there are open subsets U_1 and U_2, with $1 \in U_1$ and $1 \in U_2$, so that $U_1U_2 \subseteq U$. Let $W = U_1 \cap U_2$ and $V = W \cap W^{-1}$. Then V is an open set containing 1, and clearly $V = V^{-1}$ and $V^2 \subseteq U_1U_2 \subseteq U$. If $g \in \overline{V}$, then gV is an open set containing g (since $1 \in V$) and thus, $gV \cap V \neq \emptyset$. This means that there are some $h_1, h_2 \in V$ so that $gh_1 = h_2$, but then, $g = h_2h_1^{-1} \in VV^{-1} = VV \subseteq U$. ∎

Definition 8.2. A subset U containing 1 and such that $U = U^{-1}$ is called *symmetric*.

Proposition 8.2 is used in the proofs of many the propositions and theorems on the structure of topological groups. For example, it is key in verifying the following proposition regarding discrete topological subgroups.

Definition 8.3. A subgroup H of a topological group G is *discrete* iff the induced topology on H is discrete; that is, for every $h \in H$, there is some open subset U of G so that $U \cap H = \{h\}$.

Proposition 8.3. *If G is a topological group and H is a discrete subgroup of G, then H is closed.*

Proof. As H is discrete, there is an open subset U of G so that $U \cap H = \{1\}$, and by Proposition 8.2, we may assume that $U = U^{-1}$. Our goal is to show $H = \overline{H}$. Clearly $H \subseteq \overline{H}$. Thus it remains to show $\overline{H} \subseteq H$. If $g \in \overline{H}$, as gU is an open set containing g, we have $gU \cap H \neq \emptyset$. Consequently, there is some $y \in gU \cap H = gU^{-1} \cap H$, so $g \in yU$ with $y \in H$. We claim that $yU \cap H = \{y\}$. Note that $x \in yU \cap H$ means $x = yu_1$ with $yu_1 \in H$ and $u_1 \in U$. Since H is a subgroup of G and $y \in H$, $y^{-1}yu_1 = u_1 \in H$. Thus $u_1 \in U \cap H$, which implies $u_1 = 1$ and $x = yu_1 = y$, and we have

$$g \in yU \cap \overline{H} \subseteq \overline{yU \cap H} = \overline{\{y\}} = \{y\}.$$

since G is Hausdorff. Therefore, $g = y \in H$. ∎

Using Proposition 8.2, we can give a very convenient characterization of the Hausdorff separation property in a topological group.

Proposition 8.4. *If G is a topological group, then the following properties are equivalent:*

(1) G is Hausdorff;

(2) The set $\{1\}$ is closed;

(3) The set $\{g\}$ is closed, for every $g \in G$.

Proof. The implication (1) $\longrightarrow$ (2) is true in any Hausdorff topological space. We just have to prove that $G - \{1\}$ is open, which goes as follows: For any $g \neq 1$, since G is Hausdorff, there exists disjoint open subsets U_g and V_g, with $g \in U_g$ and $1 \in V_g$. Thus, $\bigcup U_g = G - \{1\}$, showing that $G - \{1\}$ is open. Since L_g is a homeomorphism, (2) and (3) are equivalent. Let us prove that (3) $\longrightarrow$ (1). Let $g_1, g_2 \in G$ with $g_1 \neq g_2$. Then, $g_1^{-1}g_2 \neq 1$ and if U and V are disjoint open subsets such that $1 \in U$ and $g_1^{-1}g_2 \in V$, then $g_1 \in g_1U$ and $g_2 \in g_1V$, where g_1U and g_1V are still open and disjoint. Thus, it is enough to separate 1 and $g \neq 1$. Pick any $g \neq 1$. If every open subset containing 1 also contained g, then 1 would be in the closure of $\{g\}$, which is absurd since $\{g\}$ is closed and $g \neq 1$. Therefore, there is some open subset U such that $1 \in U$ and $g \notin U$. By Proposition 8.2, we can find an open subset V containing 1, so that $VV \subseteq U$ and $V = V^{-1}$. We claim that V and gV are disjoint open sets with $1 \in V$ and $g \in gV$.

Since $1 \in V$, it is clear that $g \in gV$. If we had $V \cap gV \neq \emptyset$, then by the last sentence in the proof of Proposition 8.2 we would have $g \in VV^{-1} = VV \subseteq U$, a contradiction. ☐

If H is a subgroup of G (not necessarily normal), we can form the set of left cosets G/H, and we have the projection $p\colon G \to G/H$, where $p(g) = gH = \overline{g}$. If G is a topological group, then G/H can be given the *quotient topology*, where a subset $U \subseteq G/H$ is open iff $p^{-1}(U)$ is open in G. With this topology, p is continuous. The trouble is that G/H is not necessarily Hausdorff. However, we can neatly characterize when this happens.

Proposition 8.5. *If G is a topological group and H is a subgroup of G, then the following properties hold:*

(1) The map $p\colon G \to G/H$ is an open map, which means that $p(V)$ is open in G/H whenever V is open in G.

(2) The space G/H is Hausdorff iff H is closed in G.

(3) If H is open, then H is closed and G/H has the discrete topology (every subset is open).

(4) The subgroup H is open iff $1 \in \overset{\circ}{H}$ (i.e., there is some open subset U so that $1 \in U \subseteq H$).

Proof. (1) Observe that if V is open in G, then $VH = \bigcup_{h\in H} Vh$ is open, since each Vh is open (as right translation is a homeomorphism). However, it is clear that

$$p^{-1}(p(V)) = VH,$$

i.e., $p^{-1}(p(V))$ is open which, by definition of the quotient topology, means that $p(V)$ is open.

(2) If G/H is Hausdorff, then by Proposition 8.4, every point of G/H is closed, i.e., each coset gH is closed, so H is closed. Conversely, assume H is closed. Let $\overline{x}$ and $\overline{y}$ be two distinct point in G/H and let $x, y \in G$ be some elements with $p(x) = \overline{x}$ and $p(y) = \overline{y}$. As $\overline{x} \neq \overline{y}$, the elements x and y are not in the same coset, so $x \notin yH$. As H is closed, so is yH, and since $x \notin yH$, there is some open containing x which is disjoint from yH, and we may assume (by translation) that it is of the form Ux, where U is an open containing 1. By Proposition 8.2, there is some open V containing 1 so that $VV \subseteq U$ and $V = V^{-1}$. Thus, we have

$$V^2x \cap yH = \emptyset$$

and in fact,

$$V^2xH \cap yH = \emptyset,$$

since H is a group; if $z \in V^2xH \cap yH$, then $z = v_1v_2xh_1 = yh_2$ for some $v_1, v_2 \in V$, and some $h_1, h_2 \in H$, but then $v_1v_2x = yh_2h_1^{-1}$ so that $V^2x \cap yH \neq \emptyset$, a contradiction. Since $V = V^{-1}$, we get

$$VxH \cap VyH = \emptyset,$$

and then, since V is open, both VxH and VyH are disjoint, open, so $p(VxH)$ and $p(VyH)$ are open sets (by (1)) containing $\overline{x}$ and $\overline{y}$ respectively and $p(VxH)$ and $p(VyH)$ are disjoint (because $p^{-1}(p(VxH)) = VxHH = VxH$, $p^{-1}(p(VyH)) = VyHH = VyH$, and $VxH \cap VyH = \emptyset$). See Figure 8.1.

(3) If H is open, then every coset gH is open, so every point of G/H is open and G/H is discrete. Also, $\bigcup_{g\notin H} gH$ is open, i.e., H is closed.

(4) Say U is an open subset such that $1 \in U \subseteq H$. Then for every $h \in H$, the set hU is an open subset of H with $h \in hU$, which shows that H is open. The converse is trivial. $\square$

Recall that a topological space X is *locally compact* iff for every point $p \in X$, there is a compact neighborhood C of p; that is, there is a compact C and an open U, with $p \in U \subseteq C$. For example, manifolds are locally compact.

The next two propositions will be needed.

Proposition 8.6. *Let G be a topological group and let H be a closed subgroup of G. The following properties hold.*

(1) If G is locally compact, then so is G/H.

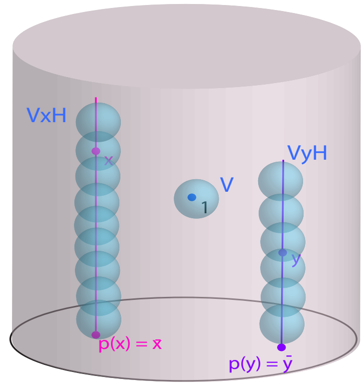


Figure 8.1: A schematic illustration of $VxH \cap VyH = \emptyset$, where G is the pink cylinder, H is the vertical edge, and G/H is the circular base. Note xH and yH are vertical fibres.

(2) If H is a normal subgroup of G, then G/H is a topological group.

Proof. (1) Since H is closed, we already know from Proposition 8.5(2) that G/H is Hausdorff. Let K be a compact neighborhood of 1 in G, so that there is an open subset U such that $1 \in U \subseteq K$ with K compact. By Proposition 8.5(1) the quotient map $p\colon G \to G/H$ is an open map, and it is continuous, so for any $g \in G$, we have $g \in gU \subseteq gK$ with gU open and gK compact, so $p(g) \in p(gU) \subseteq p(gK)$ with $p(gU)$ open and $p(gK)$ compact, which shows that G/H is locally compact.

(2) If H is a closed normal subgroup, then G/H is a group, and we already know from Proposition 8.5(2) that G/H is Hausdorff. We have to show that multiplication and inversion in G/H are continuous. For any two cosets g_1H and g_2H in G/H, if W is an open subset in G/H containing $p(g_1g_2) = p(g_1)p(g_2) = (g_1H)(g_2H) = g_1g_2H$, then because the projection map p is continuous, there are open subsets U_1 and U_2 of G with $g_1 \in U_1$ and $g_2 \in U_2$, such that $p(U_1U_2) \subseteq W$. Since p is an open map, $p(U_1)$ is an open subset containing $p(g_1) = g_1H$ and $p(U_2)$ is an open subset containing $p(g_2) = g_2H$, and we have $p(U_1)p(U_2) \subseteq W$, so multiplication in G/H is continuous. A similar proof shows that inversion is continuous in G/H. $\square$

Proposition 8.7. *Let G be a locally compact topological group and let H be a closed subgroup of G. For any compact subset K' in G/H, there is compact subset K of G such that $p(K) = K'$.*

Proof. Since G is locally compact, there is an open subset U and a compact subset V such that $1 \in U \subseteq V$. Since p is an open map, the subsets of the form $p(gU)$ for $g \in G$ form an open cover of K', and since K' is compact, there is a finite subcover $\{p(g_1U), \ldots, p(g_nU)\}$ of K'. Since p is continuous and K' is compact and thus closed (since G/H is Hausdorff), $p^{-1}(K')$ is closed and $g_1V \cup \cdots \cup g_nV$ is compact; then $K = p^{-1}(K') \cap (g_1V \cup \cdots \cup g_nV)$ is compact in G, and we have $p(K) = K'$. $\square$

We next provide a criterion relating the connectivity of G with that of G/H.

Proposition 8.8. *Let G be a topological group and H be any subgroup of G. If H and G/H are connected, then G is connected.*

Proof. It is a standard fact of topology that a space G is connected iff every continuous function f from G to the discrete space $\{0,1\}$ is constant; see Proposition A.17. Pick any continuous function f from G to $\{0,1\}$. As H is connected and left translations are homeomorphisms, all cosets gH are connected. Thus, f is constant on every coset gH. It follows that the function $f\colon G \to \{0,1\}$ induces a continuous function $\overline{f}\colon G/H \to \{0,1\}$ such that $f = \overline{f} \circ p$ (where $p\colon G \to G/H$; the continuity of $\overline{f}$ follows immediately from the definition of the quotient topology on G/H). As G/H is connected, $\overline{f}$ is constant, and so $f = \overline{f} \circ p$ is constant. $\square$

The next three propositions describe how to generate a topological group from its symmetric neighborhoods of 1.

Proposition 8.9. *If G is a connected topological group, then G is generated by any symmetric neighborhood V of 1. In fact,*

$$G = \bigcup_{n \geq 1} V^n.$$

Proof. Since $V = V^{-1}$, it is immediately checked that $H = \bigcup_{n\geq 1} V^n$ is the group generated by V. As V is a neighborhood of 1, there is some open subset $U \subseteq V$, with $1 \in U$, and so $1 \in \overset{\circ}{H}$. From Proposition 8.5 (3), the subgroup H is open and closed, and since G is connected, $H = G$. $\square$

Proposition 8.10. *Let G be a topological group and let V be any connected symmetric open subset containing 1. Then if G_0 is the connected component of the identity, we have*

$$G_0 = \bigcup_{n \geq 1} V^n,$$

and G_0 is a normal subgroup of G. Moreover, the group G/G_0 is discrete.

Proof. First, as V is open, every V^n is open, so the group $\bigcup_{n\geq 1} V^n$ is open, and thus closed, by Proposition 8.5 (3). For every $n \geq 1$, we have the continuous map

$$\underbrace{V \times \cdots \times V}_{n} \longrightarrow V^n : (g_1, \ldots, g_n) \mapsto g_1 \cdots g_n.$$

As V is connected, $V \times \cdots \times V$ is connected, and so V^n is connected; this follows from Proposition A.18 because a finite product of connected spaces is connected. Since $1 \in V^n$ for all $n \geq 1$ and every V^n is connected, we use Lemma A.19 to conclude that $\bigcup_{n\geq 1} V^n$ is connected. Now, $\bigcup_{n\geq 1} V^n$ is connected, open and closed, so it is the connected component of 1. Finally, for every $g \in G$, the group gG_0g^{-1} is connected and contains 1, so it is contained in G_0, which proves that G_0 is normal. Since G_0 is open, Proposition 8.5 (3) implies that the group G/G_0 is discrete. $\square$

Proposition 8.11. *Let G be a topological group and assume that G is connected and locally compact. Then G is countable at infinity, which means that G is the union of a countable family of compact subsets. In fact, if V is any symmetric compact neighborhood of 1, then*

$$G = \bigcup_{n\geq 1} V^n.$$

Proof. Since G is locally compact, there is some compact neighborhood K of 1. Then, $V = K \cap K^{-1}$ is also compact and a symmetric neighborhood of 1. By Proposition 8.9, we have

$$G = \bigcup_{n\geq 1} V^n.$$

An argument similar to the one used in the proof of Proposition 8.10 to show that V^n is connected if V is connected proves that each V^n compact if V is compact. $\square$

If G is a locally compact group but G is not connected, and if G_0 is the connected component of the identity, then G is the disjoint union of the cosets gG_0, and each coset gG_0 is homeomorphic to G_0, connected, and countable at infinity (σ-compact). This observation plays a crucial role in the proof of the uniqueness of the Haar measure (Theorem 8.21), because it guarantees that the use of Fubini's theorem is legitimate.

The notion of uniform continuity can be generalized to functions defined on a group.

Definition 8.4. Given a topological group G and a subset S of G, for any normed vector space F, a function $f\colon G \to F$ is *left uniformly continuous on S* if for any $\epsilon > 0$, there is an open subset U of G containing 1 such that

$$\|f(y) - f(x)\| < \epsilon \quad \text{for all } x, y \in S \text{ such that } xy^{-1} \in U.$$

The function $f\colon G \to F$ is *right uniformly continuous on S* if for any $\epsilon > 0$, there is an open subset U of G containing 1 such that

$$\|f(y) - f(x)\| < \epsilon \quad \text{for all } x, y \in S \text{ such that } x^{-1}y \in U.$$

Observe that if $xy^{-1} \in U$, then we can write $xy^{-1} = z$ for some $z \in U$, so $y = z^{-1}x$, and $\|f(y) - f(x)\| = \|f(z^{-1}x) - f(x)\| < \epsilon$.

It is customary to introduce a left action λ of G on functions $f\colon G \to F$ defined on G by

$$(\lambda_s(f))(x) = f(s^{-1}x) \quad \text{for all } x, s \in G.$$

Observe that

$$\lambda_{st}(f)(x) = f((st)^{-1}x) = f(t^{-1}s^{-1}x) = \lambda_t(f)(s^{-1}x) = \lambda_s(\lambda_t(f))(x),$$

so

$$\lambda_{st} = \lambda_s \circ \lambda_t,$$

which is the reason why we used s^{-1} instead of s in the definition of λ_s.

Then $\|f(z^{-1}x) - f(x)\| = \|\lambda_z(f)(x) - f(x)\|$, so the condition of the definition is equivalent to

$$\|\lambda_z(f)(x) - f(x)\| < \epsilon \quad \text{for all } x \in S \text{ and all } z \in U.$$

Informally, the above condition can be written as

$$\lim_{z \mapsto 1} \sup_{x \in S} \|\lambda_z(f)(x) - f(x)\| = 0.$$

It is also customary to introduce a right action ρ of G on functions $f\colon G \to F$ defined on G by

$$(\rho_s(f))(x) = f(xs) \quad \text{for all } x, s \in G.$$

Observe that

$$\rho_{st}(f)(x) = f(xst) = \rho_t(f)(xs) = \rho_s(\rho_t(f))(x),$$

so

$$\rho_{st} = \rho_s \circ \rho_t.$$

Observe that if $x^{-1}y \in U$, then we can write $x^{-1}y = z$ for some $z \in U$, so $y = xz$, and $\|f(y) - f(x)\| = \|f(xz) - f(x)\| = \|\rho_z(f)(x) - f(x)\| < \epsilon$.

Thus the condition of the definition is equivalent to

$$\|\rho_z(f)(x) - f(x)\| < \epsilon \quad \text{for all } x \in S \text{ and all } z \in U.$$

Informally, the above condition can be written as

$$\lim_{z \mapsto 1} \sup_{x \in S} \|\rho_z(f)(x) - f(x)\| = 0.$$

Proposition 8.12. *Let G be a topological group and let S be a subset of G. For any function $f\colon S \to F$, where F is any normed vector space, if f is continuous with compact support K, then f is left (resp. right) uniformly continuous on K.*

Proof. We prove that f is left uniformly continuous, the proof that f is right uniformly continuous being similar and left as an exercise. Since f is continuous, for every $y \in K$, there is some open subset U_y with $1 \in U_y$ such that

$$\|f(y) - f(x)\| < \frac{\epsilon}{2} \quad \text{for all } x \in U_y y.$$

We can find an open subset V_y containing 1 such that $V_y V_y \subseteq U_y$. The open subsets of the form $V_y y$ for $y \in K$ form an open cover of K, and since K is compact, there is a finite subcover $\{V_{y_1} y_1 \ldots, V_{y_n} y_n\}$ of K with $y_1, \ldots, y_n \in K$. Let

$$V = V_{y_1} \cap \cdots \cap V_{y_n}.$$

Consider $x, y \in K$ such that $xy^{-1} \in V$, that is, $x \in Vy$. Then $y \in V_{y_i}y_i \subseteq U_{y_i}y_i$ for some i, and so

$$x \in Vy \in VV_{y_i}y_i \subseteq V_{y_i}V_{y_i}y_i \subseteq U_{y_i}y_i,$$

which implies that

$$\|f(y) - f(x)\| \leq \|f(y) - f(y_i)\| + \|f(y_i) - f(x)\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

as desired. $\square$

We end this section by combining the various properties of a topological group G to characterize when G/G_x is homeomorphic to X. The reader should review the notion of group action and the related concepts of stabilizer and orbit; see Appendix C, Sections C.2 and C.3.

First we need two definitions.

Definition 8.5. Let G be a topological group and let X be a topological space. An action $\varphi\colon G \times X \to X$ is *continuous* (and G *acts continuously on* X) if the map φ is continuous.

If an action $\varphi\colon G \times X \to X$ is continuous, then each map $\varphi_g\colon X \to X$ is a homeomorphism of X (recall that $\varphi_g(x) = g \cdot x$, for all $x \in X$). Indeed, the map $x \mapsto g \cdot x$ is a continuous bijection whose inverse $x \mapsto g^{-1} \cdot x$ is also continuous.

Under some mild assumptions on G and X, the quotient space G/G_x is homeomorphic to X. For example, this happens if X is a Baire space.

Definition 8.6. A *Baire space* X is a topological space with the property that if $\{F\}_{i\geq 1}$ is any countable family of closed sets F_i such that each F_i has empty interior, then $\bigcup_{i\geq 1} F_i$ also has empty interior. By complementation, this is equivalent to the fact that for every countable family of open sets U_i such that each U_i is dense in X (i.e., $\overline{U}_i = X$), then $\bigcap_{i\geq 1} U_i$ is also dense in X.

Remark: A subset $A \subseteq X$ is *rare* if its closure $\overline{A}$ has empty interior. A subset $Y \subseteq X$ is *meager* if it is a countable union of rare sets. Then it is immediately verified that a space X is a Baire space iff every nonempty open subset of X is not meager.

The following theorem shows that there are plenty of Baire spaces:

Theorem 8.13. *(Baire) (1) Every locally compact topological space is a Baire space.*

(2) Every complete metric space is a Baire space.

A proof of Theorem 8.13 can be found in Bourbaki [14], Chapter IX, Section 5, Theorem 1.

Theorem 8.14. *Let G be a topological group which is locally compact and countable at infinity, X a Hausdorff topological space which is a Baire space, and assume that G acts transitively and continuously on X. Then for any $x \in X$, the map $\varphi\colon G/G_x \to X$ is a homeomorphism.*

Proof. We follow the proof given in Bourbaki [14], Chapter IX, Section 5, Proposition 6 (Essentially the same proof can be found in Mneimné and Testard [73], Chapter 2). First observe that if a topological group acts continuously and transitively on a Hausdorff topological space, then for every $x \in X$, the stabilizer G_x is a closed subgroup of G. This is because, as the action is continuous, the projection $\pi_x\colon G \longrightarrow X\colon g \mapsto g \cdot x$ is continuous, and $G_x = \pi^{-1}(\{x\})$, with $\{x\}$ closed. Therefore, by Proposition 8.5, the quotient space G/G_x is Hausdorff. As the map $\pi_x\colon G \longrightarrow X$ is continuous, the induced map $\varphi_x\colon G/G_x \to X$ is continuous, and by Proposition C.14, it is a bijection. Therefore, to prove that φ_x is a homeomorphism, it is enough to prove that φ_x is an open map. For this, it suffices to show that π_x is an open map. Given any open U in G, we will prove that for any $g \in U$, the element $\pi_x(g) = g \cdot x$ is contained in the interior of $U \cdot x$. However, observe that this is equivalent to proving that x belongs to the interior of $(g^{-1} \cdot U) \cdot x$. Therefore, we are reduced to the following case: if U is any open subset of G containing 1, then x belongs to the interior of $U \cdot x$.

Since G is locally compact, using Proposition 8.2, we can find a compact neighborhood of the form $W = \overline{V}$, such that $1 \in W$, $W = W^{-1}$ and $W^2 \subseteq U$, where V is open with $1 \in V \subseteq U$. As G is countable at infinity, $G = \bigcup_{i \geq 1} K_i$, where each K_i is compact. Since V is open, all the cosets gV are open, and as each K_i is covered by the gV's, by compactness of K_i, finitely many cosets gV cover each K_i, and so

$$G = \bigcup_{i \geq 1} g_i V = \bigcup_{i \geq 1} g_i W,$$

for countably many $g_i \in G$, where each $g_i W$ is compact. As our action is transitive, we deduce that

$$X = \bigcup_{i \geq 1} g_i W \cdot x,$$

where each $g_i W \cdot x$ is compact, since our action is continuous and the $g_i W$ are compact. As X is Hausdorff, each $g_i W \cdot x$ is closed, and as X is a Baire space expressed as a union of closed sets, one of the $g_i W \cdot x$ must have nonempty interior; that is, there is some $w \in W$, with $g_i w \cdot x$ in the interior of $g_i W \cdot x$, for some i. But then, as the map $y \mapsto g \cdot y$ is a homeomorphism for any given $g \in G$ (where $y \in X$), we see that x is in the interior of

$$w^{-1} g_i^{-1} \cdot (g_i W \cdot x) = w^{-1} W \cdot x \subseteq W^{-1} W \cdot x = W^2 \cdot x \subseteq U \cdot x,$$

as desired. ∎

By Theorem 8.13, we get the following important corollary:

Theorem 8.15. *Let G be a topological group which is locally compact and countable at infinity, X a Hausdorff locally compact topological space, and assume that G acts transitively and continuously on X. Then for any $x \in X$, the map $\varphi_x \colon G/G_x \to X$ is a homeomorphism.*

Readers who wish to learn more about topological groups may consult Sagle and Walde [81] and Chevalley [18] for an introductory account, and Bourbaki [13], Weil [105] and Pontryagin [77, 78], for a more comprehensive account (especially the last two references).

8.2 Existence of the Haar Measure; Preliminaries

Let G be a locally compact group. We are going to show there is a positive σ-regular locally finite Borel measure μ on G such that $\mu(U) > 0$ for every nonempty open subset U, and such that μ is left-invariant, which means that

$$\mu(A) = \mu(sA) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B},$$

where $\mathcal{B}$ is the σ-algebra of Borel sets on G.

Recall that for any $x \in G$, the maps $L_x \colon G \to G$ (left translation) and $R_x \colon G \to G$ (right translation) are defined by

$$L_x(z) = xz, \qquad R_x(z) = zx, \quad \text{for all } x, z \in G.$$

It is obvious that

$$L_{xy} = L_x \circ L_y \quad \text{and} \quad R_{xy} = R_y \circ R_x,$$

and that L_x and R_y commute for all $x, y \in G$.

It is customary to introduce a left action λ of G and a right action ρ of G on functions $f \colon G \to F$. We did this in the previous section, but for the sake of completeness, we repeat these definitions.

Definition 8.7. Let G be a group, and let F be any set. The *left action* λ of G on a function $f \colon G \to F$ is the function $\lambda_s(f)$ is given by

$$(\lambda_s(f))(x) = f(s^{-1}x) \quad \text{for all } x, s \in G,$$

and the *right action* ρ of G on a function $f \colon G \to F$ is the function $\rho_s(f)$ given by

$$(\rho_s(f))(x) = f(xs) \quad \text{for all } x, s \in G.$$

It might help the reader to remember that λ_s is a *left* action and that ρ_s is a *right action* by noticing that $\lambda =$ *l*ambda begins with an "l" as in *l*eft and that $\rho =$ *r*ho begin with an "r" as in *r*ight.

Observe that

$$\lambda_{st}(f)(x) = f((st)^{-1}x) = f(t^{-1}s^{-1}x) = \lambda_t(f)(s^{-1}x) = \lambda_s(\lambda_t(f))(x),$$

so

$$\lambda_{st} = \lambda_s \circ \lambda_t,$$

which is the reason why we used s^{-1} instead of s in the definition of λ_s. Observe that

$$\rho_{st}(f)(x) = f(xst) = \rho_t(f)(xs) = \rho_s(\rho_t(f))(x),$$

so

$$\rho_{st} = \rho_s \circ \rho_t.$$

Given a subset A of G , we usually write sA for $L_s(A)$ and As for $R_s(A)$.

We define a left action of λ_s and a right action ρ_s on measures and Radon functionals as follows.

Definition 8.8. Let G be a locally compact topological group. The *left action* λ of G on a measure μ on $(G, \mathcal{B})$ is the measure $\lambda_s(\mu)$ given by

$$(\lambda_s(\mu))(A) = \mu(s^{-1}A) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B},$$

and the *right action* ρ of G on a measure μ on $(G, \mathcal{B})$ is the measure $\rho_s(\mu)$ given by

$$(\rho_s(\mu))(A) = \mu(As) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B}.$$

The *left action* λ of G on a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$ is the Radon functional $\lambda_s(\Phi)$ given by

$$(\lambda_s(\Phi))(f) = \Phi(\lambda_{s^{-1}}(f)) \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_{\mathbb{C}}(G),$$

and the *right action* ρ of G on a Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$ is the Radon functional $\rho_s(\Phi)$ given by

$$(\rho_s(\Phi))(f) = \Phi(\rho_{s^{-1}}(f)) \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_{\mathbb{C}}(G).$$

If m_Φ is the Borel measure corresponding to a positive Radon functional Φ, $m_{\lambda_s(\Phi)}$ is the Borel measure corresponding to $\lambda_s(\Phi)$, and $m_{\rho_s(\Phi)}$ is the Borel measure corresponding to $\rho_s(\Phi)$, given by Theorem 7.8, then we have

$$\begin{aligned}\int f(sx)dm_\Phi(x) &= \int (\lambda_{s^{-1}}(f))(x)dm_\Phi(x) = \Phi(\lambda_{s^{-1}}(f)) \\ &= (\lambda_s(\Phi))(f) = \int f(x)d(m_{\lambda_s(\Phi)})(x)\end{aligned}$$

and

$$\begin{aligned}\int f(xs^{-1})dm_\Phi(x) &= \int (\rho_{s^{-1}}(f))(x)dm_\Phi(x) = \Phi(\rho_{s^{-1}}(f)) \\ &= (\rho_s(\Phi))(f) = \int f(x)d(m_{\rho_s(\Phi)})(x).\end{aligned}$$

Therefore, we have the change of variable formulae

$$\int f(x)d(m_{\lambda_s(\Phi)})(x) = \int f(sx)dm_\Phi(x),$$

and

$$\int f(x)d(m_{\rho_s(\Phi)})(x) = \int f(xs^{-1})dm_\Phi(x)$$

for all $f \in \mathcal{K}_{\mathbb{C}}(G)$ and all $s \in G$. See Figures 8.2 and 8.3.

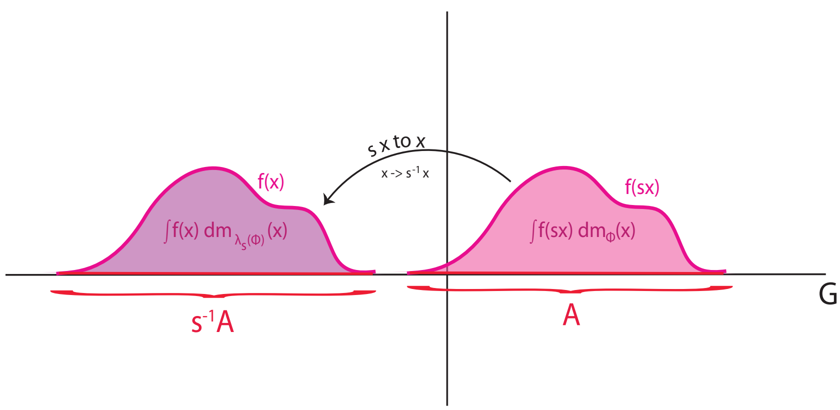


Figure 8.2: A schematic illustration of the change of variable $x \to s^{-1}x$ associated with $\int f(x)d(m_{\lambda_s(\Phi)})(x) = \int f(sx)dm_\Phi(x)$.

Definition 8.8 has been designed so that for every measure μ on G we have

$$\lambda_{st}(\mu) = \lambda_s(\lambda_t(\mu)), \qquad \rho_{st}(\mu) = \rho_s(\rho_t(\mu)) \quad \text{for all } s, t \in G.$$

For every Radon functional Φ and for every function $f \in \mathcal{K}_{\mathbb{C}}(G)$, we have

$$\begin{aligned}(\lambda_{st}(\Phi))(f) &= \Phi(\lambda_{(st)^{-1}}(f)) = \Phi(\lambda_{t^{-1}s^{-1}}(f)) \\ &= \Phi(\lambda_{t^{-1}}(\lambda_{s^{-1}}(f))) = (\lambda_t(\Phi))(\lambda_{s^{-1}}(f)) = (\lambda_s(\lambda_t(\Phi)))(f),\end{aligned}$$

and a similar computation shows that

$$(\rho_{st}(\Phi))(f) = (\rho_s(\rho_t(\Phi)))(f).$$

Therefore, for every Radon functional Φ, we have

$$\lambda_{st}(\Phi) = \lambda_s(\lambda_t(\Phi)), \qquad \rho_{st}(\Phi) = \rho_s(\rho_t(\Phi)) \quad \text{for all } s, t \in G.$$

The left actions λ_s and the right actions ρ_s are summarized in the following table.

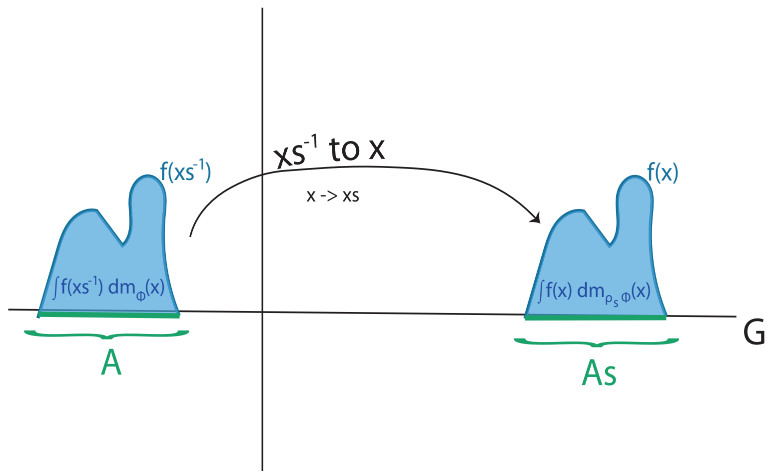


Figure 8.3: A schematic illustration of the change of variable $x \to xs$ associated with $\int f(x)d(m_{\rho_s(\Phi)})(x) = \int f(xs^{-1})dm_\Phi(x)$.

Left action	Right action
On functions	
$(\lambda_s(f))(x) = f(s^{-1}x)$	$(\rho_s(f))(x) = f(xs)$
On measures	
$(\lambda_s(\mu))(A) = \mu(s^{-1}A)$	$(\rho_s(\mu))(A) = \mu(As)$
On functionals	
$(\lambda_s(\Phi))(f) = \Phi(\lambda_{s^{-1}}(f))$	$(\rho_s(\Phi))(f) = \Phi(\rho_{s^{-1}}(f))$

Definition 8.9. Let G be a locally compact group. A *left Haar measure* μ is a σ-regular, locally finite, Borel measure on the σ-algebra $\mathcal{B}$ of Borel sets of G, such that $\mu(U) > 0$ for all nonempty open subsets $U \in \mathcal{B}$ and μ is *left-invariant*, which means that

$$\lambda_s(\mu) = \mu \quad \text{for all } s \in G.$$

The above condition means that

$$\mu(s^{-1}A) = \mu(A) \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G,$$

or equivalently,

$$\mu(sA) = \mu(A) \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G,$$

A *right Haar measure* μ is a Borel measure satisfying the same conditions as a left Haar measure, except that it is *right-invariant*, which means that

$$\rho_s(\mu) = \mu \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G.$$

The above condition means that

$$\mu(As) = \mu(A) \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G.$$

Note that according to Definition 7.5, a left (resp.) Haar measure is a σ-Radon measure which is left-invariant (resp. right-invariant), and such that $\mu(U) > 0$ for all nonempty open subsets $U \in \mathcal{B}$.

In order to prove that a left (resp. right) Haar measure exists, we will use Theorem 7.8, which motivates the following definition.

Definition 8.10. Let G be a locally compact group. A *left Haar functional* Φ is a positive non-zero Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$ which is *left-invariant*, which means that

$$\lambda_s(\Phi) = \Phi \quad \text{for all } s \in G.$$

A *right Haar functional* Φ is a positive non-zero Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$ which is *right-invariant*, which means that

$$\rho_s(\Phi) = \Phi \quad \text{for all } s \in G.$$

If m_Φ is the Borel measure associated with Φ, since

$$\int f(sx)dm_\Phi(x) = \int (\lambda_{s^{-1}}(f))(x)dm_\Phi(x) = \Phi(\lambda_{s^{-1}}(f)) = (\lambda_s(\Phi))(f),$$

then the left-invariance of Φ means that

$$(\lambda_s(\Phi))(f) = \Phi(f) = \int f(x)dm_\Phi(x),$$

so we have the change of variable formula

$$\int f(x)dm_\Phi(x) = \int f(sx)dm_\Phi(x).$$

for all $f \in \mathcal{K}_{\mathbb{C}}(G)$ and all $s \in G$. Similarly, since

$$\int f(xs^{-1})dm_\Phi(x) = \int (\rho_{s^{-1}}(f))(x)dm_\Phi(x) = \Phi(\rho_{s^{-1}}(f)) = (\rho_s(\Phi))(f),$$

the right-invariance of Φ means that

$$(\rho_s(\Phi))(f) = \Phi(f) = \int f(x)dm_\Phi(x),$$

so we have the change of variable formula

$$\int f(x)dm_\Phi(x) = \int f(xs^{-1})dm_\Phi(x),$$

for all $f \in \mathcal{K}_{\mathbb{C}}(G)$ and all $s \in G$.

The following operation will allow us to convert a left-invariant measure (resp. functional) to a right-invariant measure (resp. functional).

Definition 8.11. Let G be any locally compact group and F be any set. For any function $f\colon G \to F$, define the function $\check{f}\colon G \to F$ by

$$\check{f}(s) = f(s^{-1}) \quad \text{for all } s \in G.$$

For any Borel measure μ on $(G, \mathcal{B})$, define the Borel measure $\check{\mu}$ by

$$\check{\mu}(A) = \mu(A^{-1}) \quad \text{for all } A \in \mathcal{B}.$$

For any Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$, define the Radon functional $\check{\Phi}\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$ by

$$\check{\Phi}(f) = \Phi(\check{f}) \quad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G).$$

Observe that

$$(\lambda_s(\mu))\check{}(A) = (\lambda_s(\mu))(A^{-1}) = \mu(s^{-1}A^{-1}) = \mu((As)^{-1}) = \check{\mu}(As) = (\rho_s(\check{\mu}))(A),$$

so

$$(\lambda_s(\mu))\check{} = \rho_s(\check{\mu}).$$

Similarly,

$$(\rho_s(\mu))\check{}(A) = (\rho_s(\mu))(A^{-1}) = \mu(A^{-1}s) = \mu((s^{-1}A)^{-1}) = \check{\mu}(s^{-1}A) = (\lambda_s(\check{\mu}))(A),$$

so

$$(\rho_s(\mu))\check{} = \lambda_s(\check{\mu}).$$

For any function $f\colon G \to F$, we have

$$(\lambda_s(f))\check{}(x) = (\lambda_s(f))(x^{-1}) = f(s^{-1}x^{-1}) = f((xs)^{-1}) = \check{f}(xs) = (\rho_s(\check{f}))(x),$$

and similarly

$$(\rho_s(f))\check{}(x) = (\rho_s(f))(x^{-1}) = f(x^{-1}s) = f((s^{-1}x)^{-1}) = \check{f}(s^{-1}x) = (\lambda_s(\check{f}))(x).$$

Therefore,

$$(\lambda_s(f))\check{} = \rho_s(\check{f}), \qquad (\rho_s(f))\check{} = \lambda_s(\check{f}) \quad \text{for all } s \in G \text{ and all } f\colon G \to F.$$

Using the above equations, for every Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$, we have

$$(\lambda_s(\Phi))\check{}(f) = (\lambda_s(\Phi))(\check{f}) = \Phi(\lambda_{s^{-1}}(\check{f})) = \Phi((\rho_{s^{-1}}(f))\check{}) = \check{\Phi}(\rho_{s^{-1}}(f)) = (\rho_s(\check{\Phi}))(f).$$

Similarly,

$$(\rho_s(\Phi))\check{}(f) = (\rho_s(\Phi))(\check{f}) = \Phi(\rho_{s^{-1}}(\check{f})) = \Phi((\lambda_{s^{-1}}(f))\check{}) = \check{\Phi}(\lambda_{s^{-1}}(f)) = (\lambda_s(\check{\Phi}))(f).$$

Therefore, we have

$$(\lambda_s(\Phi))\check{} = \rho_s(\check{\Phi}), \qquad (\rho_s(\Phi))\check{} = \lambda_s(\check{\Phi}) \quad \text{for all } s \in G.$$

The definition of the cech operation $(\check{})$ is summarized in the following table.

On functions $\check{f}(s) = f(s^{-1})$
On measures $\check{\mu}(A) = \mu(A^{-1})$
On functionals $\check{\Phi}(f) = \Phi(\check{f})$

Proposition 8.16. *Let G be a locally compact group, and let μ be a σ-regular, locally finite, Borel measure on G (a σ-Radon measure). The following properties hold:*

(1) *We have*

$$(\lambda_s(\mu))\check{} = \rho_s(\check{\mu}), \qquad (\rho_s(\mu))\check{} = \lambda_s(\check{\mu}) \quad \text{for all } s \in G.$$

Consequently, μ is a left-invariant measure iff $\check{\mu}$ is a right-invariant measure. For any Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$, we have

$$(\lambda_s(\Phi))\check{} = \rho_s(\check{\Phi}), \qquad (\rho_s(\Phi))\check{} = \lambda_s(\check{\Phi}) \quad \text{for all } s \in G.$$

Consequently, Φ is left-invariant iff $\check{\Phi}$ is right-invariant.

(2) *If the Haar measure μ is left-invariant then*

$$\int \lambda_{s^{-1}}(f)\, d\mu = \int f d\mu$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$ and all $s \in G$. If the Haar measure μ is right-invariant then

$$\int \rho_{s^{-1}}(f)\, d\mu = \int f d\mu$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$ and all $s \in G$.

(3) *We have*

$$\int \lambda_{s^{-1}}(f)\, d\mu = \int f d\lambda_s(\mu) \quad \text{for all } f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C}) \text{ and all } s \in G.$$

If

$$\int \lambda_{s^{-1}}(f)\, d\mu = \int f d\mu$$

for all $f \in \mathcal{K}_{\mathbb{C}}(G)$ and all $s \in G$, then μ is left-invariant. We have

$$\int \rho_{s^{-1}}(f)\, d\mu = \int f d\rho_s(\mu) \quad \text{for all } f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C}) \text{ and all } s \in G.$$

If

$$\int \rho_{s^{-1}}(f)\, d\mu = \int f d\mu$$

for all $f \in \mathcal{K}_{\mathbb{C}}(G)$ and all $s \in G$, then μ is right-invariant.

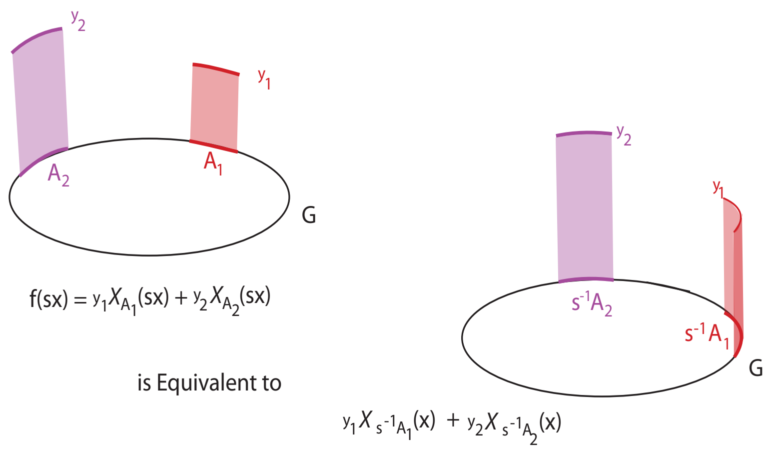


Figure 8.4: Let $G = S^1$. A step function on S^1 is represented by the top arcs of the colored vertical "rectangular" sheets. The step function $f(sx) = \sum_{k=1}^{2} y_k \chi_{A_k}(sx)$ is equivalent to $\lambda_{s^{-1}} f(x) = \sum_{k=1}^{2} y_k \chi_{s^{-1}A_k}(x)$.

Proof. We already proved (1).

(2) Let f be any μ-step function

$$f = \sum_{k=1}^{n} y_k \chi_{A_k},$$

where the A_k are measurable Borel sets of finite measure. For all $s, x \in G$, we see that $(\lambda_{s^{-1}} f)(x) = f(sx) = y_k$ iff $sx \in A_k$ iff $x \in s^{-1}A_k$, which means that

$$\lambda_{s^{-1}} f = \sum_{k=1}^{n} y_k \chi_{s^{-1}A_k},$$

so

$$\int (\lambda_{s^{-1}} f) d\mu = \sum_{k=1}^{n} y_k \mu(s^{-1}A_k) = \sum_{k=1}^{n} y_k (\lambda_s(\mu))(A_k) = \int f d\lambda_s(\mu).$$

See Figure 8.4. If μ is left-invariant, then $\lambda_s(\mu) = \mu$, so

$$\sum_{k=1}^{n} y_k (\lambda_s(\mu))(A_k) = \sum_{k=1}^{n} y_k \mu(A_k) = \int f d\mu,$$

and we deduce that

$$\int (\lambda_{s^{-1}} f) d\mu = \int f d\mu.$$

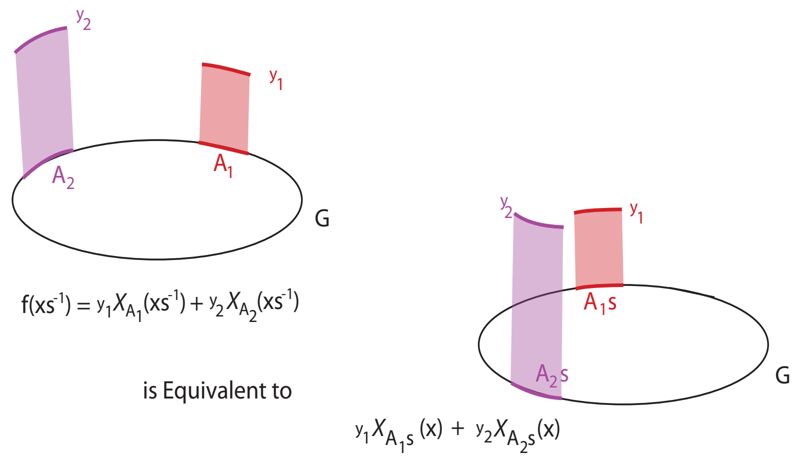


Figure 8.5: Let $G = S^1$. A step function on S^1 is represented by the top arcs of the colored vertical "rectangular" sheets. The step function $f(xs^{-1}) = \sum_{k=1}^{2} y_k \chi_{A_k}(xs^{-1})$ is equivalent to $\rho_{s^{-1}} f(x) = \sum_{k=1}^{2} y_k \chi_{A_k s}(x)$.

Every function $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$ has some approximation sequence (f_n) by μ-step functions that converges to f a.e. and in the L^1-norm. It follows that the sequence $(\lambda_{s^{-1}} f_n)$ converges a.e. to $\lambda_{s^{-1}}(f)$. We check immediately that it is a Cauchy sequence because

$$\int (\lambda_{s^{-1}} f_n) d\mu = \int f_n \, d\mu$$

for all n, and it follows that

$$\int \lambda_{s^{-1}}(f) \, d\mu = \int f d\mu.$$

If f is any μ-step function

$$f = \sum_{k=1}^{n} y_k \chi_{A_k},$$

we have $(\rho_{s^{-1}} f)(x) = f(xs^{-1}) = y_k$ iff $xs^{-1} \in A_k$ iff $x \in A_k s$, which means that

$$\rho_{s^{-1}} f = \sum_{k=1}^{m} y_k \chi_{A_k s},$$

so

$$\int (\rho_{s^{-1}} f) d\mu = \sum_{k=1}^{n} y_k \mu(A_k s) = \sum_{k=1}^{n} y_k (\rho_s(\mu))(A_k) = \int f d\rho_s(\mu).$$

See Figure 8.5.

We finish the argument as in the previous case.

(3) The proof in (2) actually shows that

$$\int \lambda_{s^{-1}}(f)\, d\mu = \int f d\lambda_s(\mu)$$

and

$$\int \rho_{s^{-1}}(f)\, d\mu = \int f d\rho_s(\mu)$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$. If

$$\int \lambda_{s^{-1}}(f)\, d\mu = \int f d\mu$$

then

$$\int f d\lambda_s(\mu) = \int f d\mu$$

for all $f \in \mathcal{K}_\mathbb{C}(G)$ and all $s \in G$, and by the uniqueness of the Borel measure corresponding to the Radon functional $f \mapsto \int f d\mu$ from Theorem 7.8, we see that $\lambda_s(\mu) = \mu$ for all $s \in G$, which means that μ is left-invariant. The right-invariant case is similar. $\square$

The condition

$$\int \lambda_{s^{-1}}(f)\, d\mu = \int f d\mu \quad \text{for all } s \in G$$

is also written as

$$\int f(sx) d\mu(x) = \int f(x) d\mu(x) \quad \text{for all } s \in G.$$

The condition

$$\int \rho_{s^{-1}}(f)\, d\mu = \int f d\mu \quad \text{for all } s \in G$$

is also written as

$$\int f(xs^{-1}) d\mu(x) = \int f(x) d\mu(x) \quad \text{for all } s \in G.$$

Since G is a group and s is any arbitrary element of G, the above condition is also equivalent to

$$\int f(xs) d\mu(x) = \int f(x) d\mu(x) \quad \text{for all } s \in G.$$

8.3 Existence of the Haar Measure

We are now going to sketch the proof that a left-invariant Haar measure exists on any locally compact group. All proofs we are aware of (Weil [105], Halmos [44], Bourbaki [7], Dieudonné [24], Lang [62], Folland [33]) make use of Haar's original clever idea (1933). Except for Halmos who constructs directly a measure (as Haar did), all the other proofs are essentially André Weil's proof (which constructs a Haar functional) from his famous little book [105] first published in 1940.

As we noted just after Proposition 7.4, there is a bijection between the space $\mathrm{M}^+(X)$ of positive linear functionals $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ and the space of positive linear functionals $\Psi\colon \mathcal{K}_{\mathbb{R}}(X) \to \mathbb{R}$, so it is enough to construct a left (or right) real Haar functional on $\mathcal{K}_{\mathbb{R}}(G)$.

Theorem 8.17. *(Haar) Every locally compact group G possesses a left-invariant Haar measure.*

Proof sketch. Folland [33] (Chapter 2, Section 2.2) is kind enough to provide the intuition behind the construction. In this method a measure is not constructed directly. Instead, a left Haar functional is constructed. Then Theorem 7.8 is used to obtain a left-invariant Borel measure which is a left Haar measure.

Suppose we have positive function $\varphi \in \mathcal{K}_{\mathbb{R}}(G)$ bounded by 1, equal to 1 on a small open set U, and whose support is a compact subset slightly larger than U. If $f \in \mathcal{K}_{\mathbb{R}}(G)$ is any other function slowly varying so that it is essentially constant on the left translates of U, then f can be approximated by a linear combination $f \approx \sum c_j \lambda_{s_j}(\varphi)$. If μ were a left Haar measure, then we would have

$$\int f d\mu \approx \left(\sum_j c_j\right) \int \varphi \, d\mu.$$

See Figure 8.6.

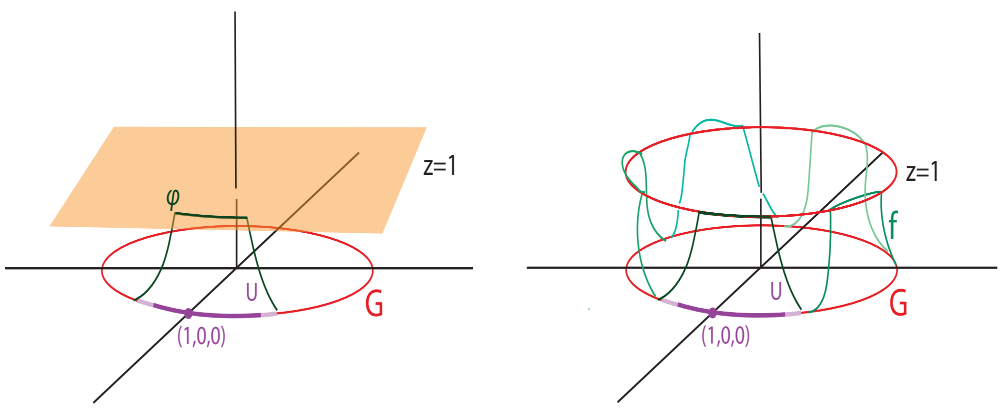


Figure 8.6: Let G be the unit circle $\mathbb{T}$ in the xy-plane. The left figure shows the "bump" function φ, while the right figure illustrates f as five translates of φ, namely $f \approx \sum_{j=1}^{5} c_j \lambda_{s_j}(\varphi)$.

This approximation gets better and better as the support of φ shrinks to a point, and if we introduce a normalization to cancel out the factor $\int \varphi \, d\mu$, then we obtain $\int f d\mu$ as the limit of the sums $\sum_j c_j$. If $G = \mathbb{R}$, and if φ is the characteristic function of a small interval, this is reminiscent of the approximation of $\int f d\mu$ by Riemann sums. The issue is to make this idea precise and formal.

The first step is to pick some positive nonzero function $\varphi \in \mathcal{K}_{\mathbb{R}}(G)$. This function remains fixed until Proposition 8.18. Then we claim that for every function $f \in \mathcal{K}_{\mathbb{R}}(G)$, there exists a finite set $\{s_1, \ldots, s_n\}$ of elements of G and a finite sequence $(c_1, \ldots, c_n)$ of reals $c_j \in \mathbb{R}$, such that

$$f \leq \sum_{j=1}^{n} c_j \lambda_{s_j}(\varphi).$$

This is because f has compact support, so its support can be covered by finitely many translates of the open subset U given by

$$U = \left\{ s \in G \mid \varphi(s) > \frac{1}{2} \|\varphi\|_\infty \right\},$$

and if we pick $c_j = \|f\|_\infty / a$ where $a = (n/2) \|\varphi\|_\infty$, then on each translate $s_j U$ we have

$$\lambda_{s_j}(\varphi)(x) > \frac{1}{2} \|\varphi\|_\infty, \quad x \in s_j U,$$

which implies

$$\begin{aligned} \sum_{j=1}^{n} c_j \lambda_{s_j}(\varphi) &= \sum_{j=1}^{n} \frac{2 \|f\|_\infty}{n \|\varphi\|_\infty} \lambda_{s_j}(\varphi) \\ &> \|f\|_\infty, \end{aligned}$$

so

$$f \leq \sum_{j=1}^{n} c_j \lambda_{s_j}(\varphi).$$

Define $(f : \varphi)$ as the greatest lower bound of the sums $\sum_{j=1}^{n} c_j$, over all sets $\{s_1, \ldots, s_n\}$ of elements of G and all finite sequences $(c_1, \ldots, c_n)$ of reals $c_j \in \mathbb{R}$ such that

$$f \leq \sum_{j=1}^{n} c_j \lambda_{s_j}(\varphi).$$

Then it is not hard to show that the quantity $(f : \varphi)$ has the following properties:

$$\begin{aligned}
(f : \varphi) &= (\lambda_s(f) : \varphi) && \text{for all } s \in G \\
(f_1 + f_2 : \varphi) &\leq (f_1 : \varphi) + (f_2 : \varphi) && \text{for all } f_1, f_2 \in \mathcal{K}_{\mathbb{R}}(G) \\
(cf : \varphi) &= c(f : \varphi) && \text{for all } c \geq 0 \\
(f_1 : \varphi) &\leq (f_2 : \varphi) && \text{whenever } f_1 \leq f_2 \\
(f : \varphi) &\geq \|f\|_\infty / \|\varphi\|_\infty && \\
(f : \varphi) &\leq (f : \psi)(\psi : \varphi) && \text{for all positive } \psi \in \mathcal{K}_{\mathbb{R}}(G) \\
0 < \frac{1}{(f_0 : f)} \leq \frac{(f : \varphi)}{(f_0 : \varphi)} &\leq (f : f_0) && \text{for all positive } f, f_0 \in \mathcal{K}_{\mathbb{R}}(G).
\end{aligned}$$

We now make a normalization by fixing some positive nonzero $f_0 \in \mathcal{K}_{\mathbb{R}}(G)$, and defining

$$I_\varphi(f) = \frac{(f : \varphi)}{(f_0 : \varphi)}$$

for every *positive* function $f \in \mathcal{K}_{\mathbb{R}}(G)$. The above properties show that I_φ is a functional which is left-invariant, subadditive, homogeneous of degree 1, and monotone. It also satisfies the following property:

$$\frac{1}{(f_0 : f)} \leq I_\varphi(f) \leq (f : f_0). \tag{$*$}$$

If I_φ were additive rather than subadditive, it would be the restriction to the positive functions in $\mathcal{K}_{\mathbb{R}}(G)$ of a positive linear functional on $\mathcal{K}_{\mathbb{R}}(G)$, and we would be done. To make a linear functional, we need to shrink to the domain of φ, and this is the part of the argument which is the most subtle. Let $\mathcal{K}^+_{\mathbb{R}}(G)$ denote the set of positive functions in $\mathcal{K}_{\mathbb{R}}(G)$.

The following technical proposition whose proof is given in Folland [33] (Chapter 2, Lemma 2.18) is needed.

Proposition 8.18. *If f_1 and f_2 are any two positive functions in $\mathcal{K}^+_{\mathbb{R}}(G)$ and if $\epsilon > 0$, then there is an open subset V containing 1 such that $I_\varphi(f_1) + I_\varphi(f_2) \leq I_\varphi(f_1 + f_2) + \epsilon$, whenever* $\mathrm{supp}(\varphi) \subseteq V$.

To shrink the domain of φ we use a compactness argument. For every positive function $f \in \mathcal{K}^+_{\mathbb{R}}(G)$, let X_f be the interval

$$X_f = \left[\frac{1}{(f_0 : f)}, (f : f_0)\right].$$

Let $X = \prod_f X_f$. More precisely, X is the set of all functions from $\mathcal{K}^+_{\mathbb{R}}(G)$ to $(0, +\infty)$ mapping f into X_f. We put the topology of Definition 2.2 on X. Since each X_f is compact, by Tychonoff's theorem X is also compact. By $(*)$, we have $I_\varphi(f) \in X$ for all positive nonzero $\varphi \in \mathcal{K}_{\mathbb{R}}(G)$. For every compact neighborhood V containing 1, let $K(V)$ be the closure in X of the subset $\{I_\varphi \mid \mathrm{supp}(\varphi) \subseteq V\}$. The family of subsets $K(V)$ has the finite intersection property since $K(\bigcap_{j=1}^n V_j) \subseteq \bigcap_{j=1}^n K(V_j)$. Since X is compact, there is some $I \in X$ which lies in every $K(V)$. This means that every neighborhood of I in X contains some I_φ with $\mathrm{supp}(\varphi)$ arbitrarily small. In other words, for any open subset V containing 1, any $\epsilon > 0$, and for any positive functions $f_1, \ldots, f_n \in \mathcal{K}_{\mathbb{R}}(G)$, there exist some positive nonzero $\varphi \in \mathcal{K}_{\mathbb{R}}(G)$ with $\mathrm{supp}(\varphi) \subseteq V$ such that $|I(f_j) - I_\varphi(f_j)| < \epsilon$ for all j. By the properties of I_φ listed above and by Proposition 8.18, we conclude that I commutes with left translation, addition, and multiplication by positive scalars.

We can extend I to arbitrary functions $f \in \mathcal{K}_{\mathbb{R}}(G)$ as follows. We can write $f = f_1 - f_2$ with f_1, f_2 positive functions in $\mathcal{K}_{\mathbb{R}}(G)$, and we let $I(f) = I(f_1) - I(f_2)$. If we also have $f = f_1' - f_2'$ with f_1', f_2' positive functions in $\mathcal{K}_{\mathbb{R}}(G)$, then

$$f_1 + f_2' = f_2 + f_1',$$

so by linearity of I on positive functions we get

$$I(f_1) + (f_2') = I(f_2) + I(f_1'),$$

thus

$$I(f_1) - I(f_2) = I(f_1') - I(f_2'),$$

which means that $I(f)$ is well defined. The functional I is a left Haar functional, and we are done. By Proposition 8.16(3), since the Haar functional I is left-invariant, the corresponding σ-Radon measure is also left-invariant. $\square$

Remark: The proof in Bourbaki [7] uses an argument involving an ultrafilter instead of Tychonoff's theorem, but otherwise it is identical. Dieudonné [24] assumes that the locally compact group G is separable and metrizable. This allows him to avoid using Tychonoff's theorem, but does not make the proof simpler.

Let μ be the left Haar measure associated with the left Haar functional I given by Theorem 8.17. Here is an immediate consequence of Theorem 8.17.

Proposition 8.19. *If μ is a left Haar measure on G, then for every nonempty open subset U, we have $\mu(U) > 0$. For every positive nonzero function $f \in \mathcal{K}_{\mathbb{R}}(G)$, we have $\int f d\mu > 0$.*

Proof. Assume U is a nonempty open set with $\mu(U) = 0$. Then since μ is left-invariant $\mu(gU) = 0$ for all $g \in G$, and since any compact subset K can be covered by finitely many translates of U, we have $\mu(K) = 0$. Since μ is a σ-regular Borel measure, it is σ-inner regular, that is,

$$\mu(G) = \sup\{\mu(K) \mid K \subseteq G,\ K \text{ compact}\}$$

so $\mu(G) = 0$, contradicting the fact that μ is not the zero measure because it arises from a non-zero left Haar functional by Radon–Riesz I.

For any positive nonzero function $f \in \mathcal{K}_{\mathbb{R}}(G)$, let $U = \{g \in G \mid f(g) > \frac{1}{2}\left\|f\right\|_\infty\}$. Then $\int f d\mu > \frac{1}{2}\left\|f\right\|_\infty \mu(U) > 0$. $\square$

Remark: If G is a Lie group, there are much simpler methods for obtaining a left Haar measure on G. Suppose G has dimension n. Pick an n-differential form ω_0 on $\mathfrak{g}$, and transport it on all tangent spaces by left translation, obtaining a left-invariant volume form ω. Then $f \mapsto \int f\omega$ is a left-invariant Haar functional that induces a left Haar measure.

We now turn to the uniqueness of the Haar measure.

8.4 Uniqueness of the Haar Measure

Any two left Haar measures on a locally compact group are proportional up to a positive factor. All the proofs we are aware of use tricks involving a double integration and Fubini's theorem. These proofs are attributed to von Neumann. In our opinion, the proof using the least devious trick is that of Dieudonné [24] (Chapter XIV, Section 1), also used in Bourbaki [7] in a slightly more concise form. Since Dieudonné uses a theory of integration based on Radon functionals rather than on measure theory, some minor adaptations need to be made; specifically, Proposition 7.12 is needed instead of Proposition 13.15.3 in Dieudonné [24]. The first step is the following crucial result.

Proposition 8.20. *Given a left Haar functional Φ and a right Haar functional Ψ on a locally compact group G, if ν is the corresponding right Haar measure, for any function $f \in \mathcal{K}_{\mathbb{R}}(G)$, if $\Phi(f) \neq 0$, then the function D_f given by*

$$D_f(s) = \Phi(f)^{-1} \int f(t^{-1}s)\, d\nu(t)$$

for all $s \in G$ is continuous.

Proof. It suffices to show that the function

$$s \mapsto \int f(t^{-1}s)\, d\nu(t)$$

is continuous. Let K be the compact subset of G which is the support of f. Pick any $s_0 \in G$, and let V_0 by any compact neighborhood of s_0. For every $\epsilon > 0$, we have to find an open subset V containing s_0 such that $V \subseteq V_0$ and

$$\left| \int (f(t^{-1}s) - f(t^{-1}s_0)) d\nu(t) \right| < \epsilon$$

for all $s \in V$. In order to have $t^{-1}s, t^{-1}s_0 \in K$, since $s_0, s \in V_0$, it suffices that $t \in V_0K^{-1}$. If we let $L = V_0K^{-1}$, then for $s_0 \in V_0$ and $s \in V \subseteq V_0$,

$$\int (f(t^{-1}s) - f(t^{-1}s_0)) d\nu(t) = \int_L (f(t^{-1}s) - f(t^{-1}s_0)) d\nu(t).$$

By Proposition 8.12, the function f is right uniformly continuous, so there is some open subset W containing 1 such that

$$|f(t^{-1}s) - f(t^{-1}s_0)| < \frac{\epsilon}{\nu(L)}$$

for all $(t^{-1}s_0)^{-1}t^{-1}s \in W$, that is, $s_0^{-1}s \in W$, namely $s \in Ws_0$, and all $t \in G$. If we take $V = V_0 \cap Ws_0$ (see Figure 8.7), then

$$\left| \int_L (f(t^{-1}s) - f(t^{-1}s_0)) d\nu(t) \right| < \int_L |f(t^{-1}s) - f(t^{-1}s_0)| d\nu(t) < \epsilon,$$

as desired. $\square$

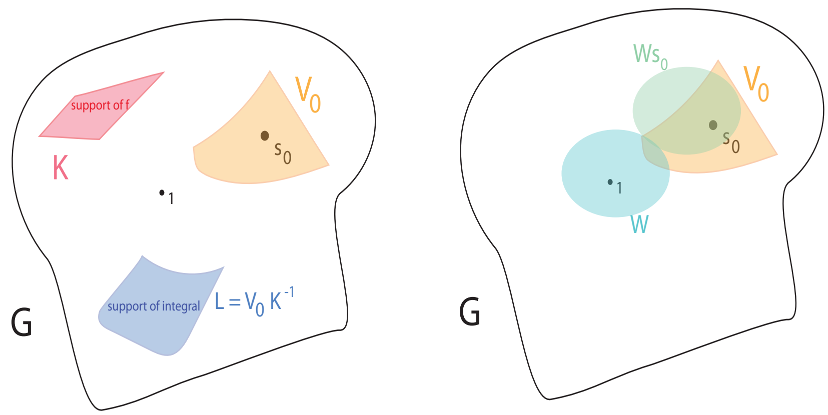


Figure 8.7: A schematic representation of the sets used in the proof of Proposition 8.20. Observe that $V = V_0 \cap W s_0$ is the intersection of the light green ellipse and peach triangle.

We are now ready to prove our uniqueness result.

Theorem 8.21. *(Haar) If μ and ν are any two left-invariant Haar measures on a locally compact group G, then there is some $c > 0$ such that $\mu = c\nu$.*

Proof. Since a Haar functional Ψ is right-invariant iff $\check{\Psi}$ is left-invariant, it suffices to prove that if Φ is a left-invariant Haar functional and if Ψ is a right-invariant Haar functional, then there is some $c > 0$ such that $\Phi = c\check{\Psi}$. Let μ be the left Haar measure associated with Φ and let ν be right Haar measure associated with Ψ.

Let $f \in \mathcal{K}_{\mathbb{R}}(G)$ be any function such that $\Phi(f) \neq 0$ and let $g \in \mathcal{K}_{\mathbb{R}}(G)$ be any other function. The function from $G \times G$ to $\mathbb{R}$ given by $(s, t) \mapsto f(s)g(ts)$ is continuous and has compact support. Recall that D_f is given by

$$D_f(s) = \Phi(f)^{-1} \int f(t^{-1}s)\, d\nu(t).$$

For $s = 1$, we have

$$D_f(1) = \Phi(f)^{-1}\check{\Psi}(f).$$

Therefore, if we can show that D_f is independent of f, we are done. We evaluate $\Phi(f)\Psi(g)$

using Fubini's theorem.

$$
\begin{aligned}
\Phi(f)\Psi(g) &= \left(\int f(s)d\mu(s)\right)\left(\int g(t)\,d\nu(t)\right) \\
&= \int f(s)\left(\int g(t)\,d\nu(t)\right)d\mu(s) \\
&= \int \left(\int f(s)g(t)\,d\nu(t)\right)d\mu(s) \\
&= \int \left(\int f(s)g(ts)\,d\nu(t)\right)d\mu(s) && \text{by right-invariance of } \nu \\
&= \int \left(\int f(s)g(ts)\,d\mu(s)\right)d\nu(t) && \text{by Fubini} \\
&= \int \left(\int f(t^{-1}s)g(s)\,d\mu(s)\right)d\nu(t) && \text{by left-invariance of } \mu \\
&= \int \left(\int f(t^{-1}s)g(s)\,d\nu(t)\right)d\mu(s) && \text{by Fubini} \\
&= \int g(s)\left(\int f(t^{-1}s)\,d\nu(t)\right)d\mu(s) \\
&= \int g(s)\Phi(f)D_f(s)\,d\mu(s) && \text{by definition of } D_f \\
&= \Phi(f)\Phi(D_f\cdot g),
\end{aligned}
$$

where $D_f \cdot g$ is the function given by $(D_f \cdot g)(s) = D_f(s)g(s)$ for all $s \in G$. Since $\Phi(f) \neq 0$, we deduce that

$$\Psi(g) = \Phi(D_f \cdot g).$$

The above equation shows that D_f is independent of f because if f' is another function $f' \in \mathcal{K}_{\mathbb{R}}(G)$ such that $\Phi(f') \neq 0$, then

$$\Phi(D_f \cdot g) = \int D_f(s)g(s)\,d\mu(s) = \int D_{f'}(s)g(s)\,d\mu(s) = \Phi(D_{f'} \cdot g) \quad \text{for all } g \in \mathcal{K}_{\mathbb{R}}(G).$$

By Proposition 7.12, we deduce that D_f and $D_{f'}$ are equal a.e. (The version of Proposition 7.12 is stated for complex-valued functions, but it also holds for real-valued functions). However, D_f and $D_{f'}$ are continuous and the subset N where they differ is open and a null set, thus empty by Proposition 8.19. Therefore $D_f = D_{f'} = D$, and by definition of D we have

$$\Phi(f) = D(1)^{-1}\check{\Psi}(f) \quad \text{for all } f \in \mathcal{K}_{\mathbb{R}}(G) \text{ with } \Phi(f) \neq 0.$$

Now Φ and $\check{\Psi}$ are two linear functionals that agree in the complement of the hyperplane H in $\mathcal{K}_{\mathbb{R}}(G)$ of equation $\Phi(f) = 0$, so they agree everywhere. To see this, pick a basis of $\mathcal{K}_{\mathbb{R}}(G)$ consisting of a basis $(h_j)_{j\in J}$ of H and a function v not in H. We claim that the family

consisting of $(h_j+v)_{j\in J}$ and v is a basis of $\mathcal{K}_{\mathbb{R}}(G)$. This family obviously spans $\mathcal{K}_{\mathbb{R}}(G)$ (since every h_j is obtained as h_j+v-v), and it is linearly independent because if we have a finite linear combination

$$\sum_{i\in I}\lambda_i(h_i+v)+\mu v=0.$$

for any finite subset I of J, then

$$\sum_{i\in I}\lambda_i h_i+\left(\mu+\sum_{i\in I}\lambda_i\right)v=0,$$

and by linear independence, $\lambda_i=0$ for all $i\in I$ and $\mu+\sum_{i\in I}\lambda_i=0$, which implies $\mu=0$, and since this holds for any finite subset I of J, the family consisting of $(h_j+v)_{j\in J}$ and v is linearly independent. Since Φ and Ψ are linear and they agree on a basis, they must be identical.

Since $\Psi\neq 0$, we must have $D(1)\neq 0$, thus, $\Phi=D(1)^{-1}\check{\Psi}$. Since Φ and Ψ are positive functionals, we must have $D(1)>0$.

As we observed earlier, since a locally compact group is the disjoint union of σ-compact cosets, it is legitimate to use Fubini's theorem. $\square$

8.5 Examples of Haar Measures

Here are some examples of Haar measures on various locally compact groups. In most cases, a Haar measure μ on a locally compact group G is defined indirectly by a Haar functional $f\mapsto\int f d\mu$, for all $f\in\mathcal{K}_{\mathbb{C}}(G)$. This Haar functional is denoted by $d\mu$.

Example 8.1. The additive group $\mathbb{R}$ is a locally compact group, and the Lebesgue measure μ_L is a left (and right) Haar measure on it. For a proof see Lang [62] (Chapter VI, Theorem 9.7). An alternative is to use Proposition 8.16(3). By the simple change of variable $x=t+s$, for any function $f\in\mathcal{K}_{\mathbb{C}}(\mathbb{R})$, we have

$$\int_{-\infty}^{\infty}f(t)dt=\int_{-\infty}^{\infty}f(t+s)dt.$$

Example 8.2. The additive group $\mathbb{R}^n$ is a locally compact group, and the product Lebesgue measure μ_L on it (see Section 5.13) is a left (and right) Haar measure on it. This will be shown as an application of Proposition 8.37.

Example 8.3. The multiplicative group $\mathbb{R}_+^*$ is a locally compact group. We claim that $d\mu=dx/x$ is a left Haar measure, where dx is the restriction of the Lebesgue measure to $\mathbb{R}_+^*$. Indeed, using the change of variable $t\mapsto st$, for any function $f\in\mathcal{K}_{\mathbb{C}}(\mathbb{R}_+^*)$, we have

$$\int_0^{\infty}\frac{f(t)dt}{t}=\int_0^{\infty}\frac{f(st)sd(t)}{st}=\int_0^{\infty}\frac{f(st)d(t)}{t},$$

establishing left-invariance. One might wonder what is the measure $\mu([a,b])$ of a closed interval, with $0 < a < b$. We have

$$\mu([a,b]) = \int \chi_{[a,b]} d\mu = \int_{[a,b]} d\mu = \int_a^b \frac{dt}{t} = [\log t]_a^b = \log \frac{b}{a}.$$

For any $s > 0$, we have $s \cdot [a,b] = [sa, sb]$, and

$$\mu(s \cdot [a,b]) = \mu([sa, sb]) = \log \frac{sb}{sa} = \log \frac{b}{a}.$$

This measure is indeed left invariant.

Example 8.4. Let $\mathbb{T} = \mathbf{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\}$, the circle group, that is, the group of complex numbers of unit length. Let $\sigma\colon \mathbb{T} \to \mathbb{R}$ be the injection given by

$$\sigma(e^{i\theta}) = \theta, \quad -\pi \leq \theta < \pi.$$

Define the measure ν_1 on $\mathbb{T}$ by

$$\nu_1(A) = \mu_L(\sigma(A)),$$

on the σ-algebra $\sigma^{-1}(\mathcal{B}(\mathbb{R}))$ defined in Proposition 5.2(2) (where μ_L is the Lebesgue measure on $\mathbb{R}$). For any $f \in \mathcal{L}_{\nu_1}(\mathbb{T})$, we have

$$\int_{\mathbb{T}} f d\nu_1 = \int_{-\pi}^{\pi} f(e^{i\theta})\, d\mu_L(\theta),$$

also written as $\int_{-\pi}^{\pi} f(e^{i\theta})\, d\theta$. Observe that $\int_{\mathbb{T}} d\nu_1 = 2\pi$. It is easy to check that ν_1 is left-invariant. Indeed, for $\theta_0 \in [-\pi, \pi)$, if we let $\varphi = \theta + \theta_0$, we have

$$\int_{-\pi}^{\pi} f(e^{i(\theta+\theta_0)})\, d\theta = \int_{-\pi+\theta_0}^{\pi+\theta_0} f(e^{i\varphi})\, d\varphi = \int_{-\pi+\theta_0}^{\pi} f(e^{i\varphi})\, d\varphi + \int_{\pi}^{\pi+\theta_0} f(e^{i\varphi})\, d\varphi.$$

Using the change of variable $\varphi = u + 2\pi$ in the second integral, we get

$$\int_{\pi}^{\pi+\theta_0} f(e^{i\varphi})\, d\varphi = \int_{-\pi}^{-\pi+\theta_0} f(e^{iu})\, du,$$

and so

$$\int_{-\pi}^{\pi} f(e^{i(\theta+\theta_0)})\, d\theta = \int_{-\pi+\theta_0}^{\pi} f(e^{i\varphi})\, d\varphi + \int_{-\pi}^{-\pi+\theta_0} f(e^{iu})\, du = \int_{-\pi}^{\pi} f(e^{i\theta})\, d\theta.$$

Example 8.5. Let $G = \mathbf{GL}(n, \mathbb{R})$, the group of invertible $n \times n$ real matrices. It can be shown that a left (and right) Haar measure on $\mathbf{GL}(n, \mathbb{R})$ is given by

$$d\mu = \frac{dA}{|\det(A)|^n} = |\det(A)|^{-n} \bigotimes_{i,j} da_{ij}$$

with $A = (a_{ij})$, where da_{ij} is the Lebesgue measure on $\mathbb{R}$, and dA is the Lebesgue measure on $\mathbb{R}^{n^2}$.

In the next section, we explore the relationship between a left Haar measure μ and the left Haar measure $\rho_s(\mu)$.

8.6 The Modular Function

Let μ be a left Haar measure on the locally compact group G. For all $s, t \in G$, since λ_t and ρ_s commute (on functions, measures, and Radon functionals), since μ is a left Haar measure, $\lambda_t(\mu) = \mu$, so we have

$$\lambda_t(\rho_s(\mu)) = \rho_s(\lambda_t(\mu)) = \rho_s(\mu),$$

which means that $\rho_s(\mu)$ is also a left Haar measure. By the uniqueness result of Theorem 8.21, there is a constant $a > 0$ such that

$$\rho_s(\mu) = a\mu.$$

If ν is another left Haar measure, again by theorem Theorem 8.21, we have $\nu = c\mu$ for some $c > 0$, but $\rho_s(\nu)(A) = \nu(As) = c\mu(As) = c\rho_s(\mu)(A)$ for all $A \in \mathcal{B}$, that is, $\rho_s(\nu) = c\rho_s(\mu)$, so

$$\rho_s(\nu) = c\rho_s(\mu) = ca\mu = ac\mu = a\nu.$$

Therefore, the number a such that $\rho_s(\mu) = a\mu$ is independent of μ. It is customary to denote this number by $\Delta(s)$.

Definition 8.12. Let G be a locally compact group. For every $s \in G$, there is a unique positive number $\Delta(s)$ such that

$$\rho_s(\mu) = \Delta(s)\mu \tag{$*$}$$

for all left Haar measures μ. The function $\Delta\colon G \to \mathbb{R}_+^*$ (given by $\Delta(s)$ for every $s \in G$) is called the *modular function* of G (if necessary, we denote it by Δ_G to avoid ambiguities).

Observe that $(*)$ can be expressed as

$$\mu(As) = \Delta(s)\mu(A) \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G.$$

Proposition 8.22. *Let G be a locally compact group and let μ be a left Haar measure on G. For any $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$, we have*

$$\int \rho_s(f)\, d\mu = \Delta(s^{-1}) \int f d\mu.$$

The function $\Delta\colon G \to \mathbb{R}_+^*$ is a continuous homomorphism.

Proof sketch. Let $f = \sum_{i=1}^n y_i \chi_{A_i}$ be a μ-step function. For all $x \in G$, we have $\rho_s(f)(x) = f(xs) = y_i$ iff $xs \in A_i$ iff $x \in A_i s^{-1}$, which shows that

$$\rho_s(f) = \sum_{i=1}^n y_i \chi_{A_i s^{-1}};$$

see Figure 8.8.

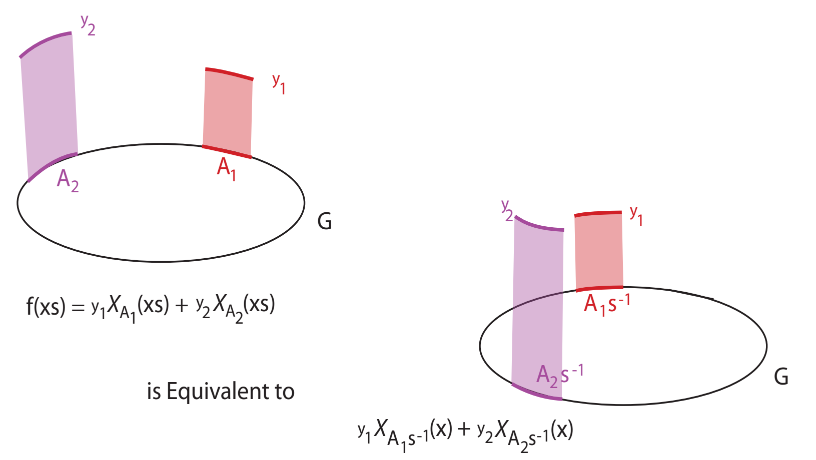


Figure 8.8: Let $G = S^1$. A step function on S^1 is represented by the top arcs of the colored vertical "rectangular" sheets. The step function $f(xs) = \sum_{k=1}^2 y_k\chi_{A_k}(xs)$ is equivalent to $\rho_s f(x) = \sum_{k=1}^2 y_k\chi_{A_k s^{-1}}(x)$.

Consequently,

$$\int \rho_s(f)\,d\mu = \sum_{i=1}^n y_i\mu(A_i s^{-1}) = \Delta(s^{-1})\sum_{i=1}^n y_i\mu(A_i) = \Delta(s^{-1})\int f d\mu,$$

by $(*)$. As in the proof of Proposition 8.16, every function $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$ has some approximation sequence (f_n) by μ-step functions that converges to f a.e. and in the L^1-norm, which allows us to conclude that

$$\int \rho_s(f)\,d\mu = \Delta(s^{-1})\int f d\mu$$

for every $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$.

By $(*)$ we have

$$\Delta(st)\mu(A) = \mu(Ast) = \Delta(t)\mu(As) = \Delta(t)\Delta(s)\mu(A)$$

for all $A \in \mathcal{B}$, and since $\mu(A) > 0$ if A is open and nonempty (and $\mathbb{R}_+$ is commutative under multiplication!), we deduce that

$$\Delta(st) = \Delta(s)\Delta(t).$$

Thus Δ is a homomorphism from G to the multiplicative group $\mathbb{R}^*_+$. Proposition 8.12 implies that the map $s \mapsto \rho_s(f)$ is uniformly continuous, and so it can be shown that the map $s \mapsto \int \rho_s(f)\,d\mu$ is continuous, and since

$$\int \rho_s(f)\,d\mu = \Delta(s^{-1})\int f d\mu,$$

we deduce that Δ is continuous. $\square$

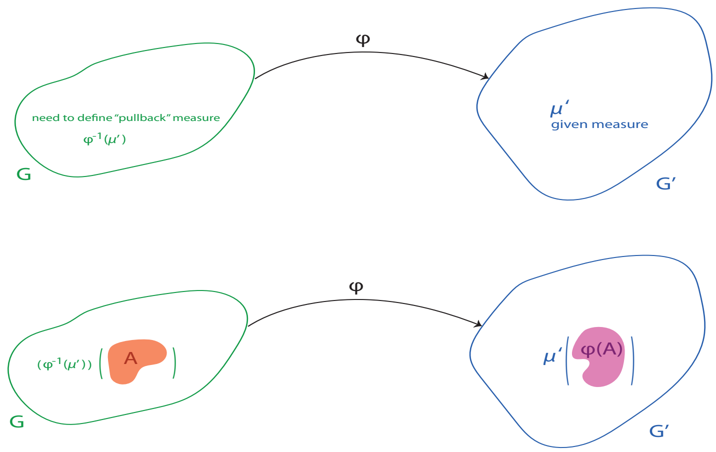


Figure 8.9: A schematic representation of Definition 8.14.

The equation

$$\int \rho_s(f)\, d\mu = \Delta(s^{-1}) \int f d\mu$$

is also written as

$$\int f(xs) d\mu(x) = \Delta(s^{-1}) \int f(x) d\mu(x),$$

or equivalently as

$$\int f(xs^{-1}) d\mu(x) = \Delta(s) \int f(x) d\mu(x).$$

Since $\Delta\colon G \to \mathbb{R}_+^*$ is a group homomorphism, we have $\Delta(s^{-1}) = (\Delta(s))^{-1}$.

Definition 8.13. We write Δ^{-1} for the function given by $\Delta^{-1}(s) = \Delta(s^{-1})$ for all $s \in G$.

Let G and G' be two locally compact groups. An *isomorphism* is map $\varphi\colon G \to G'$ which is a group isomorphism and a homeomorphism. Then it is easy to check that φ maps Borel sets of G to Borel sets of G'.

Definition 8.14. Let G and G' be two locally compact groups, and $\varphi\colon G \to G'$ be an isomorphism. Given a measure μ' on G', we define the map $\varphi^{-1}(\mu')$ with domain $\mathcal{B}(G)$ by

$$(\varphi^{-1}(\mu'))(A) = \mu'(\varphi(A)) \quad \text{for all } A \in \mathcal{B}(G);$$

see Figure 8.9.

Proposition 8.23. *Let G and G' be two locally compact groups, and $\varphi\colon G \to G'$ be an isomorphism. For any left Haar measure μ' on G', the map $\varphi^{-1}(\mu')$ is a left Haar measure on G. We have*

$$\Delta_G = \Delta_{G'} \circ \varphi.$$

Proof sketch. The fact that $\varphi^{-1}(\mu')$ is a measure follows from the fact that φ maps Borel sets to Borel sets and is a bijection, so it preserves union and disjointness. The details are left as an exercise. Since μ' is a left Haar measure and φ is a homomorphism,

$$\varphi^{-1}(\mu')(sA) = \mu'(\varphi(sA)) = \mu'(\varphi(s)\varphi(A)) = \mu'(\varphi(A)) = \varphi^{-1}(\mu')(A),$$

so $\varphi^{-1}(\mu')$ is a left Haar measure.

We have

$$\begin{aligned}(\varphi^{-1}(\mu'))(As) &= \Delta_G(s)(\varphi^{-1}(\mu'))(A) \\ &= \Delta_G(s)\mu'(\varphi(A)),\end{aligned}$$

and

$$\begin{aligned}(\varphi^{-1}(\mu'))(As) &= \mu'(\varphi(As)) \\ &= \mu'(\varphi(A)\varphi(s)) \\ &= \Delta_{G'}(\varphi(s))\mu'(\varphi(A)),\end{aligned}$$

which implies

$$\Delta_{G'}(\varphi(s)) = \Delta_G(s) \quad \text{for all } s \in G$$

since we can pick a nonempty open subset A of G, and $\varphi(A)$ is a nonempty open subset of G'. $\square$

Corollary 8.24. *If $G' = G$, that is, $\varphi\colon G \to G$ is an automorphism, then $\Delta \circ \varphi = \Delta$.*

Definition 8.15. Let G be a locally compact group. We say that G is *unimodular* if $\Delta(s) = 1$ for all $s \in G$, equivalently, if and only if a left Haar measure is also a right Haar measure.

Luckily, many familiar groups are unimodular (but unfortunately, not the affine groups of rigid motions). Obviously, abelian locally compact groups are unimodular.

Given a group G, recall that its *commutator subgroup* $[G, G]$ is the subgroup generated by all elements $[s, t] = sts^{-1}t^{-1}$. The group $[G, G]$ is a normal subgroup of G.

Proposition 8.25. *Let G be a locally compact group.*

(1) *If there is a compact neighborhood V of 1 such that $s^{-1}Vs = V$ for all $s \in G$, then G is unimodular. Consequently, if G is compact, discrete, or commutative, then G is unimodular.*

(2) *If K is any compact subgroup of G, then $\Delta \mid K \equiv 1$.*

(3) *If $G/[G, G]$ is compact, then G is unimodular. As a consequence, every connected semisimple Lie group is unimodular. Recall that semisimple Lie group is a Lie group G such that the Killing form on its Lie algebra $\mathfrak{g}$ is nondegenerate.*

Proof. (1) Let μ be any left Haar measure. Since μ is left-invariant

$$\mu(V) = \mu(s^{-1}Vs) = \mu(Vs) = \Delta(s)\mu(V),$$

but $\mu(V) > 0$ because V contains a nonempty open subset, so $\Delta(s) = 1$ for all $s \in G$. The corollaries are left as an easy exercise.

(2) Since Δ is continuous, $\Delta(K)$ is a compact subgroup of $\mathbb{R}_+^*$, which implies $\Delta(K) = \{1\}$.

(3) Since $\mathbb{R}_+^*$ is abelian, we have

$$\Delta([s,t]) = \Delta(sts^{-1}t^{-1}) = \Delta(s)\Delta(t)\Delta(s)^{-1}\Delta(t)^{-1} = \Delta(s)\Delta(s)^{-1}\Delta(t)\Delta(t)^{-1} = 1,$$

so Δ vanishes on $[G,G]$. It follows that Δ factors through $G/[G,G]$ as $\Delta = \pi \circ \theta$ where $\theta\colon G/[G,G] \to \mathbb{R}_+^*$ is a continuous homomorphism. Since $G/[G,G]$ is compact, we have $\theta(G/[G,G]) = \{1\}$, so $\Delta(G) = \{1\}$.

If G is a connected semisimple Lie group, it is known that $G = [G,G]$, so G is unimodular. $\square$

In order to discuss the behavior of the operator $\mu \mapsto \check{\mu}$ we need the following proposition.

Proposition 8.26. *Let μ and ν be two Radon measures on a locally compact topological space X. If there is a continuous function $g\colon X \to \mathbb{R}_+^*$ such that*

$$\int f d\nu = \int fg\, d\mu \quad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(X),$$

and if $\widetilde{\nu}$ is the Radon measure given by

$$\widetilde{\nu}(E) = \int_E g\, d\mu \quad \text{for all } E \in \mathcal{B}(X),$$

then $\nu = \widetilde{\nu}$.

A proof of Proposition 8.26 is given in Folland [33] (Chapter 2, Proposition 2.23).

We propose to denote the Radon measure $\widetilde{\nu}$ by $g \cdot \mu$, by analogy with the definition of the Radon functional $g \cdot \Phi$ in Example 7.1(3). The notation $g\, d\mu$ is also used.

The following proposition shows the behavior of the operator $\mu \mapsto \check{\mu}$.

Proposition 8.27. *Let G be a locally compact group. For every left Haar measure μ, the measure $\check{\mu}$ is a right Haar measure, and we have*

$$\int \check{f}\, d\mu = \int f d\check{\mu}, \quad \check{\mu} = \Delta^{-1} \cdot \mu,$$

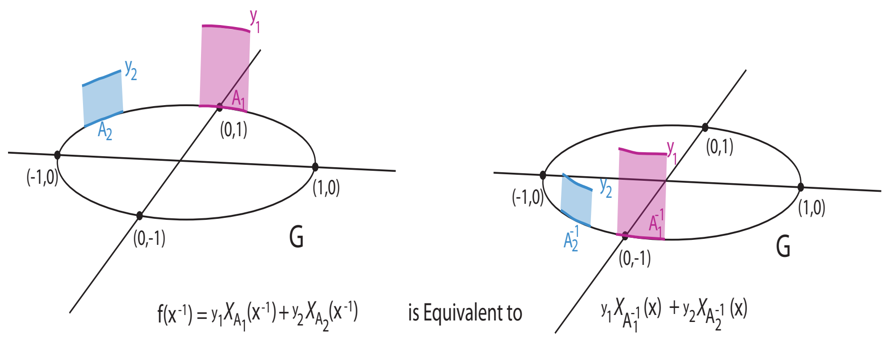


Figure 8.10: Let $G = S^1$. A step function on S^1 is represented by the top arcs of the colored vertical "rectangular" sheets. The step function $f(x^{-1}) = \sum_{k=1}^{2} y_k \chi_{A_k}(x^{-1})$ is equivalent to $\check{f}(x) = \sum_{k=1}^{2} y_k \chi_{A_k-1}(x)$.

equivalently

$$\int f(t^{-1})\, d\mu(t) = \int f(t)\Delta(t^{-1})\, d\mu(t),$$

and

$$\int f(s) d\check{\mu}(s) = \int f(s)\Delta(s^{-1}) d\mu(s), \quad \int f(s^{-1})\Delta(s^{-1}) d\mu(s) = \int f(s) d\mu(s)$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$.

Proof sketch. For every μ-step function

$$f = \sum_{i=1}^{n} y_i \chi_{A_i},$$

since $\check{f}(s) = f(s^{-1})$, we immediately obtain

$$\check{f} = \sum_{i=1}^{n} y_i \chi_{A_i^{-1}},$$

(see Figure 8.10) and since $\check{\mu}(A_i) = \mu(A_i^{-1})$, we get

$$\int \check{f}\, d\mu = \int f d\check{\mu}.$$

Then by a familiar argument using approximations sequences, we deduce that

$$\int \check{f}\, d\mu = \int f d\check{\mu} \tag{$*_{cech}$}$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$.

Using the fact that Δ is a group homomorphism, for any $f \in \mathcal{K}_\mathbb{C}(G)$, we have

$$\begin{aligned}\int (\rho_{t^{-1}}(f))(s)\Delta(s^{-1})d\mu(s) &= \Delta(t^{-1}) \int f(st^{-1})\Delta(t)\Delta(s^{-1})d\mu(s) && \text{by Definition of } \rho_{t^{-1}}(f)\\ &= \Delta(t^{-1}) \int f(st^{-1})\Delta^{-1}(st^{-1})d\mu(s)\\ &= \Delta(t^{-1})\Delta(t) \int f(s)\Delta^{-1}(s)d\mu(s) && \text{by Proposition 8.22}\\ & && \text{applied to } f\Delta^{-1}\\ &= \int f(s)\Delta(s^{-1})d\mu(s),\end{aligned}$$

which shows that the Radon functional $f \mapsto \int f(s)\Delta(s^{-1})d\mu(s)$ is right-invariant. The corresponding Haar measure ν is a right Haar measure, and since $\check{\mu}$ is a right Haar measure, there is some $a > 0$ such that $a\check{\mu} = \nu$. Then we have

$$a \int f(s)d\check{\mu}(s) = \int f(s)d\nu = \int f(s)\Delta(s^{-1})d\mu(s),$$

for all $f \in \mathcal{K}_\mathbb{C}(G)$, and since Δ^{-1} is a positive continuous function, by Proposition 8.26, $\nu = \Delta^{-1} \cdot \mu$, so

$$a\check{\mu} = \Delta^{-1} \cdot \mu.$$

It remains to show that $a = 1$.

Assume that $a \neq 1$. Since Δ is continuous, there is a symmetric neighborhod U of 1 such that $|\Delta(s^{-1}) - 1| \leq 1/2|a - 1|$ on U. Since U is symmetric, $\mu(U) = \check{\mu}(U)$, and we have

$$|a-1|\mu(U) = |a\check{\mu}(U) - \mu(U)| = \left|\int_U (\Delta(s^{-1}) - 1)d\mu(s)\right| \leq \frac{1}{2}|a-1|\mu(U),$$

a contradiction.

Therefore,

$$\int f(s)d\check{\mu}(s) = \int f(s)\Delta(s^{-1})d\mu(s),$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$, so

$$\int f(s)\Delta(s^{-1})d\mu(s) = \int \check{f}(s)\, d\mu(s),$$

and by changing f to $\check{f}$, we obtain the desired equation. $\square$

As a corollary, if G is unimodular, then we have

$$\int f(sx)d\mu(x) = \int f(xs)d\mu(x) = \int f(x^{-1})d\mu(x) = \int f(x)d\mu(x)$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$, and

$$\mu(A) = \mu(As) = \mu(sA) = \mu(A^{-1}),$$

for all $A \in \mathcal{B}$.

Remark: If G is a Lie group, then it can be shown that the modular function Δ is given by

$$\Delta(s) = |\det \mathrm{Ad}(s^{-1})|;$$

see Gallier and Quaintance [39] (Chapter 6, Proposition 6.25).

8.7 More Examples of Haar Measures

In the examples of Section 8.5, the groups under consideration were unimodular. The groups of the next examples are not unimodular.

Example 8.6. Let $G = \mathbf{GA}(n, \mathbb{R})$, the affine group of $\mathbb{R}^n$, which consists of pairs (A, u) with $A \in \mathbf{GL}(n, \mathbb{R})$ and $u \in \mathbb{R}^n$, acting on $\mathbb{R}^n$ by $(A, u)(X) = Ax + u$. It can be shown that a left Haar measure on $\mathbf{GA}(n, \mathbb{R})$ is given by

$$d\mu_L = |\det(A)|^{-n-1} \bigotimes_{i,j} da_{ij} \otimes \bigotimes_{i} du_i$$

with $A = (a_{ij})$, and $u = (u_i)$, where da_{ij} and du_i is the Lebesgue measure on $\mathbb{R}$. A right Haar measure is given by

$$d\mu_R = |\det(A)|^{-n} \bigotimes_{i,j} da_{ij} \otimes \bigotimes_{i} du_i,$$

and the modular function is given by

$$\Delta((A, u)) = |\det(A)|^{-1}.$$

A proof of these facts can be found in Bourbaki [7] (Chapter VII, Section 2, no. 10, Proposition 14, and Section 3, no. 3, Example 2). In particular, if $n = 1$, then an affine bijection is a map $x \mapsto ax + b$ with $a \neq 0$, and we have $d\mu_L = da\, db/a^2$, $d\mu_R = da\, db/|a|$, and $\Delta((a, b)) = |a|^{-1}$.

Remark: In view of Proposition 8.27, the value of the modular function is not unexpected.

Example 8.7. Let $G = \mathbf{T}(n, \mathbb{R})$, the group of invertible upper triangular matrices. It can be shown that a left Haar measure on $\mathbf{T}(n, \mathbb{R})$ is given by

$$d\mu_L = \prod_{i=1}^{n} |a_{ii}|^{i-n-1} \bigotimes_{i \le j} da_{ij}$$

with $A = (a_{ij})$, and da_{ij} is the Lebesgue measure on $\mathbb{R}$. A right Haar measure is given by

$$d\mu_R = \prod_{i=1}^{n} |a_{ii}|^{-i} \bigotimes_{i \le j} da_{ij},$$

and the modular function is given by

$$\Delta(A) = \prod_{i=1}^{n} |a_{ii}|^{2i-n-1}.$$

A proof of these facts can be found in Bourbaki [7] (Chapter VII, Section 2, no. 10, Proposition 14, and Section 3, no. 3, Example 4).

Remark: In view of Proposition 8.27, the value of the modular function is not unexpected.

More examples can be found in Bourbaki [7] (Chapter VII, Section 3, no. 3). The group $\mathbf{SL}(n, \mathbb{R})$ is unimodular, but finding a Haar measure for it is nontrivial.

8.8 The Modulus of an Automorphism

We now consider the effect of an automorphism $u\colon G \to G$ on a Haar measure. Recall that u is a group isomorphism and a homeomorphism.

Definition 8.16. Let G be a locally compact group and let $u\colon G \to G$ be an automorphism of G. For every function $f\colon G \to \mathbb{C}$, define the function $u(f)$ by

$$(u(f))(s) = f(u^{-1}(s)), \quad \text{for all } s \in G,$$

(see Figure 8.11), and for every left Haar measure μ, define the measure $u^{-1}(\mu)$ by

$$(u^{-1}(\mu))(A) = \mu(u(A)), \quad \text{for all } A \in \mathcal{B};$$

see Figure 8.12.

It is immediately verified that if u and v are two automorphisms of G, then

$$(u \circ v)(f) = (u(v(f)), \quad (u \circ v)(\mu) = (u(v(\mu)).$$

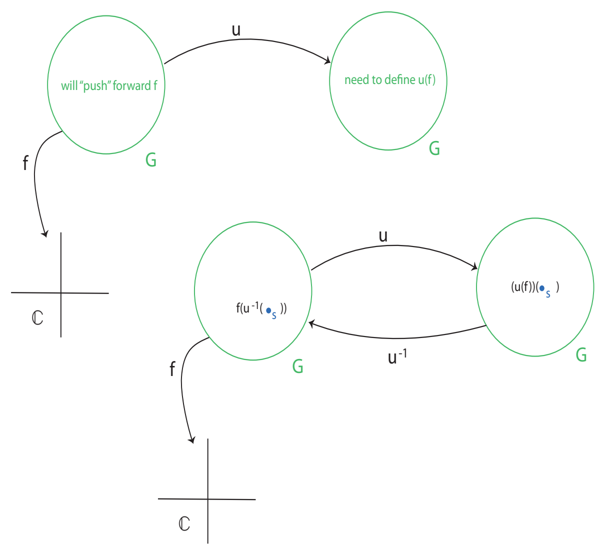


Figure 8.11: A schematic illustration of the "push forward" function $u(f)$.

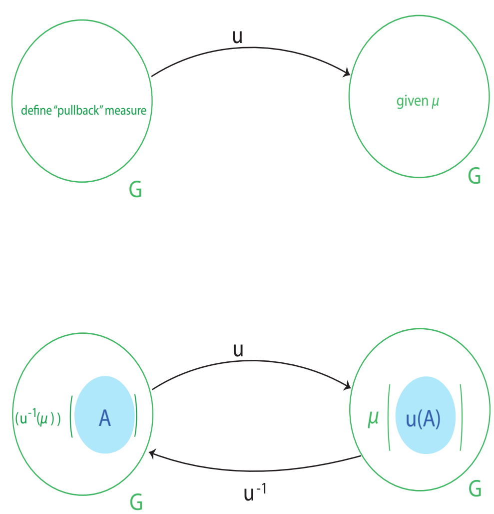


Figure 8.12: A schematic illustration of the "pullback" measure $u^{-1}(\mu)$.

Also observe that since μ is left-invariant and u is an automorphism,

$$(u^{-1}(\mu))(sA) = \mu(u(sA)) = \mu(u(s)u(A)) = \mu(u(A)) = u^{-1}(\mu)(A),$$

so $u^{-1}(\mu)$ is left-invariant. By the uniqueness of a left Haar measure up to a constant, there is a real $a > 0$ such that $u^{-1}(\mu) = a\mu$. For any other left Haar measure $\nu = c\mu$, we have

$$(u^{-1}(\nu))(A) = \nu(u(A)) = c\mu(u(A)) = cu^{-1}(\mu)(A) = ca\mu(A) = ac\mu(A) = a\nu(A).$$

Therefore, the constant a is independent of the left Haar measure μ.

Definition 8.17. Let G be a locally compact group. For every automorphism $u\colon G \to G$, there is a unique positive number $\mathrm{mod}(u)$ such that

$$u^{-1}(\mu) = \mathrm{mod}(u)\mu$$

for all left Haar measures μ. The number $\mathrm{mod}(u)$ is called the *modulus of the automorphism* u.

Note that the condition of Definition 8.17 can also be expressed as

$$\mu(u(A)) = \mathrm{mod}(u)\mu(A) \quad \text{for all } A \in \mathcal{B}. \tag{$**$}$$

Proposition 8.28. *Let G be a locally compact group and let μ be any left Haar measure on G. For every automorphism $u\colon G \to G$, we have*

$$\int u(f)d\mu = \int f du^{-1}(\mu) = \mathrm{mod}(u) \int f d\mu.$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$.

Proof sketch. For every μ-step function

$$f = \sum_{i=1}^n y_i \chi_{A_i},$$

since $(u(f))(s) = f(u^{-1}(s))$, we have $f(u^{-1}(s)) = y_i$ iff $u^{-1}(s) \in A_i$ iff $s \in u(A_i)$, which means that

$$u(f) = \sum_{i=1}^n y_i \chi_{u(A_i)};$$

see Figure 8.13.

Thus

$$\int u(f)\, d\mu = \sum_{i=1}^n y_i \mu(u(A_i)) = \sum_{i=1}^n y_i u^{-1}(\mu)(A_i) = \int f du^{-1}(\mu).$$

Then by a familiar argument using approximations sequences, we deduce that

$$\int u(f)d\mu = \int f du^{-1}(\mu)$$

for all $f \in \mathcal{L}^1_\mu(G, \mathcal{B}, \mathbb{C})$. Since $u^{-1}(\mu) = \mathrm{mod}(u)\mu$, we get the second equation. $\square$

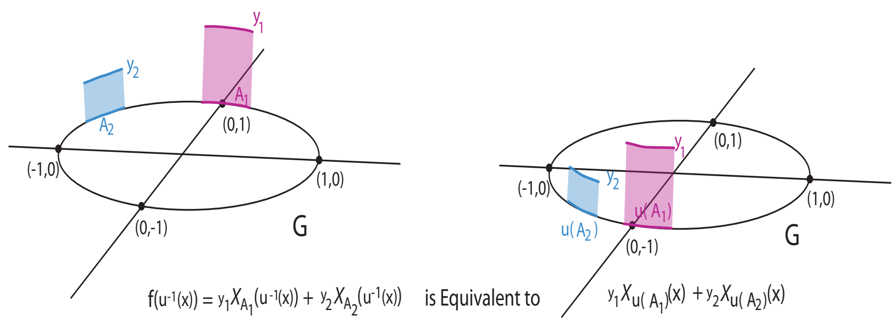


Figure 8.13: Let $G = S^1$. A step function on S^1 is represented by the top arcs of the colored vertical "rectangular" sheets. The step function $f(u^{-1}(x)) = \sum_{k=1}^{2} y_k \chi_{A_k}(u^{-1}(x))$ is equivalent to $(u(f))(x) = \sum_{k=1}^{2} y_k \chi_{u(A_i)}(x)$.

Proposition 8.28 can also be stated as

$$\int f(u^{-1}(s))d\mu(s) = \mathrm{mod}(u) \int f(s)d\mu(s).$$

Suppose that μ is a right Haar measure. As in the case of a left Haar measure, we define the measure $u^{-1}(\mu)$ by

$$(u^{-1}(\mu))(A) = \mu(u(A)) \quad \text{for all } A \in \mathcal{B}(G).$$

The measure $u^{-1}(\mu)$ is a right Haar measure because

$$(u^{-1}(\mu))(As) = \mu(u(As)) = \mu(u(A)u(s)) = \mu(u(A)) = (u^{-1}(\mu))(A).$$

As in the left-invariant case, for every automorphism $u \colon G \to G$, there a constant $c > 0$ such that $u^{-1}(\mu) = c\mu$. Interestingly, $c = \mathrm{mod}(u)$, so there is *no difference* between the left modulus and the right modulus of an automorphism.

Proposition 8.29. *Let G be a locally compact group and let $u \colon G \to G$ be an automorphism. Then for all left Haar measures and all right Haar measures μ on G, we have*

$$u^{-1}(\mu) = \mathrm{mod}(u)\mu,$$

where $\mathrm{mod}(u)$ *is the modulus of u defined for left Haar measures (see Definition 8.17).*

Proof. We use Corollary 8.24 which implies that $\Delta \circ u^{-1} = \Delta$, since u^{-1} is also an automorphism when u is an automorphism. As a consequence,

$$\Delta(s^{-1}) = \Delta(u^{-1}(s^{-1})) = \Delta((u^{-1}(s))^{-1}) = \Delta^{-1}(u^{-1}(s)),$$

that is,

$$\Delta(s^{-1}) = \Delta^{-1}(u^{-1}(s)). \tag{$\dagger$}$$

Recall that if μ is a left Haar measure, then $\check{\mu}$ is a right Haar measure, and by Proposition 8.27 we have $\check{\mu} = \Delta^{-1} \cdot \mu$. Then for every $f \in \mathcal{K}_{\mathbb{C}}(G)$, since $(u(f))(s) = f(u^{-1}(s))$, we have

$$\begin{aligned}
\int f(s)\, du^{-1}(\Delta^{-1} \cdot \mu)(s) &= \int (u(f))(s) d(\Delta^{-1} \cdot \mu)(s) && \text{by Proposition 8.28} \\
&= \int f(u^{-1}(s))\Delta(s^{-1}) d\mu(s) \\
&= \int f(u^{-1}(s))\Delta^{-1}(u^{-1}(s)) d\mu(s) && \text{by } (\dagger) \\
&= \mathrm{mod}(u) \int f(s)\Delta^{-1}(s) d\mu(s) && \text{by Proposition 8.28} \\
&= \mathrm{mod}(u) \int f(s)\Delta(s^{-1}) d\mu(s) \\
&= \mathrm{mod}(u) \int f(s)\, d(\Delta^{-1} \cdot \mu)(s).
\end{aligned}$$

By the uniqueness of the Radon measure associated with a Radon functional, this proves that

$$u^{-1}(\Delta^{-1} \cdot \mu) = \mathrm{mod}(u)\, \Delta^{-1} \cdot \mu,$$

and by Proposition 8.27, we obtain, $u^{-1}(\check{\mu}) = \mathrm{mod}(u)\check{\mu}$. Since every right Haar measure is of the form $\check{\mu}$ for some left Haar measure μ, we proved our result. $\square$

For every $s \in G$, if C_s is the automorphism conjugation by s, namely $C_s(t) = sts^{-1}$, then we have the following result.

Proposition 8.30. *Let G be a locally compact group. For every $s \in G$, we have*

$$\mathrm{mod}(C_s) = \Delta(s^{-1}).$$

Proof. We prove that $\mathrm{mod}(C_{s^{-1}}) = \Delta(s)$, which is equivalent to the equation of the Proposition. By Definition 8.16, for any left Haar measure μ,

$$(C_{s^{-1}}^{-1}(\mu))(A) = \mu(C_{s^{-1}}(A)) = \mu(s^{-1}As) = (\rho_s(\lambda_s(\mu)))(A).$$

Since μ is left-invariant, $\lambda_s(\mu) = \mu$, and by definition of the modulus $\rho_s(\mu) = \Delta(s)\mu$, so

$$C_{s^{-1}}^{-1}(\mu) = \rho_s(\lambda_s(\mu)) = \rho_s(\mu) = \Delta(s)\mu,$$

which by Definition 8.17 shows that

$$\mathrm{mod}(C_{s^{-1}}) = \Delta(s),$$

as claimed. $\square$

Proposition 8.31. *Let G be a locally compact group.*

(1) If G is compact or discrete, then $\mathrm{mod}(u) = 1$ *for any automorphism $u\colon G \to G$.*

(2) For any two automorphisms $u\colon G \to G$ and $v\colon G \to G$, we have

$$\mathrm{mod}(u \circ v) = \mathrm{mod}(u)\,\mathrm{mod}(v).$$

Proof. (1) Since u is an automorphism $u(G) = G$ and $u(\{1\}) = \{1\}$. If G is compact, let $A = G$ in Equation $(**)$ to obtain

$$\mu(G) = \mu(u(G)) = \mathrm{mod}(u)\mu(G),$$

and if G is discrete, let $A = \{1\}$, where μ is any left Haar measure.

(2) Using Equation $(**)$, we have

$$\mu((u \circ v)(A)) = \mu(u(v(A))) = \mathrm{mod}(u)\mu(v(A)) = \mathrm{mod}(u)\,\mathrm{mod}(v)\mu(A)),$$

and we choose A to be any open nonempty subset. $\square$

If $G = \mathbb{R}^n$ (which is a locally compact group under addition with the topology induced by any norm), and u a linear automorphism of $\mathbb{R}^n$, that is, an invertible linear map of $\mathbb{R}^n$, then we have the following interesting characterization of $\mathrm{mod}(u)$.

Proposition 8.32. *Let $u\colon \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear map, with $\mathbb{R}^n$ as an additive group with the Lebesgue measure. Then*

$$\mathrm{mod}(u) = |\det(u)|.$$

Sketch of proof. This result makes use of the following fact from linear algebra which is stated in Gallier and Quaintance [40] Chapter 7, Proposition 7.18, which can be restated as stating that every real $n \times n$ invertible matrix can be expressed as the product of elementary matrices $E_{i,j;\beta} = I_n + \beta E_{ij}$ and $I_n + (\alpha - 1)E_{nn}$. Then one must check the formula of the proposition,

$$\int u(f)(x_1, \ldots, x_n) d\mu(x_1, \ldots, x_n) = \mathrm{mod}(u) \int f(x_1, \ldots, x_n) d\mu(x_1, \ldots, x_n),$$

by integrating the functions of the form

$$f(x_1, \ldots, x_{n-1}, \alpha x_n)$$

and

$$f(x_1, \ldots, x_j + \beta x_i, \ldots, x_n),$$

with $f \in \mathcal{K}_{\mathbb{R}}(\mathbb{R}^n)$, using a change of variables. Details can be found in Dieudonné [24] (Chapter XIV, Proposition 14.3.9.1). $\square$

As an application of Proposition 8.32, we obtain formulae for the measure (volume) of a parallelotope and of a simplex in $\mathbb{R}^n$.

Let $(v_1, \ldots, v_n)$ be n linearly independent vectors in $\mathbb{R}^n$. Then the set

$$P = \{\lambda_1 v_1 + \cdots + \lambda_n v_n \mid (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n,\ 0 \leq \lambda_i \leq 1\}$$

is called a *parallelotope*; see Figure 8.14. The set

$$S = \{\lambda_1 v_1 + \cdots + \lambda_n v_n \mid (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n,\ \lambda_i \geq 0,\ \lambda_1 + \cdots + \lambda_n \leq 1\}$$

is called a *simplex*; see Figure 8.15.

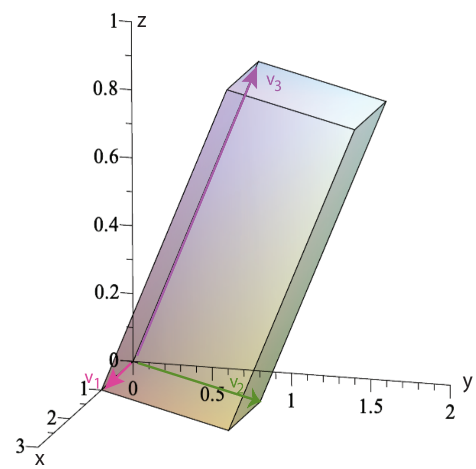


Figure 8.14: The parallelotope in $\mathbb{R}^3$ spanned by the vectors $v_1 = (1, 0, 0)$, $v_2 = (1, 1, 0)$, and $v_3 = (1, 1, 1)$.

Proposition 8.33. *Let $(v_1, \ldots, v_n)$ be n linearly independent vectors in $\mathbb{R}^n$, and let P be the parallelotope and S be the simplex determined by $(v_1, \ldots, v_n)$. If μ is the Lebesgue measure on $\mathbb{R}^n$, then*

$$\mu(P) = |\det(v_1, \ldots, v_n)|, \quad \mu(S) = \frac{1}{n!}|\det(v_1, \ldots, v_n)|.$$

Proof sketch. Since $(v_1, \ldots, v_n)$ are linearly independent, there is a unique linear map u such that $u(e_i) = v_i$, for $i = 1, \ldots, n$, where e_i is the canonical basis vector of $\mathbb{R}^n$. Then $P = u(K)$, where K is the n-cube determined by $(e_1 \ldots, e_n)$, and

$$\mu(P) = \int \chi_P d\mu = \int u(\chi_K) d\mu = \mathrm{mod}(u) \int \chi_K d\mu = |\det(u)|\mu(K).$$

But under the Lebesgue measure, $\mu(K) = 1$, and we get

$$\mu(P) = |\det(u)| = |\det(v_1, \ldots, v_n)|,$$

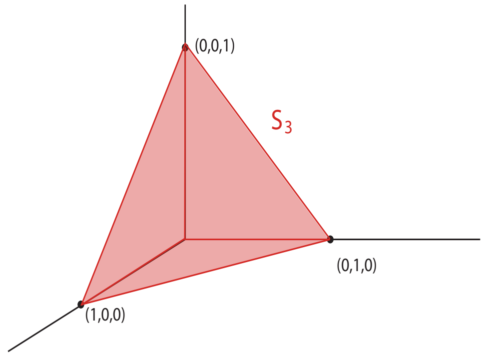


Figure 8.15: The standard simplex S_3 is the solid tetrahedron spanned by the basis vectors e_1, e_2, and e_3.

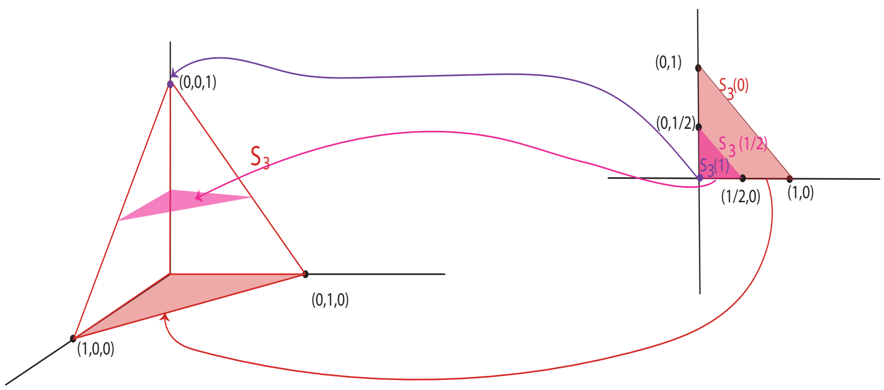


Figure 8.16: The standard simplex S_3 with its embedded cross sections $S_n(\lambda)$, where $0 \leq \lambda \leq 1$.

as claimed.

For the simplex, write S_m for simplex determined by the canonical basis vectors $e_1, \ldots, e_m$, and write μ_m for the Lebesgue measure in $\mathbb{R}^m$. Then $S = u(S_n)$, so by a similar reasoning

$$\mu(S) = |\det(u)|\mu(S_n),$$

and we are reduced to computing $\mu(S_n)$. We view $\mathbb{R}^n$ as $\mathbb{R}^{n-1} \times \mathbb{R}$ and we consider the section $S_n(\lambda)$ of S_n consisting of the set of points $(x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$ such that

$$x_1, \ldots, x_{n-1} \geq 0, \quad x_1 + \cdots + x_{n-1} \leq 1 - \lambda,$$

where $0 \leq \lambda \leq 1$; see Figure 8.16.

This section is the image of $S_n(0)$ under the scaling by $1 - \lambda$, so we get

$$\mu_{n-1}(S_n(\lambda)) = (1 - \lambda)^{n-1}\mu_{n-1}(S_{n-1}).$$

Then by Fubini we get

$$\mu_n(S_n) = \int_0^1 (1-\lambda)^{n-1}\mu_{n-1}(S_{n-1})d\lambda = \frac{1}{n}\mu_{n-1}(S_{n-1}).$$

By induction

$$\mu_{n-1}(S_{n-1}) = \frac{1}{(n-1)!},$$

so we get

$$\mu_n(S_n) = \frac{1}{n!}$$

as claimed. $\square$

As another application of Proposition 8.32, the computation of the measure (volume) of a closed ball in $\mathbb{R}^n$ can be found in Dieudonné [24] (Chapter XIV, Proposition 14.3.11).

More on the modular function and the modulus of an automorphism can be found in Bourbaki [7], Chapter VII, Section 1.

8.9 Some Properties and Applications of the Haar Measure

Since the Haar measure is σ-regular and locally finite, Theorem 7.11 implies the following result which will be needed in Chapter 12.

Theorem 8.34. *Let G be a locally compact, metrizable, separable group equipped with a left Haar measure. Then $\mathrm{L}^p_\mu(G, \mathbb{C})$ is separable for $p = 1, 2$.*

The following result is proven in Dieudonné [24] (Chapter XIV, Proposition 14.2.3).

Proposition 8.35. *Let G be a locally compact group, and let μ be a left Haar measure on G. Then G is discrete if and only if $\mu(\{1\}) > 0$, and G is compact if and only if $\mu(G) < +\infty$.*

An interesting and important application of the Haar measure is the construction of a Hermitian inner product invariant under the representation of a compact group. The idea of such a construction originates with Hurwitz and was generalized by H. Weyl.

Let K be a topological group, and let H be a (complex) finite-dimensional Hermitian space (with inner product $\langle -, - \rangle$ and corresponding norm $\| \, \|$). A *representation* of K in H is a group homomorphism $U \colon K \to \mathbf{GL}(H)$, where $\mathbf{GL}(H)$ is the group of invertible linear maps on H.

Theorem 8.36. *Let K be a compact group, let H be a finite-dimensional Hermitian space, and let $U\colon K \to \mathbf{GL}(H)$ be a representation of K which is continuous when $\mathbf{GL}(H)$ is equipped with the operator norm. Then there is a Hermitian inner product φ on H such that*

$$\varphi(U_s(x), U_s(y)) = \varphi(x, y) \quad \text{for all } s \in K \text{ and all } x, y \in H.$$

In other words, the linear maps U_s are unitary transformations with respect to φ. Furthermore, the norms $\| \ \|$ and $x \mapsto \sqrt{\varphi(x, x)}$ on H are equivalent.

Proof. Let μ be a right Haar measure on K (which is also a left Haar measure since K is compact). By hypothesis the map

$$s \mapsto \langle U_s(x), U_s(y) \rangle$$

is continuous for all $x, y \in H$. Define φ by

$$\varphi(x, y) = \int \langle U_s(x), U_s(y) \rangle \, d\mu(s).$$

It is immediately verified that φ is a sesquilinear form on H. Since H is finite-dimensional, the sphere $S = \{x \in H \mid \|x\| = 1\}$ is compact. Since U is continuous, the map $\theta\colon K \times S \to \mathbb{R}$ given by

$$\theta(s, x) = \|U_s(x)\|$$

is continuous, and since K and S are compact, $K \times S$ is compact so θ achieves a minimum $m > 0$ and a maximum $M > 0$ (every map U_s is invertible and for $x \in S$, $U_s(x) \neq 0$ since $x \neq 0$). We deduce that for every $x \neq 0$,

$$m \|x\| \leq \|x\| \left\| U_s\left(\frac{x}{\|x\|}\right) \right\| \leq M \|x\|,$$

that is,

$$m \|x\| \leq \|U_s(x)\| \leq M \|x\|.$$

As a consequence, we get

$$m^2 \mu(K) \|x\|^2 \leq \varphi(x, x) \leq M^2 \mu(K) \|x\|^2,$$

which shows that φ is indeed positive definite, and that $\| \ \|$ is equivalent to the norm induced by φ. Finally, for every $t \in K$, since μ is right-invariant we have

$$\begin{aligned}
\varphi(U_t(x), U_t(y)) &= \int \langle U_s(U_t(x)), U_s(U_t(y)) \rangle \, d\mu(s) \\
&= \int \langle U_{st}(x)), U_{st}(y) \rangle \, d\mu(s) \\
&= \int \langle U_s(x)), U_s(y) \rangle \, d\mu(s) \\
&= \varphi(x, y),
\end{aligned}$$

as claimed. $\square$

Theorem 8.36 is a basic tool in representation theory. For example, if G is a Lie group and if V is a finite-dimensional vector space, for any representation $\rho\colon G \to \mathbf{GL}(V)$, there is a G-invariant inner product on V iff $\overline{\rho(G)}$ is compact; see Gallier and Quaintance [38] (Chapter 21, Theorem 21.5).

Theorem 8.36 will also be used in Section 12.2 to show that every linear representation of a compact group is the sum of irreducible representations.

Regarding the product of Haar measures, we have the following result.

Proposition 8.37. *Let G_1 and G_2 be two locally compact groups, and let μ_1 be a left Haar measure on G_1 and μ_2 be a left Haar measure on G_2. Then the linear functional $\Phi_1 \otimes \Phi_2\colon \mathcal{K}_{\mathbb{C}}(G_1 \times G_2) \to \mathbb{C}$ given by*

$$(\Phi_1 \otimes \Phi_2)(f) = \int f(x_1, x_2) d\mu_1(x_1) \otimes d\mu_2(x_2)$$

is a left-invariant positive Radon functional. If G_1 and G_2 are σ-compact, then the Radon measure $\mu_{\Phi_1 \otimes \Phi_2}$ on $G_1 \times G_2$ associated with $\Phi_1 \otimes \Phi_2$ given by Theorem 7.8 is a left Haar measure extending the product measure $\mu_1 \otimes \mu_2$. Furthermore, if G_1 and G_2 are also second-countable, then $\mu_{\Phi_1 \otimes \Phi_2} = \mu_1 \otimes \mu_2$.

Proof sketch. By Fubini's Theorem (which applies since f vanishes outside of a compact subset), we have

$$\begin{aligned}
(\lambda_{(s_1,s_2)}(\Phi_1 \otimes \Phi_2))(f) &= (\Phi_1 \otimes \Phi_2)(\lambda_{(s_1^{-1},s_2^{-1})}(f)) \\
&= \int (\lambda_{(s_1^{-1},s_2^{-1})} f)(x_1, x_2) d\mu_1(x_1) \otimes d\mu_2(x_2) \\
&= \int f(s_1x_1, s_2x_2) d\mu_1(x_1) \otimes d\mu_2(x_2) \\
&= \int \left(\int f(s_1x_1, s_2x_2) d\mu_1(x_1) \right) d\mu_2(x_2) \\
&= \int \left(\int f(x_1, s_2x_2) d\mu_1(x_1) \right) d\mu_2(x_2) \\
&= \int \left(\int f(x_1, s_2x_2) d\mu_2(x_2) \right) d\mu_1(x_1) \\
&= \int \left(\int f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) \\
&= \int \left(\int f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2) \\
&= \int f(x_1, x_2) d\mu_1(x_1) \otimes d\mu_2(x_2) = (\Phi_1 \otimes \Phi_2)(f).
\end{aligned}$$

Therefore, $\Phi_1 \otimes \Phi_2$ is left-invariant. The other two statements are explained in Folland [33] (Chapter 2, Section 2.2). $\square$

Remark: Since $\Phi_1 \otimes \Phi_2$ is a positive linear functional, by Theorem 7.8, the corresponding Radon measure $\mu_{\Phi_1\otimes\Phi_2}$ is a left Haar measure on $G_1 \times G_2$. But if G_1 or G_2 is not σ-compact then the product measure $\mu_1\otimes\mu_2$ is not defined, and if G_1 or G_2 is not second-countable, then the σ-algebra associated with $\mu_{\Phi_1\otimes\Phi_2}$ has more Borel subsets than the σ-algebra associated with $\mu_1 \otimes \mu_2$ (see Definition 5.23).

As an application of Proposition 8.37, since the Lebesgue measure μ_L on $\mathbb{R}$ is both left and right-invariant, we see that the product measure $\mu_{L,n}$ of n copies of μ_L is a left and a right Haar measure on $\mathbb{R}^n$. To simplify notation, we may write μ_n instead of $\mu_{L,n}$, and $\mathcal{L}^1(\mu_n)$ instead of $\mathcal{L}^1_{\mu_n}(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{C})$.

As a Haar measure, μ_n is both inner and outer regular.

8.10 G-Invariant Measures on Homogeneous Spaces

Let X be a locally compact space and let G be a locally compact group. Suppose we have a continuous left action $\varphi\colon G\times X \to X$ of G on X (which means that the map φ is continuous, see Definition 8.5). As usual, we write $g \cdot x$ instead of $\varphi(g, x)$. We would like to generalize the notion of left-invariance of a measure on G to the notion of G-invariance of a measure μ on X. This is easily done by replacing multiplication in G by the action of G on X.

Definition 8.18. Let G be a locally compact group, let X be a locally compact space, and let $\cdot\colon G\times X \to X$ be a continuous left action of G on X. For every $s \in G$, define $L_s\colon X \to X$ by

$$L_s(x) = s \cdot x \quad \text{for all } x \in X.$$

For every subset A of X and every $s \in G$, let

$$s \cdot A = \{s \cdot a \mid a \in A\}.$$

For every function $f\colon X \to \mathbb{C}$, the function $\lambda_s(f)$ is given by

$$(\lambda_s(f))(x) = f(s^{-1} \cdot x) \quad \text{for all } x \in X \text{ and } s \in G,$$

For every Borel measure μ on $(X, \mathcal{B}(X))$, the measure $\lambda_s(\mu)$ given by

$$(\lambda_s(\mu))(A) = \mu(s^{-1} \cdot A) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B}(X).$$

For every Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$, the Radon functional $\lambda_s(\Phi)$ given by

$$(\lambda_s(\Phi))(f) = \Phi(\lambda_{s^{-1}}(f)) \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_{\mathbb{C}}(X).$$

It is immediately verified that

$$L_{st} = L_s \circ L_t, \quad \lambda_{st}(f) = \lambda_s(\lambda_t(f)),$$

and

$$\lambda_{st}(\mu) = \lambda_s(\lambda_t(\mu)), \quad \lambda_{st}(\Phi) = \lambda_s(\lambda_t(\Phi)).$$

The proof of Proposition 8.16 is immediately adapted to show that

$$\int_X (\lambda_{s^{-1}}(f))(x)\, d\mu(x) = \int_X f(s \cdot x)\, d\mu(x) = \int_X f(x)\, d\lambda_s(\mu)(x),$$

for every $s \in G$, every $f \in \mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$, and every Borel measure μ on X.

Definition 8.19. Let G be a locally compact group, let X be a locally compact space, and let $\cdot\colon G \times X \to X$ be a continuous left action of G on X. A Borel measure μ on X is G-*invariant* if

$$\lambda_s(\mu) = \mu \quad \text{for all } s \in G.$$

A Radon functional $\Phi\colon \mathcal{K}_{\mathbb{C}}(X) \to \mathbb{C}$ on X is G-*invariant* if

$$\lambda_s(\Phi) = \Phi \quad \text{for all } s \in G.$$

If μ is G-invariant, then

$$\int_X f(s \cdot x)\, d\mu(x) = \int_X f(x)\, \mu(x),$$

for every $f \in \mathcal{L}^1_\mu(X, \mathcal{B}, \mathbb{C})$ and all $s \in G$. The proof of Proposition 8.16 is immediately adapted to show that if

$$\int_X \lambda_{s^{-1}}(f)(x)\, d\mu(x) = \int_X f(s \cdot x)\, d\mu(x) = \int_X f(x)\, d\mu(x),$$

for all $f \in \mathcal{K}_{\mathbb{C}}(X)$ and all $s \in G$, then μ is G-invariant.

Our goal is to find sufficient conditions to ensure that X has some G-invariant measure. We will consider the case where $X = G/H$, with the left action of G on G/H given by

$$a \cdot (bH) = abH, \qquad a, b \in G;$$

see Figure 8.17. In this case, by Proposition 8.6, the space X is also locally compact (and Hausdorff).

A G-invariant measure on G/H does not always exist. For example, if G is the affine "$ax + b$" group (with $a \neq 0$) and $X = \mathbb{R}$, obviously G acts transitively on $\mathbb{R}$ (see Example 8.6 for the definition of the action) and the stabilizer of 0 is $H = \mathbb{R}$. However, the only Borel measure on $\mathbb{R}$ invariant under translation is the Lebesgue measure, but it is not invariant under scaling transformations $x \mapsto ax$ with $a \neq 0, 1$.

It turns out that there is a necessary and sufficient condition for a G-invariant σ-Radon measure to exist on G/H in terms of Δ_G and Δ_H: Δ_H must the equal to the restriction of

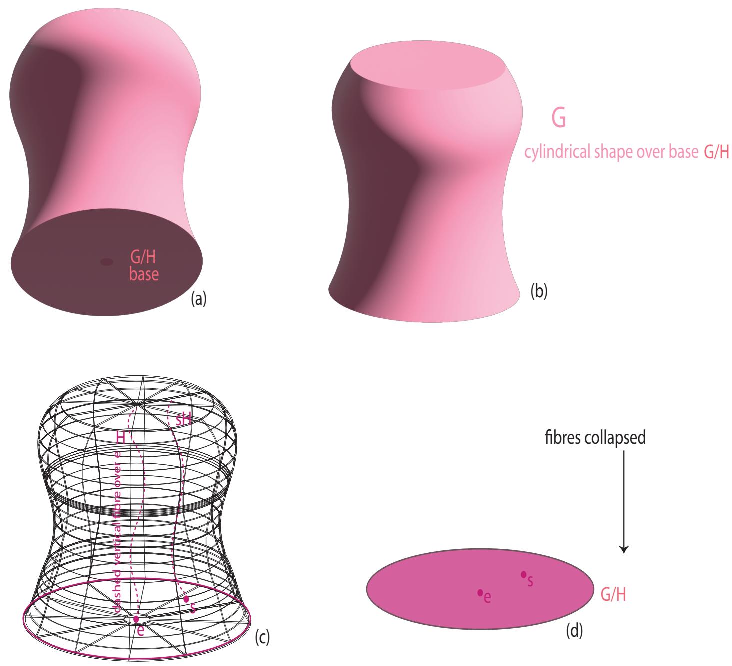


Figure 8.17: Let G be the solid pink "cylindrical" shape; see Figures (a) and (b). The fibres sH are represented by wavy vertical lines over the circular base; see Figure (c). When these fibres are identified to the base point, we have effectively "collapsed" G to the circular base G/H; see Figure (d).

Δ_G on H. We proceed to explain this following Folland's exposition [33] (Chapter 2, Section 2.6).

Suppose μ is a left Haar measure on G and ξ is a left Haar measure on H. The group G is locally compact and σ-compact, and H is a closed subgroup of G. Denote the quotient map by $\pi\colon G \to G/H$. The first step is to define a map P from $\mathcal{K}_{\mathbb{C}}(G)$ to $\mathcal{K}_{\mathbb{C}}(G/H)$.

Definition 8.20. With (G, μ) and (H, ξ) as above, let $P\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G/H)$ be the function defined as follows: for every $f \in \mathcal{K}_{\mathbb{C}}(G)$, for every $s \in G$, let

$$(P(f))(sH) = \int_H f(sh)\, d\xi(h);$$

see Figure 8.18.

We need to check that the map P is well-defined, that is, if $sH = tH$, then $(P(f))(sH) = (P(f))(tH)$, but this follows from the left-invariance of ξ since

$$(P(f))(sH) = \int_H f(sh)\, d\xi(h) = \int_H f(h)\, d\xi(h) = \int_H f(th)\, d\xi(h) = (P(f))(tH).$$

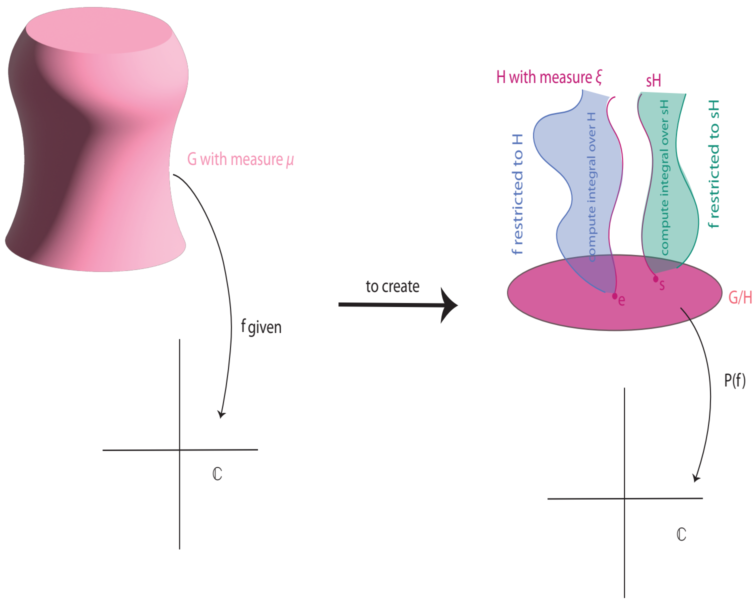


Figure 8.18: Let G, H, and G/H be as in Figure 8.17. The right figure is a schematic interpretation of $P\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G/H)$. For each $f \in \mathcal{K}_{\mathbb{C}}(G)$, restrict the domain of f to be over the fibre sH, and then integrate over that fibre using the measure ξ. This integral is represented as the shaded "area" between the restricted function image and the fibre.

Roughly speaking, $P(f)(sH)$ is obtained by averaging over H. The following properties are immediately verified.

Proposition 8.38. *The function $P\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G/H)$ satisfies the following properties:*

(1) The function $P(f)$ is continuous.

(2) We have $\mathrm{supp}(P(f)) \subseteq \pi(\mathrm{supp}(f))$.

(3) For any $\varphi \in \mathcal{K}_{\mathbb{C}}(G/H)$, we have

$$P((\varphi \circ \pi)f) = \varphi\, P(f),$$

where $\varphi\, P(f)$ denotes the function defined by pointwise multiplication on G/H.

Our next goal is to show that P is surjective. The following technical result is needed.

Proposition 8.39. *For any compact subset F of G/H, there is a positive function $f \in \mathcal{K}_{\mathbb{C}}(G)$ such that $P(f) \equiv 1$ on F.*

Proof. Let E be a compact neighborhood of F in G/H. By Proposition 8.7 there is a compact subset K of G such that $\pi(K) = E$. Since G and G/H are locally compact, by

Proposition A.39 we can find a positive function $g \in \mathcal{K}_{\mathbb{C}}(G)$ such that g is strictly positive on K and a function $\varphi \in \mathcal{K}_{\mathbb{C}}(G/H)$ such that $\varphi \equiv 1$ on F and $\mathrm{supp}(\varphi) \subseteq E$. Define f by

$$f(s) = \begin{cases} \dfrac{\varphi(\pi(s))}{(P(g))(\pi(s))}\, g(s) & \text{if } (P(g))(\pi(s)) \neq 0 \\ 0 & \text{if } (P(g))(\pi(s)) = 0. \end{cases}$$

Since $P(g) > 0$ on $\mathrm{supp}(\varphi)$, the function f is continuous, we have $\mathrm{supp}(f) \subseteq \mathrm{supp}(g)$, and by Proposition 8.38(3) applied to $\left(\frac{\varphi}{P(g)} \circ \pi\right) g$, we have $P(f) = (\varphi/P(g))P(g) = \varphi$, as desired. $\square$

Using Proposition 8.39, we obtain the surjectivity of P.

Proposition 8.40. *For any function $\varphi \in \mathcal{K}_{\mathbb{C}}(G/H)$, there is some function $f \in \mathcal{K}_{\mathbb{C}}(G)$ such that $P(f) = \varphi$. Furthermore, $\pi(\mathrm{supp}(f)) = \mathrm{supp}(\varphi)$, and if $\varphi \geq 0$, then $f \geq 0$.*

Proof. If $\varphi \in \mathcal{K}_{\mathbb{C}}(G/H)$, by Proposition 8.39 there is some function $g \geq 0$ in $\mathcal{K}_{\mathbb{C}}(G)$ such that $P(g) \equiv 1$ on $\mathrm{supp}(\varphi)$. Let $f = (\varphi \circ \pi)g$. Then by Proposition 8.38(3), we have $P(f) = \varphi P(g) = \varphi$ since $P(g) \equiv 1$ on $\mathrm{supp}(\varphi)$. The other properties are immediately verified. $\square$

If G is a locally compact group and if H is a closed normal subgroup of G, then by Proposition 8.6, the group G/H is also a locally compact group. As application of the surjectivity of the map $P\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G/H)$, the following proposition shows how to integrate on G by integrating on H and G/H.

Proposition 8.41. *Let G be a locally compact group and let H be a closed normal subgroup of G. If ξ is a left Haar measure on H and if γ is a left Haar measure on G/H, then the functional*

$$f \mapsto \int_{G/H} P(f)(sH)\, d\gamma(sH) = \int_{G/H}\int_H f(sh)\, d\xi(h) d\gamma(sH), \quad f \in \mathcal{K}_{\mathbb{C}}(G)$$

is a left Haar functional on G. Consequently, for any left Haar measure μ on G, by rescaling ξ or γ, we have

$$\int_G f(s) d\mu(s) = \int_{G/H}\int_H f(sh)\, d\xi(h) d\gamma(sH).$$

Proof skech. The verification that the given functional is a positive and left-invariant functional is left as an easy exercise. The fact that the functional is not the zero functional follows immediately from the surjectivity of P. By Theorem 7.8, there is a left Haar measure corresponding to this positive Radon functional, and by uniqueness of the left Haar measure up to a scalar, we can rescale ξ or γ as desired. $\square$

We now come to our main theorem.

Theorem 8.42. *Let G be a locally compact group, and let H be closed subgroup of G. Suppose μ is a left Haar measure on G and ξ is a left Haar measure on H. There is a G-invariant σ-Radon measure γ on G/H if and only if Δ_H is equal to the restriction of Δ_G to H. In this case, γ is unique up to a scalar, and with a suitable choice of this factor, we have*

$$\int_G f(s)d\mu(s) = \int_{G/H} P(f)(sH)\, d\gamma(sH) = \int_{G/H}\int_H f(sh)\, d\xi(h)\, d\gamma(sH), \tag{$\dagger$}$$

for all $f \in \mathcal{K}_{\mathbb{C}}(G)$; see Figure 8.19.

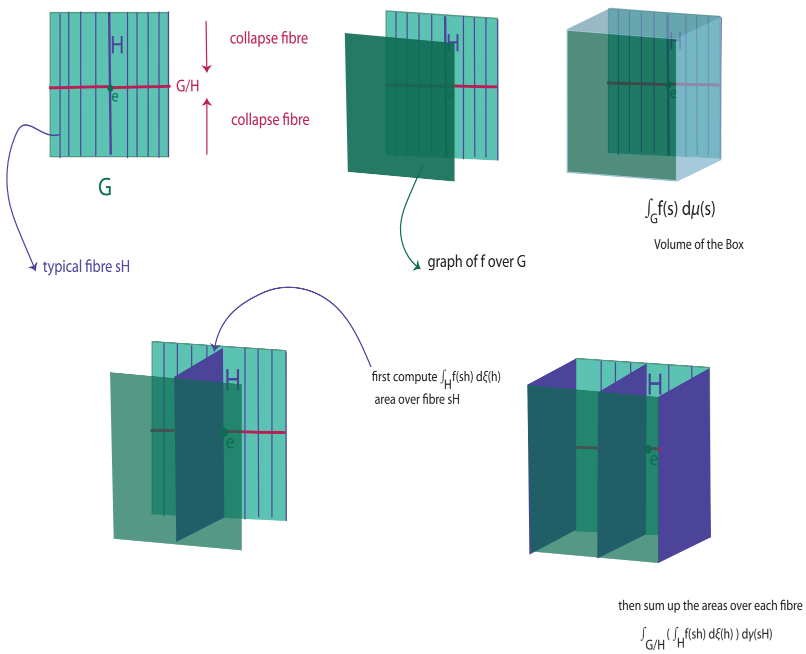


Figure 8.19: The group G is represented by the green square, the fibres are the vertical purple lines, and G/H is represented by the horizontal red line. The graph of the function $f \in \mathcal{K}_{\mathbb{R}}(G)$ is represented by a second green square "floating" above G. The "volume" below the graph of f is computed by $\int_G f(s)d\mu(s)$. This volume can also be computed in an iterative manner by first compute the "area" over a fibre sH, and then "summing" up the areas by varying the fibre over G/H. Algebraically, this iterative process corresponds to calculating $\int_{G/H}\int_H f(sh)\, d\xi(h)\, d\gamma(sH)$.

Proof. First suppose that a G-invariant σ-Radon measure γ on G/H exists. The map $f \mapsto \int P(f)\, d\gamma$ is a nonzero left-invariant positive linear functional on $\mathcal{K}_{\mathbb{C}}(G)$, so by the uniqueness of Haar measure on G there is some $c > 0$ such that

$$\int P(f)(sH)\, d\gamma(sH) = c\int f(s)d\mu(s). \tag{$*$}$$

By Radon-Riesz I (Theorem 7.8), the measure γ is uniquely determined by the functional $\varphi \mapsto \int \varphi(sH)\, d\gamma(sH)$ (with $\varphi \in \mathcal{K}_{\mathbb{C}}(G/H)$). By Proposition 8.40, since $P\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G/H)$ is surjective, $(*)$ determines this functional completely, and thus γ is also completely determined, so γ is unique up to the scalar determining a left Haar measure. By replacing γ by $c^{-1}\gamma$, equation $(\dagger)$ holds. Then for any $\eta \in H$ and $f \in \mathcal{K}_{\mathbb{C}}(G)$, we have

$$
\begin{aligned}
\Delta_G(\eta) \int_G f(s) d\mu(s) &= \int_G \rho_{\eta^{-1}}(f)(s)\, d\mu(s) && \text{by Proposition 8.22}\\
&= \int_{G/H}\int_H \rho_{\eta^{-1}}(f)(sh)\, d\xi(h)\, d\gamma(sH) && \text{by } (\dagger)\\
&= \int_{G/H}\int_H f(sh\eta^{-1})\, d\xi(h)\, d\gamma(sH) && \text{by definition of } \rho_{\eta^{-1}}(f)\\
&= \Delta_H(\eta) \int_{G/H}\int_H f(sh)\, d\xi(h)\, d\gamma(sH) && \text{by Proposition 8.22}\\
&= \Delta_H(\eta) \int_G f(s)\, d\mu(s), && \text{by } (\dagger)
\end{aligned}
$$

which implies that $\Delta_G(\eta) = \Delta_H(\eta)$.

Conversely, assume that Δ_H is equal to the restriction of Δ_G to H. We claim that if $f \in \mathcal{K}_{\mathbb{C}}(G)$ and if $P(f) = 0$, then $\int f(s) d\mu(s) = 0$.

By Proposition 8.39 there is a positive function $\varphi \in \mathcal{K}_{\mathbb{C}}(G)$ such that $P(\varphi) \equiv 1$ on $\pi(\mathrm{supp}(f))$. Therefore we have

$$
\begin{aligned}
0 = (P(f))(sH) &= \int_H f(sh) d\xi(h) && \text{by definition of } P(f)\\
&= \int_H f(sh^{-1}) \Delta_H(h^{-1}) d\xi(h) && \text{by Proposition 8.27}\\
&= \int_H f(sh^{-1}) \Delta_G(h^{-1}) d\xi(h), && \text{since } \Delta_H = \Delta_G \mid H
\end{aligned}
$$

which implies

$$
\begin{aligned}
0 &= \int_G \int_H \varphi(s) f(sh^{-1}) \Delta_G(h^{-1}) d\xi(h)\, d\mu(s) \\
&= \int_H \int_G \varphi(s) f(sh^{-1}) \Delta_G(h^{-1}) d\mu(s)\, d\xi(h) && \text{by Fubini} \\
&= \int_H \Delta_G(h^{-1}) \int_G \varphi(s) f(sh^{-1}) d\mu(s)\, d\xi(h) \\
&= \int_H \int_G \varphi(sh) f(s) d\mu(s)\, d\xi(h) && \text{by Proposition 8.22} \\
&= \int_G f(s) \int_H \varphi(sh) d\xi(h)\, d\mu(s) && \text{by Fubini} \\
&= \int_G (P(\varphi))(sH) f(s) d\mu(s) = \int f(s) d\mu(s) && \text{since } P(\varphi) \equiv 1 \text{ on } \pi(\mathrm{supp}(f)).
\end{aligned}
$$

What we just showed implies that if $P(f) = P(g)$ then $\int f(s)d\mu(s) = \int g(s)d\mu(s)$. Since by Proposition 8.40 the map $P\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G/H)$ is surjective, we define a functional Φ on $\mathcal{K}_{\mathbb{C}}(G/H)$ as follows: for every $\varphi \in \mathcal{K}_{\mathbb{C}}(G/H)$, let

$$
\Phi(\varphi) = \int_G f(s) d\mu(s) \quad \text{for any } f \in \mathcal{K}_{\mathbb{C}}(G) \text{ such that } P(f) = \varphi.
$$

Since $P(f) = P(g)$ implies that $\int f(s)d\mu(s) = \int g(s)d\mu(s)$, the functional Φ is well-defined, and it is immediately verified that Φ is a G-invariant positive linear functional on $\mathcal{K}_{\mathbb{C}}(G/H)$. By Radon–Riesz I (Theorem 7.8), this functional induces the desired G-invariant σ-Radon measure on G/H. $\square$

If H is compact then by Proposition 8.25(2), we have $\Delta_H = \Delta_G \mid H = 1$, so we obtain the following useful corollary.

Proposition 8.43. *If G is a locally compact group, for any compact subgroup H of G, the space G/H admits a G-invariant σ-Radon measure (unique up to a scalar). In fact, if $\pi\colon G \to G/H$ is the quotient map, then for any left Haar measure μ on G, there is a unique G-invariant σ-Radon measure γ on G/H such that*

$$
\int_{G/H} f(x)\, d\gamma(x) = \int_G (f \circ \pi)(s)\, d\mu(s), \qquad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G/H).
$$

Proof. If $f \in \mathcal{K}_{\mathbb{C}}(G/H)$ has compact support K, then $f \circ \pi$ has support homeomorphic to $K \times H$, and since K and H are compact, it is compact. Thus $f \circ \pi \in \mathcal{K}_{\mathbb{C}}(G)$. The functional $\Phi\colon \pi(f) \mapsto \int_G (f \circ \pi)(s)\, d\mu(s)$ is well-defined, clearly a positive linear functional on $\mathcal{K}_{\mathbb{C}}(G/H)$, and since μ is left-invariant, it is G-invariant. By Radon–Riesz I, there is a unique G-invariant σ-Radon measure γ corresponding to Φ. $\square$

Proposition 8.43 applies to the projective spaces and the Grassmannians with $G = \mathbf{SO}(n)$ and a suitable compact subgroup H. It also applies to $G = \mathbf{GL}(n, \mathbb{R})$ and $X = \mathbf{SPD}(n)$, where $\mathbf{SPD}(n)$ is the set of positive symmetric definite matrices discussed in Example C.11.

Example 8.8. Recall that the group $\mathbf{GL}(n) = \mathbf{GL}(n, \mathbb{R})$ acts on $\mathbf{SPD}(n)$ as follows: for all $A \in \mathbf{GL}(n)$ and all $S \in \mathbf{SPD}(n)$,

$$A \cdot S = ASA^{\top}.$$

This action is transitive, and in Section C.3, Example (d), we show that the stabilizer of I is $\mathbf{O}(n)$, so

$$\mathbf{GL}(n)/\mathbf{O}(n) = \mathbf{SPD}(n).$$

It can be shown that the unique (up to a scalar) $\mathbf{GL}(n, \mathbb{R})$-invariant measure on $\mathbf{SPD}(n)$ is given by

$$d\mu = (\det(H))^{-(n+1)/2} d\eta(H),$$

where η is the Haar measure on the additive group $\mathbf{S}(n)$ of real symmetric matrices. For a proof, see Bourbaki [7] (Chapter VII, Section 3, no. 3, Example 8).

When the condition $\Delta_H = \Delta_G \mid H$ fails, it is possible relax the notion of G-invariance and to obtain sufficient conditions for the existence of measures on G/H satisfying weaker invariance conditions. Such notions are relative invariance, quasi-invariance, and strong quasi-invariance. The interested reader should consult Folland [33] (Chapter 2, Section 2.6) and Bourbaki [7] (Chapter VII, Section 2).

8.11 Convolution of Measures

Let G be a locally compact group equipped with a left Haar measure λ. Recall that $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(G)$ denotes the Banach space of complex regular Borel measures on G (see Definition 7.22), and that $\mathrm{L}^1_\lambda(G, \mathcal{B}, \mathbb{C})$ denotes the space of integrable functions on the measure space $(G, \mathcal{B}, \lambda)$, where $\mathcal{B}$ is the σ-algebra of Borel sets of G. To simplify notation, from now on *we write $\mathcal{M}^1(G)$ for $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(G)$, and $\mathrm{L}^1(G)$ for $\mathrm{L}^1_\lambda(G, \mathcal{B}, \mathbb{C})$.*

The vector space $\mathcal{M}^1(G)$ is a Banach space with the norm $\|\mu\| = |\mu|(G)$, and $\mathrm{L}^1(G)$ is a Banach space with the L^1-norm. There are three flavors of convolutions but we use mainly two of them:

1. Convolutions of two measures $\mu, \nu \in \mathcal{M}^1(G)$. This makes $\mathcal{M}^1(G)$ into a Banach algebra with identity and with an involution.

2. Convolution of two functions $f, g \in \mathrm{L}^1(G)$, which makes $\mathrm{L}^1(G)$ into a Banach algebra with involution, but without a multiplicative unit element, unless G is discrete.

3. There is also a notion of convolution of a measure $\mu \in \mathcal{M}^1(G)$ and of a function $f \in \mathrm{L}^1(G)$, and of a function $f \in \mathrm{L}^1(G)$ and a measure $\mu \in \mathcal{M}^1(G)$.

Convolution applied to functions can be used as a regularization (or filtering) process. We begin with the convolution of measures.

Let $\mu, \nu \in \mathcal{M}^1(G)$ be two complex measures; then for any function $f \in \mathcal{C}_0(G, \mathbb{C})$ (recall that $\mathcal{C}_0(G, \mathbb{C})$ is the space of continuous functions that tend to zero at infinity), we have a linear functional $\Phi\colon \mathcal{C}_0(G, \mathbb{C}) \to \mathbb{C}$ given by

$$\Phi(f) = \int\int f(st)\, d\mu(s)\, d\nu(t), \quad f \in \mathcal{C}_0(G; \mathbb{C}).$$

Observe that

$$|\Phi(f)| \leq \|f\|_\infty \|\mu\| \|\nu\|,$$

so Φ is a bounded linear functional. By Radon–Riesz III (Theorem 7.30), there is a unique measure $\mu * \nu \in \mathcal{M}^1(G)$ such that

$$\Phi(f) = \int\int f(st)\, d\mu(s)\, d\nu(t) = \int f\, d(\mu * \nu), \quad f \in \mathcal{C}_0(G; \mathbb{C}),$$

with $\|\Phi\| = \|\mu * \nu\|$. Since

$$\|\Phi\| = \sup\{|\Phi(f)| \mid f \in \mathcal{C}_0(G; \mathbb{C}),\ \|f\|_\infty = 1\}$$

and $|\Phi(f)| \leq \|f\|_\infty \|\mu\| \|\nu\|$, we deduce that $\|\mu * \nu\| \leq \|\mu\| \|\nu\|$.

Definition 8.21. Let G be locally compact group. If $\mu, \nu \in \mathcal{M}^1(G)$ are two measures, then the measure $\mu * \nu$, called *convolution* of μ and ν, is the unique measure such that

$$\int f\, d(\mu * \nu) = \int\int f(st)\, d\mu(s)\, d\nu(t) \quad \text{for all } f \in \mathcal{C}_0(G, \mathbb{C}).$$

We have $\|\mu * \nu\| \leq \|\mu\| \|\nu\|$.

Observe that by interchanging s and t, $\iint f(st)\, d\mu(s)\, d\nu(t) = \iint f(ts)\, d\mu(t)\, d\nu(s)$, and by Fubini's theorem, $\iint f(ts)\, d\mu(t)\, d\nu(s) = \iint f(ts)\, d\nu(s)\, d\mu(t)$, so we have

$$\int\int f(st)\, d\mu(s)\, d\nu(t) = \int\int f(ts)\, d\nu(s)\, d\mu(t).$$

Recall the Dirac measure δ_s given by $\delta_s(E) = 1$ iff $s \in E$, and $\delta_s(E) = 0$ otherwise (see Example 4.7).

Proposition 8.44. *Let G be locally compact group. The following properties hold.*

(1) Convolution is associative; that is, if $\mu, \nu, \sigma \in \mathcal{M}^1(G)$, then

$$(\mu * \nu) * \sigma = \mu * (\nu * \sigma).$$

(2) The measure δ_1 (where 1 is the identity element of G) is an identity for convolution; that is,

$$\mu * \delta_1 = \delta_1 * \mu = \mu \quad \textit{for all } \mu \in \mathcal{M}^1(G).$$

We also have

$$\delta_s * \mu = \lambda_s(\mu), \qquad \mu * \delta_s = \rho_{s^{-1}}(\mu).$$

(3) For all $s, t \in G$, we have $\delta_s * \delta_t = \delta_{st}$, and convolution is commutative ($\mu * \nu = \nu * \mu$) if and only if G is abelian.

Most of Proposition 8.44 is proven in Folland [33] (Chapter 2, Section 2.5), and the other parts are proven using Proposition 8.16. See also Dieudonné [24] (Chapter XIV, Section 6).

We need to define $\check{\mu}$ for complex measures. We simply use Definition 8.11.

Definition 8.22. Let G be locally compact group. For any complex measure $\mu \in \mathcal{M}^1(G)$, define $\check{\mu}$ by

$$\check{\mu}(A) = \mu(A^{-1}), \qquad \text{for all } A \in \mathcal{B}(G).$$

We also set $\mu^* = \overline{\check{\mu}}$ and call it the *adjoint* of μ.

The complex measure $\check{\mu}$ can be characterized by a property similar to the property of Proposition 8.27. Recall that the complex measure μ has a unique Jordan decomposition

$$\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-),$$

where the measures $\mu_1^+, \mu_1^-, \mu_2^+$, and μ_2^- are positive measures; see Theorem 7.22.

Proposition 8.45. *Let G be locally compact group. For any complex measure $\mu \in \mathcal{M}^1(G)$, for every function $f \in \mathcal{L}^1_{|\mu|}(X, \mathcal{A}, \mathbb{C})$, we have*

$$\int f d\check{\mu} = \int f(s^{-1}) d\mu(s).$$

Consequently, the complex measure $\check{\mu}$ is the unique measure in $\mathcal{M}^1(G)$ such that

$$\int \varphi d\check{\mu} = \int \varphi(s^{-1}) d\mu(s) \quad \textit{for all } \varphi \in \mathcal{C}_0(G; \mathbb{C}).$$

Proof. We know that μ can be expressed uniquely as

$$\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-),$$

where the measures $\mu_1^+, \mu_1^-, \mu_2^+$, and μ_2^- are positive measures, and we have

$$\check{\mu} = \check{\mu_1^+} - \check{\mu_1^-} + i(\check{\mu_2^+} - \check{\mu_2^-}).$$

Then we have

$$\int f d\check{\mu} = \int f d\check{\mu_1^+} - \int f d\check{\mu_1^-} + i \int f d\check{\mu_2^+} - i \int f d\check{\mu_2^-}.$$

By Proposition 8.27,

$$\int f d\check{\mu_i^+} = \int f(s^{-1}) d\mu_i^+, \qquad \int f d\check{\mu_i^-} = \int f(s^{-1}) d\mu_i^-,$$

so we get

$$\begin{aligned}\int f d\check{\mu} &= \int f d\check{\mu_1^+} - \int f d\check{\mu_1^-} + i \int f d\check{\mu_2^+} - i \int f d\check{\mu_2^-} \\ &= \int f(s^{-1}) d\mu_1^+ - \int f(s^{-1}) d\mu_1^- + i \int f(s^{-1}) d\mu_2^+ - i \int f(s^{-1}) d\mu_2^- \\ &= \int f(s^{-1}) d\mu,\end{aligned}$$

as claimed. The second fact is an immediate consequence of Radon–Riesz III theorem. $\square$

Recall from Proposition 7.24 that $\overline{\mu}$ is the unique measure in $\mathcal{M}^1(G)$ satisfying the equation

$$\int \varphi \, d\overline{\mu} = \overline{\int \overline{\varphi(s)} \, d\mu(s)}, \qquad \text{for all } \varphi \in \mathcal{C}_0(G; \mathbb{C}).$$

Since for any function φ, the function $\check{\varphi}$ is given by $\check{\varphi}(s) = \varphi(s^{-1})$ for all $s \in G$, if we define φ^* by $\varphi^*(s) = \overline{\varphi(s^{-1})} = \overline{\check{\varphi}}(s)$, observe that the measure $\mu^* = \overline{\check{\mu}}$ is characterized by the equation

$$\int \varphi \, d\mu^* = \overline{\int \varphi^*(s) \, d\mu(s)}, \qquad \text{for all } \varphi \in \mathcal{C}_0(G; \mathbb{C}). \tag{$*$}$$

The verification of Equation $(*)$ is as follows:

$$\begin{aligned}\int \varphi(s) \, d\mu^*(s) &= \int \varphi(s) \overline{\check{\mu}}(s) \\ &= \overline{\int \overline{\varphi(s)} \, d\check{\mu}(s)}, \qquad \text{by Proposition 7.24} \\ &= \overline{\int \overline{\varphi(s^{-1})} \, d\mu(s)}, \qquad \text{by Proposition 8.45} \\ &= \overline{\int \varphi^*(s) \, d\mu(s)}.\end{aligned}$$

Proposition 8.46. *Let G be locally compact group. For any measures $\mu, \nu \in \mathcal{M}^1(G)$, we have*

$$
\begin{aligned}
(\mu+\nu)^* &= \mu^* + \nu^* \\
(\alpha\mu)^* &= \overline{\alpha}\mu^* \quad (\alpha \in \mathbb{C}) \\
(\overline{\mu * \nu}) &= \overline{\mu} * \overline{\nu} \\
(\mu * \nu)\check{} &= \check{\nu} * \check{\mu} \\
(\mu * \nu)^* &= \nu^* * \mu^* \\
(\mu^*)^* &= \mu \\
\|\mu^*\| &= \|\mu\|\,.
\end{aligned}
$$

Proof. We prove the second, third, and fourth equations, leaving the others as easy exercises. For every $\varphi \in \mathcal{C}_0(G;\mathbb{C})$ we have

$$
\begin{aligned}
\int \varphi\, d(\alpha\mu)^* &= \overline{\int \overline{\varphi(s^{-1})} d(\alpha\mu)(s)}, && \text{by Equation } (*) \\
&= \overline{\alpha}\,\overline{\int \overline{\varphi(s^{-1})} d\mu(s)} \\
&= \overline{\alpha} \int \varphi\, d\mu^*,
\end{aligned}
$$

which, by Radon–Riesz III, shows that $(\alpha\mu)^* = \overline{\alpha}\mu^*$. We also have

$$
\begin{aligned}
\int \varphi\, d(\overline{\mu * \nu}) &= \overline{\int \overline{\varphi}\, d(\mu * \nu)}, && \text{by Proposition 7.24} \\
&= \overline{\int \left(\int \overline{\varphi(st)}\, d\mu(s) \right) d\nu(s)}, && \text{definition of } d(\mu * \nu) \\
&= \overline{\int \left(\overline{\int \varphi(st)\, d\overline{\mu}(s)} \right) d\nu(s)} \\
&= \int \left(\int \varphi(st)\, d\overline{\mu}(s) \right) d\overline{\nu}(s), && \text{two applications of Proposition 7.24} \\
&= \int \varphi\, d(\overline{\mu} * \overline{\nu}),
\end{aligned}
$$

which, by Radon–Riesz III, shows that $\overline{(\mu * \nu)} = \overline{\mu} * \overline{\nu}$. Finally, we have

$$\begin{aligned}
\int \varphi\, d(\mu * \nu)\check{} &= \int \varphi(s^{-1})\, d(\mu * \nu), && \text{by Proposition 8.45} \\
&= \int\int \varphi((st)^{-1})\, d\mu(s)\, d\nu(t), && \text{definition of } d(\mu * \nu) \\
&= \int \left(\int \varphi(t^{-1}s^{-1})\, d\nu(t) \right) d\mu(s), && \text{Fubini's theorem} \\
&= \int \left(\int \varphi(ts^{-1})\, d\check{\nu}(t) \right) d\mu(s) \\
&= \int \left(\int \varphi(ts)\, d\check{\nu}(t) \right) d\check{\mu}(s), && \text{two applications of Proposition 8.45} \\
&= \int \varphi\, d(\check{\nu} * \check{\mu}),
\end{aligned}$$

which, by Radon–Riesz III, shows that $(\mu * \nu)\check{} = \check{\nu} * \check{\mu}$. ☐

The identities of Proposition 8.46 show that $\mathcal{M}^1(G)$ is a *normed algebra with involution*; see Example 9.6. Furthermore, $\mathcal{M}^1(G)$ has an identity element δ_1 such that $\delta_1^* = \delta_1$, and it is complete. The algebra $\mathcal{M}^1(G)$ is called the *measure algebra* of G.

Remark: In general the identity

$$\|\mu^* * \mu\| = \|\mu\|^2$$

fails.

Observe that until now, we had no need for a Haar measure on G. Now we make use of the Haar measure.

8.12 Convolution of Functions

We know from Proposition 7.32 that every $f \in \mathrm{L}^1(G)$ can be viewed as a complex measure $fd\lambda$ in $\mathcal{M}^1(G)$, where $fd\lambda$ is the unique complex regular Borel measure such that

$$\int fg\, d\lambda = \int g\, (fd\lambda) \quad \text{for all } g \in \mathcal{C}_0(G; \mathbb{C}).$$

Thus we can see what happens when we convolve two complex measures of the form $fd\lambda$ and $gd\lambda$, where $f, g \in \mathrm{L}^1(G)$. We need to figure out what

$$\Phi(h) = \int\int h(ts)f(t)g(s)\, d\lambda(s)\, d\lambda(t)$$

is for any $h \in \mathcal{C}_0(G; \mathbb{C})$, where we used the fact that

$$\int\int h(st)\, d\mu(s)\, d\nu(t) = \int\int h(ts)\, d\nu(s)\, d\mu(t),$$

with $\mu = f d\lambda$ and $\nu = g d\lambda$. Using the left-invariance of the Haar measure (changing s to $t^{-1}s$) and Fubini's theorem, we have

$$\begin{aligned}\Phi(h) = \int\int h(ts)f(t)g(s)\, d\lambda(s)\, d\lambda(t) &= \int \left(\int h(ts)f(t)g(s)\, d\lambda(s) \right) d\lambda(t) \\ &= \int \left(\int h(s)f(t)g(t^{-1}s)\, d\lambda(s) \right) d\lambda(t) \\ &= \int \left(\int h(s)f(t)g(t^{-1}s)\, d\lambda(t) \right) d\lambda(s) \\ &= \int \left(\int f(t)g(t^{-1}s)\, d\lambda(t) \right) h(s)\, d\lambda(s).\end{aligned}$$

Fubini's theorem implies that the integral $\int f(t)g(t^{-1}s)\, d\lambda(t)$ is defined for almost all s. We also have

$$\left\| \int f(t)g(t^{-1}s)\, d\lambda(t) \right\|_1 \leq \|f\|_1 \|g\|_1 .$$

Indeed, by Fubini and by left-invariance of the Haar measure (changing s to ts), we have

$$\begin{aligned}\left\| \int f(t)g(t^{-1}s)\, d\lambda(t) \right\|_1 &= \int \left| \int f(t)g(t^{-1}s)\, d\lambda(t) \right| d\lambda(s) \\ &\leq \int\int |f(t)g(t^{-1}s)|\, d\lambda(t)\, d\lambda(s) \\ &= \int |f(t)| \left(\int |g(t^{-1}s)|\, d\lambda(s) \right) d\lambda(t) \\ &= \int |f(t)| \left(\int |g(s)|\, d\lambda(s) \right) d\lambda(t) \\ &= \|f\|_1 \|g\|_1 .\end{aligned}$$

Then by Proposition 7.31, if the function $f * g$ is the function in $\mathrm{L}^1(G)$ given by

$$(f * g)(s) = \int f(t)g(t^{-1}s)\, d\lambda(t),$$

and since $h \in \mathcal{C}_0(G; \mathbb{C})$, by Proposition 7.31, the equation

$$\Phi(h) = \int\int h(ts)f(t)g(s)\, d\lambda(s)\, d\lambda(t) = \int \left(\int f(t)g(t^{-1}s)\, d\lambda(t) \right) h(s)\, d\lambda(s)$$

shows that Φ is a bounded linear functional, and by Theorem 7.30, there is a unique complex measure $m = (f * g)d\lambda$ such that

$$\Phi(h) = \int \left(\int f(t)g(t^{-1}s)\, d\lambda(t) \right) h(s)\, d\lambda(s) = \int (f * g)(s)h(s)\, d\lambda(s) = \int h(s)\, dm(s),$$

which is the convolution of the measures $fd\lambda$ and $gd\lambda$. This suggests defining the convolution of functions as follows.

Definition 8.23. Let G be a locally compact group equipped with a left Haar measure λ. For any two functions $f, g \in \mathrm{L}^1(G)$, the function $f * g$ called the *convolution* of f and g is the function defined for all almost all s by

$$(f * g)(s) = \int f(t)g(t^{-1}s)\, d\lambda(t).$$

It satisfies the inequality $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$.

The following expression for the convolution $f * g$ of two functions f and g may shed some light on what convolution does. Note that

$$\begin{aligned}(f * g)(s) &= \int_G f(t)g(t^{-1}s)\, d\lambda_G(t) = \int_G f(st)g(t^{-1})\, d\lambda_G(t) \\ &= \int_G (\lambda_{s^{-1}} f)(t)\check{g}(t)\, d\lambda_G(t).\end{aligned}$$

Think of $\check{g}$ as the function g flipped about the y-axis, which is what happens when $G = \mathbb{R}$, since in this case $\check{g}(t) = g(-t)$. Also think of $\lambda_{s^{-1}} f$ as the function f shifted along the x-axis by the amount s, which is what happens when $G = \mathbb{R}$, since $(\lambda_{s^{-1}} f)(t) = f(st) = f(t + s)$. Then $\int_G (\lambda_{s^{-1}} f)(t)\check{g}(t)\, d\lambda_G(t)$ is the "area" around s corresponding to overlapping the functions $\lambda_{s^{-1}} f$ and $\check{g}$ for all t. We can think of f as some kind of filter, and $f * g$ is the result of filtering, or smoothing g, using f. If f is tall and narrow and decays quickly near the origin, $(f * g)(s)$ is almost $g(s)$; see Figure 8.20. This is the idea behind the Dirac delta-function. If f is wider, it tends to perform a better smoothing effect.

The following result will be needed later and is easy to prove. It is a version of Proposition 8.46 functions in $\mathrm{L}^1(G)$.

Proposition 8.47. *The involution $\mu \mapsto \mu^*$ on $\mathcal{M}^1(G)$ yields an involution $f \mapsto f^*$ on $\mathrm{L}^1(G)$, with*

$$f^*(s) = \Delta(s^{-1})\overline{f(s^{-1})},$$

where Δ is the modular function of G. If G is unimodular, then we have the simpler formula

$$f^*(s) = \overline{f(s^{-1})}.$$

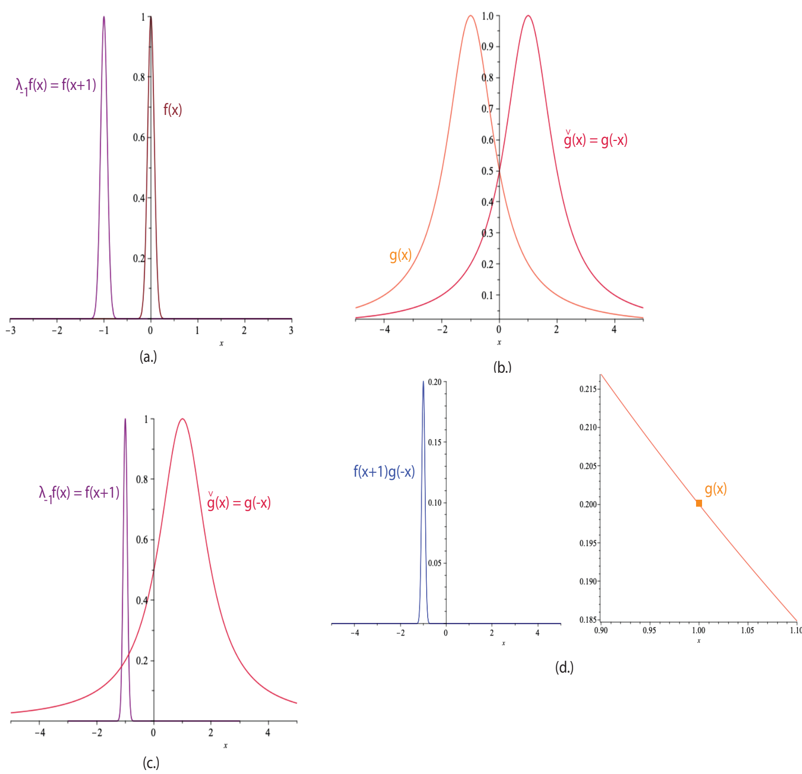


Figure 8.20: Let $G = \mathbb{R}$, $f(x) = \exp(-100x^2)$, and $g(x) = \frac{1}{(x+1)^2+1}$. Figure (a) shows the graphs of $f(x)$ and $\lambda_{-1}f(x)$, while Figure (b) shows the graphs of $g(x)$ and $\check{g}(x)$. Figure (d) shows the graph of the integrand of $(f * g)(1)$ and the value of $g(1)$. The graph of the integrand (in blue) is a narrow peak whose apex has a y-value which is extremely close to the value of $g(1)$, which is denoted by the orange square in the second graph. Hence, the value of $(f * g)(1)$, which is the area under the blue curve, is almost $g(1)$.

Furthermore, for any functions $f, g \in \mathrm{L}^1(G)$, we have

$$\begin{aligned}
(f+g)^* &= f^* + g^* \\
(\alpha f)^* &= \overline{\alpha} f^* \quad (\alpha \in \mathbb{C}) \\
\overline{(f * g)} &= \overline{f} * \overline{g} \\
(f * g)\check{} &= \check{g} * \check{f} \\
(f * g)^* &= g^* * f^* \\
(f^*)^* &= f \\
\|f^*\| &= \|f\| \, .
\end{aligned}$$

It is also easy to see that the convolution $f * g$ of two functions is given by the following

equivalent equations:

$$\begin{aligned} f * g(s) &= \int f(t)g(t^{-1}s)\, d\lambda(t) \\ &= \int f(st)g(t^{-1})\, d\lambda(t) \\ &= \int f(t^{-1})g(ts)\Delta(t^{-1})\, d\lambda(t) \\ &= \int f(st^{-1})g(t)\Delta(t^{-1})\, d\lambda(t). \end{aligned}$$

We go from the first to the second equation using left-invariance by changing t to st. We go from the first to the third equation by changing t to t^{-1} and using the second equation of Proposition 8.27. We go from the second to the fourth equation by changing t to t^{-1} and using the second equation of Proposition 8.27.

Folland gives the following tips to remember how to arrange the variables:

1. The variable of integration t appears as t in one factor and as t^{-1} in the other.

2. The two occurrences of the variable of integration are *adjacent* to each other, not separated by the variable s.

When G is unimodular, the factor $\Delta(t^{-1})$ disappears. If G is abelian, it is customary to use an additive notation, and we have

$$f * g(x) = \int f(y)g(x-y)\, d\lambda(y) = \int f(x-y)g(y)\, d\lambda(y).$$

If $G = \mathbb{R}$, then the function $x \mapsto f(x-y)$ is the function f translated along the x-axis by the amount y (and similarly the function $x \mapsto g(x-y)$ is the function g translated along the x-axis by the amount y). Thus we can think of $f * g(x)$ as a continuous superposition of translates of g, or as a continuous superposition of translates of f. We can interpret these continuous superpositions as moving weighted averages. For example, $f * g(x) = \int f(y)g(x-y)\, d\lambda(y)$ is the weighted average of f (on the whole line) with respect to the weight function $w(y) = g(x-y)$. In particular, if $g(x) = 0$ for all x such that $|x| > a$, then $g(x-y) = 0$ for all x such that $|x-y| > a$, and then $f * g(x)$ is a weighted average of f on the interval $[x-a, x+a]$; see Figure 8.21. In particular, if g is given by

$$g(x) = \begin{cases} \frac{1}{2a} & \text{if } -a < x < a; \\ 0 & \text{if } |x| \geq a, \end{cases} \tag{$*_1$}$$

then

$$f * g(x) = \frac{1}{2a} \int_{x-a}^{x+a} f(y)\, dy.$$

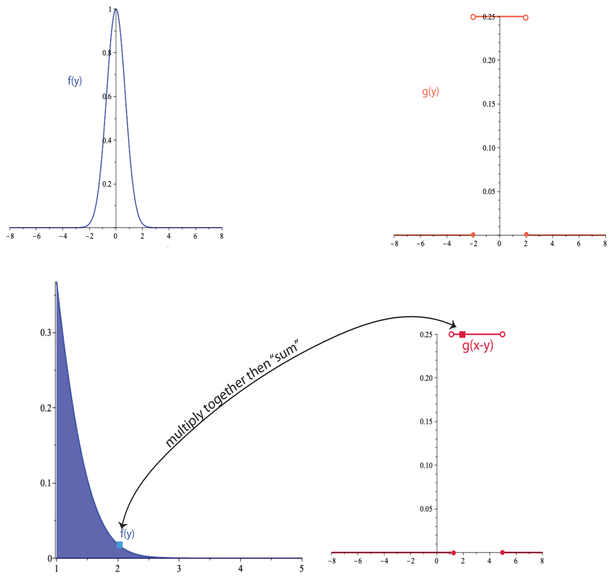


Figure 8.21: Let $f(y) = \exp(-y^2)$, and in the definition of $g(y)$ provided by $(*_1)$ set $a = 2$. The graphs of these two functions are shown in the top row. The graph of $g(x - y)$ (with $x = 3$) is shown in the bottom row. This graph is obtained by reflecting the graph of $g(y)$ over the vertical axis and then shifting this reflection x units to the right. To compute $f * g(x)$, for any y such that $x - a \le y \le x + a$, multiply the values $f(y)g(x - y)$ and take the "infinite" sum of these values. For our particular case, since $f * g(x) = \frac{1}{2a}\int_{x-a}^{x+a} f(y)\, dy$, this "weighted" sum procedure is the area of the shaded blue region in the left figure of the bottom row multiplied by a scaling factor of $\frac{1}{2a}$.

If a is very small, then $f * g(x)$ is approximately $f(x)$. This corresponds to letting g be an approximation to the Dirac function.

With its involution operation, $\mathrm{L}^1(G)$ is a normed Banach algebra (with involution), but generally without a multiplicative unit. It is called the *L^1 group algebra of G*. When G is discrete or finite, $\mathrm{L}^1(G)$ is isomorphic to the algebra $\mathbb{C}[G]$ consisting of all *finite* formal linear combinations of the form

$$\sum_{s\in G} a_s s, \quad a_s \in \mathbb{C},$$

where $a_s = 0$ for all but finitely many $s \in G$, with the multiplication given by

$$\left(\sum_{s_1\in G} a_{s_1} s_1\right)\left(\sum_{s_2\in G} b_{s_2} s_2\right) = \sum_{s\in G}\left(\sum_{tu=s} a_t b_u\right) s = \sum_{s\in S}\left(\sum_{t\in G} a_t b_{t^{-1}s}\right) s,$$

where the last expression is the discrete convolution $a * b$ of $a = (a_{s_1})_{s_1\in G}$ and $b = (b_{s_2})_{s_2\in G}$, with

$$(a * b)_s = \sum_{t\in G} a_t b_{t^{-1}s}.$$

In particular, if $G = \mathbb{Z}$ and if $a = \sum_{i=0}^{m} a_i\, i$ and $b = \sum_{j=0}^{n} b_j\, j$, with no negative elements from $\mathbb{Z}$, we have

$$(a * b)_k = \sum_{i=0}^{k} a_i b_{k-i}, \quad 0 \leq k \leq m + n.$$

We can view $a = \sum_{i=0}^{m} a_i\, i$ as the polynomial $a(X) = \sum_{i=0}^{m} a_i X^i$ and $b = \sum_{j=0}^{n} b_j\, j$ as the polynomial $b(X) = \sum_{j=0}^{n} b_j X^j$.

Convolution can be extended from L^1 to the spaces L^2 and L^∞.

Proposition 8.48. *Let G be a locally compact group equipped with a left Haar measure λ. For any $f \in \mathrm{L}^1(G)$ and for any $g \in \mathrm{L}^p(G)$ with $p = 1, 2, \infty$, the following facts hold:*

(1) The integral $\int f(t)g(t^{-1}s)\, d\lambda(t)$ (defining $f * g$) converges absolutely for almost all s, and we have $f * g \in \mathrm{L}^p(G)$ and $\|f * g\|_p \leq \|f\|_1 \|g\|_p$.

(2) If G is unimodular, then (1) holds with $f * g$ replaced by $g * f$.

(3) If G is not unimodular but f has compact support, then $g * f \in \mathrm{L}^p(G)$.

(4) When $p = \infty$, the function $f * g$ is continuous and under the conditions of (2) or (3), so is $g * f$.

Proposition 8.48 is proven in Folland [33] (Chapter 2, Section 2.5).

Observe that Proposition 8.48 *does not* say anything when $f \in \mathrm{L}^2(G)$. The following proposition takes care of this case.

Proposition 8.49. *Let G be a locally compact group equipped with a left Haar measure λ. For any $f, g \in \mathrm{L}^2(G)$, we have $f * g \in \mathcal{C}_0(G; \mathbb{C})$, and $\|f * g\|_\infty \leq \|f\|_2 \|g\|_2$.*

Proposition 8.49 is proven in Folland [33] (Chapter 2, Section 2.5).

8.13 Convolution of Measures and Functions

One of the applications of convolution is regularization. In order to prove that a complex measure can be approximated by complex measures with compact support we need to define the convolution of a complex measure μ and of a function g. The process of deriving a formula for $\mu * g$ is very similar to the process used to derive a formula for $f * g$, so we proceed quickly.

Let μ be a complex measure in $\mathcal{M}^1(G)$ and let g be a function in $\mathrm{L}^1(G)$. For any $h \in \mathcal{C}_0(G;\mathbb{C})$, we view g as the complex measure $gd\lambda$, and we have

$$\begin{aligned}\Phi(h) = \int\int h(ts)g(s)\,d\lambda(s)\,d\mu(t) &= \int\left(\int h(ts)g(s)\,d\lambda(s)\right)d\mu(t)\\ &= \int\left(\int h(s)g(t^{-1}s)\,d\lambda(s)\right)d\mu(t)\\ &= \int\left(\int h(s)g(t^{-1}s)\,d\mu(t)\right)d\lambda(s)\\ &= \int\left(\int g(t^{-1}s)\,d\mu(t)\right)h(s)\,d\lambda(s).\end{aligned}$$

Fubini's theorem implies that the integral $\int g(t^{-1}s)\,d\mu(t)$ is defined for almost all s. This leads to the following definition.

Definition 8.24. Let G be a locally compact group equipped with a left Haar measure λ. For any complex measure $\mu \in \mathcal{M}^1(G)$ and any function $g \in \mathrm{L}^1(G)$, the function $\mu * g$ called the *convolution function* of μ and g is the function defined for all almost all s by

$$(\mu * g)(s) = \int g(t^{-1}s)\,d\mu(t).$$

It satisfies the inequality $\|\mu * g\|_1 \leq \|\mu\|\,\|g\|_1$. The complex measure $(\mu * g)d\lambda$ is the *convolution* of μ and g. We have

$$\int\int h(ts)g(s)\,d\lambda(s)\,d\mu(t) = \int ((\mu * g)(s))\,h(s)\,d\lambda(s) \qquad \text{for all } h \in \mathcal{C}_0(G;\mathbb{C}).$$

The convolution of a complex measure $\mu \in \mathcal{M}^1(G)$ and any function $g \in \mathrm{L}^2(G)$ is also defined by the same formula, and $\|\mu * g\|_2 \leq \|\mu\|\,\|g\|_2$. Observe that

$$(\delta_s * f)(t) = f(s^{-1}t) = (\lambda_s f)(t),$$

so

$$\delta_s * f = \lambda_s f. \tag{$*_{\lambda_s}$}$$

Similarly, let f be a function in $\mathrm{L}^1(G)$ and let μ be a complex measure in $\mathcal{M}^1(G)$. For any $h \in \mathcal{C}_0(G;\mathbb{C})$, we view f as the complex measure $fd\lambda$, and using Proposition 8.22 and Fubini's theorem, we have

$$\begin{aligned}\Phi(h) = \int\int h(ts)f(t)\,d\lambda(t)\,d\mu(s) &= \int\left(\int h(ts)f(t)\,d\lambda(t)\right)d\mu(s)\\ &= \int \Delta(s^{-1})\left(\int h(t)f(ts^{-1})\,d\lambda(t)\right)d\mu(s)\\ &= \int\left(\int f(ts^{-1})\Delta(s^{-1})\,d\mu(s)\right)h(t)\,d\lambda(t).\end{aligned}$$

Consequently, the convolution $f * \mu$ of a function $f \in \mathrm{L}^1(G)$ and a complex measure $\mu \in \mathcal{M}^1(G)$ is defined as follows.

Definition 8.25. Let G be a locally compact group equipped with a left Haar measure λ. For any function $f \in \mathrm{L}^1(G)$ and any complex measure $\mu \in \mathcal{M}^1(G)$, the function $f * \mu$ called the *convolution function* of f and μ is the function defined for all almost all s by

$$(f * \mu)(s) = \int f(st^{-1})\Delta(t^{-1})\, d\mu(t).$$

The complex measure $(f * \mu)d\lambda$ is the *convolution* of f and μ. The same definition applies if $f \in \mathrm{L}^2(G)$. The inequality $\|f * \mu\|_p \leq \|\mu\| \, \|f\|_p$ holds for $p = 1, 2$.

Observe that

$$(f * \delta_s)(t) = \Delta(s^{-1})f(ts^{-1}) = \Delta(s^{-1})(\rho_{s^{-1}} f)(t),$$

so

$$f * \delta_s = \Delta(s^{-1})\rho_{s^{-1}} f. \tag{$*_{\rho_{s^{-1}}}$}$$

8.14 Regularization

Given a function $g\colon G \to \mathbb{C}$, typically continuous, for some "well chosen" function f, the convolution $f * g$ might be more regular than g, for example, $f * g$ could become a polynomial function, or a sum of trigonometric functions, a C^2 functions, *etc.* In many cases there exists a sequence (f_n) of functions such that each $f_n * g$ is more "regular" than g, and the sequence $(f_n * g)$ converges uniformly to g, at least on every compact subset of G. It is argued in Folland that such sequences always exist; see Folland [33] (Chapter 2, Section 2.5).

Sequences of functions f_n as above can be thought of as approximations of the infamous Dirac δ function[1] in the sense that the sequence of integrals $\int |f_n| \, d\lambda$ is bounded, $\int f_n \, d\lambda$ tends to 1, and that for each open subset V containing 1, the integral $\int_{G-V} |f_n| \, d\lambda$ tends to zero. In fact, Lang calls such sequences *Dirac sequences*; see Lang [62] (Chapter VIII, Section 3).

There are various formulations of the regularization theorem. Here is a version due to Dieudonné; see [24] (Chapter XIV, Section 11).

Proposition 8.50. *Let G be a locally compact group (with identity element e) equipped with a left Haar measure λ. Let (f_n) be a sequence of functions $f_n \in \mathcal{L}^1(G)$ whose supports are contained in a fixed compact subset K and which satisfy the following conditions:*

(1) The sequence of integrals $\int |f_n| \, d\lambda$ is bounded.

[1]The Dirac δ "function" is characterized by the following two properties: (1) $\delta(x) = 0$ for all $x \in \mathbb{R} - \{0\}$, $\delta(0) = +\infty$; (2) $\int \delta(x)dx = 1$. There is no such function. To make sense of it, one has to view δ as a distribution.

(2) The sequence of integrals $\int f_n \, d\lambda$ *tends to* 1.

(3) For each open subset V *containing* e, *the sequence of integrals* $\int_{G-V} |f_n| \, d\lambda$ *tends to zero.*

Then the following properties hold:

(i) For every bounded continuous function g *on* G, *the sequence* $(f_n * g)$ *converges uniformly to* g *on every compact subset of* G.

(ii) If $p = 1$ *or* 2, *and if* $g \in \mathcal{L}^p(G)$, *the sequence of norms* $\|(f_n * g) - g\|_p$ *tends to* 0 *as* n *goes to infinity.*

Proof. We follow Dieudonné's proof [24] (Chapter XIV, Section 11, Theorem 14.11.1).

(i) For every $x \in G$ and for every compact neighborhood V of e, we have from the definitions

$$\begin{aligned} g(x) - (f_n * g)(x) = g(x) &\left(1 - \int_V f_n(s) \, d\lambda(s)\right) \\ &+ \int_V f_n(s)(g(x) - g(s^{-1}x)) \, d\lambda(s) \\ &- \int_{G-V} f_n(s) g(s^{-1}x) \, d\lambda(s). \end{aligned}$$

Next we specialize V. Let L be any compact subset of G, and let V_0 be a compact neighborhood of e. By continuity of the group operations, $V_0^{-1}L$ is a compact, so the restriction of g to $V_0^{-1}L$ is uniformly continuous. Hence, by Definition 8.4, for every $\epsilon > 0$, there is a compact neighborhood $V \subseteq V_0$ such that

$$|g(x) - g(s^{-1}x)| \leq \epsilon \qquad \text{for all } x \in L \text{ and all } s \in V. \tag{$*$}$$

Conditions (2) and (3) imply that we can pick n_0 such that

$$\int_{G-V} |f_n(s)| \, d\lambda(s) \leq \epsilon, \qquad \left|1 - \int f_n(s) \, d\lambda(s)\right| \leq \epsilon, \tag{$**$}$$

for all $n \geq n_0$. Since

$$\int |f_n(s)| \, d\lambda(s) = \int_V |f_n(s)| \, d\lambda(s) + \int_{G-V} |f_n(s)| \, d\lambda(s),$$

we have

$$1 - \int_V |f_n(s)| \, d\lambda(s) = 1 - \int |f_n(s)| \, d\lambda(s) + \int_{G-V} |f_n(s)| \, d\lambda(s),$$

so

$$\left|1 - \int_V f_n(s) \, d\lambda(s)\right| \leq \left|1 - \int f_n(s) \, d\lambda(s)\right| + \int_{G-V} |f_n(s)| \, d\lambda(s) \leq 2\epsilon, \tag{$\dagger$}$$

and therefore, since g is a bounded function, for all $x \in L$, Equations (†), $(**)$, and $(*)$ imply that

$$\begin{aligned}
\left| g(x)\left(1 - \int_V f_n(s)\, d\lambda(s)\right)\right| &\leq \|g\|_\infty \left|1 - \int_V f_n(s)\, d\lambda(s)\right| \leq 2\,\|g\|_\infty\, \epsilon \\
\left|\int_{G-V} f_n(s) g(s^{-1}x)\, d\lambda(s)\right| &\leq \|g\|_\infty\, \epsilon \\
\left|\int_V f_n(s)(g(x) - g(s^{-1}x))\, d\lambda(s)\right| &\leq \int_V |f_n(s)||(g(x) - g(s^{-1}x)|\, d\lambda(s) \\
&\leq \epsilon \int_V |f_n(s)|\, d\lambda(s) \leq c\epsilon,
\end{aligned}$$

where

$$c = \sup_n \int |f_n(s)|\, d\lambda(s).$$

Note that Property (1) guarantees the existence of c. Consequently,

$$|g(x) - (f_n * g)(x)| \leq (c + 3\,\|g\|_\infty)\epsilon,$$

which proves the uniform convergence on the compact L.

(ii) By Theorem 7.10, since $\mathcal{K}_{\mathbb{C}}(G)$ is dense in $\mathcal{L}^p(G)$ for $p = 1, 2$, for every $\epsilon > 0$ there is some $h \in \mathcal{K}_{\mathbb{C}}(G)$ such that $\|g - h\|_p \leq \epsilon$. By Proposition 8.48(1), we have

$$\|(f_n * g) - (f_n * h)\|_p \leq \|f_n\|_1\, \|g - h\|_p \leq c\epsilon,$$

since by Property (1) the $\|f_n\|_1$ are bounded by some constant $c > 0$. Since

$$\begin{aligned}
\|f_n * g - g\|_p &= \|f_n * (g - h + h) - (g - h + h)\|_p \\
&= \|f_n * (g - h) - (g - h) + (f_n * h - h)\|_p \\
&\leq \|f_n * g - f_n * h\|_p + \|g - h\|_p + \|f_n * h - h\|_p \\
&\leq c\epsilon + \epsilon + \|f_n * h - h\|_p\,.
\end{aligned}$$

Therefore we are reduced to proving (ii) for functions in $\mathcal{K}_{\mathbb{C}}(G)$. If $S = \mathrm{supp}(g)$, then it is easy to see that $\mathrm{supp}(f * g) \subseteq KS$, where K is a compact set such that $\mathrm{supp}(f_n) \subseteq K$ for all n, which exists by hypothesis. By (i), the sequence $(f_n * g)$ converges uniformly to g on $S \cup KS$ and vanishes (as does g) outside this set. To conclude that $\lim_{n \mapsto \infty} \|(f_n * g) - g\|_1 = 0$, we use Proposition 5.24(2), which says that

$$\left|\int_A f\, d\lambda\right| \leq \int_A |f|\, d\lambda \leq \|f\|_\infty\, \lambda(A).$$

Here set $f = (f_n * g) - g$ and $A = S \cup KS$, which is a compact subset, so $\lambda(A) < \infty$. By Proposition 8.48(1), if $g \in \mathcal{L}^1(G)$, then $f \in \mathcal{L}^1(G)$. When $g \in \mathcal{L}^2(G)$, we use Proposition 8.48(1) to conclude that $f \in \mathcal{L}^2(G)$. Then $\int_A |f|^2 \, d\lambda$ is well defined with

$$\int_A |f|^2 \, d\lambda \leq \|f\|_\infty^2 \, \lambda(A),$$

which implies that $\lim_{n \mapsto \infty} \|(f_n * g) - g\|_2 = 0$. Alternatively, see Dieudonné [24] (Chapter XIII, Section 12, Theorem 13.12.2.2). $\square$

A sequence (f_n) of functions satisfying the conditions of Proposition 8.50 is called a *regularizing sequence*.

A neat application of Proposition 8.50 is a quick proof of the Weierstrass theorem on the uniform approximation of continuous functions on $[-1/2, 1/2]$ by polynomials.

Example 8.9. Let $G = \mathbb{R}$, with the Lebesgue measure, and let g be a continuous function with support in $[-1/2, 1/2]$. Consider the *Landau functions* f_n given by

$$f_n(x) = \begin{cases} \dfrac{1}{a_n}(1 - x^2)^n & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1, \end{cases}$$

where $a_n = \int_{-1}^{1} (1 - x^2)^n \, dx$; see Figure 8.22.

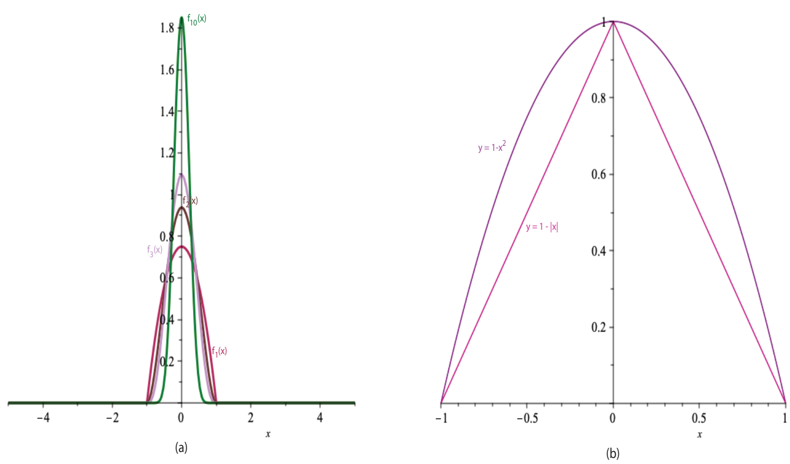


Figure 8.22: Figure (a) shows the graphs of $f_1(x)$, $f_2(x)$, $f_3(x)$, and $f_{10}(x)$. As n increases the "peak" becomes higher and thinner. Figure (b) graphically shows that $1 - x^2 \geq 1 - |x|$ when $1 \leq x \leq -1$.

Since $1 - x^2 \geq 1 - |x|$ for $1 \leq x \leq -1$, we have

$$a_n \geq 2 \int_0^1 (1 - x)^n \, dx = \frac{2}{n+1},$$

which implies that

$$f_n(x) \leq (n+1)(1-x^2)^n$$

for all $x \in [-1, +1]$. Thus $f_n(x)$ tends to 0 uniformly on every compact interval not containing 0, which implies Property (3) of a regularizing sequence. Since by construction $\int |f_n(x)| dx = \int f_n(x) dx = 1$, (f_n) is a regularizing sequence. We have

$$f_n * g(x) = \frac{1}{a_n} \int_{-1/2}^{1/2} f_n(x-y) g(y)\, dy = \frac{1}{a_n} \int_{-1/2}^{1/2} (1-(x-y)^2)^n g(y)\, dy$$

and by using the binomial formula we see that $(1-(x-y)^2)^n$ is a polynomial in y,

$$(1-(x-y)^2)^n = \sum_{j=0}^{2n} u_j(x) y^j,$$

for some polynomials $u_j(x)$ in x, and so

$$\begin{aligned} f_n * g(x) &= \frac{1}{a_n} \int_{-1/2}^{1/2} (1-(x-y)^2)^n g(y)\, dy = \frac{1}{a_n} \int_{-1/2}^{1/2} \sum_{j=0}^{2n} u_j(x) y^j g(y)\, dy \\ &= \frac{1}{a_n} \sum_{j=0}^{2n} u_j(x) \int_{-1/2}^{1/2} y^j g(y)\, dy, \end{aligned}$$

a polynomial in x. Proposition 8.50 shows that the sequence of polynomials $(f_n * g)$ converges uniformly to g on the compact interval $[-1/2, +1/2]$, giving another proof of the Weierstrass approximation theorem.

In this example a continuous function g, which could be much more complicated than a polynomial, and in particular, could lack derivatives of order ≥ 1, becomes a polynomial when convolved with f_n. This is a perfect example of regularization.

Here is another example from Lang [62] (Chapter VIII, Section 3), the *Cesàro summation* of Fourier series of continuous functions on the unit circle $\mathbb{T} = \mathbf{U}(1) = \{e^{i\theta} \mid -\pi \leq \theta < \pi\}$.

8.15 Dirichlet Kernels, Fejér Kernels, Poisson Kernels

Example 8.10. From now on, we will use the normalized Haar measure $dx/2\pi$ on $\mathbb{T}$ so that $\mathbb{T}$ has measure 1. With this normalized measure, the most important results come out cleaner (without an extra factor $1/2\pi$). With this measure, the integral of $f \in \mathrm{L}^1(\mathbb{T})$ is

$$\int_{-\pi}^{\pi} f(\theta) \frac{dx(\theta)}{2\pi},$$

and the convolution of two functions $f, g \in \mathrm{L}^1(\mathbb{T})$ is

$$(f * g)(\theta) = \int_{-\pi}^{\pi} f(\theta - \varphi) g(\varphi) \, \frac{dx(\varphi)}{2\pi} = \int_{-\pi}^{\pi} f(\varphi) g(\theta - \varphi) \, \frac{dx(\varphi)}{2\pi}.$$

Let g be a continuous periodic function (of period 2π) (equivalently, a function on $\mathbb{T} = \mathbf{U}(1)$). The nth partial sum $S_{n,g}$ of the *Fourier series* for g is given by

$$S_{n,g}(x) = \sum_{k=-n}^{n} c_k e^{ikx}, \qquad \text{with} \qquad c_k = \int_{-\pi}^{\pi} g(t) e^{-ikt} \, \frac{dx(t)}{2\pi}.$$

where c_k is called the kth *Fourier coefficient* of g. Let $A_{n,g}$ be the average of these partial sums, that is,

$$A_{n,g} = \frac{1}{n}(S_{0,g} + \cdots + S_{n-1,g}).$$

The average sums $A_{n,g}$ are known as *Cesàro sums* (or *Cesàro means*). Since g is continuous and bounded, $g \in \mathrm{L}^2(\mathbb{T})$, and although the partial sums $S_{n,g}$ converge to g in the $\| \, \|_2$-norm (see Theorem 6.2(3)), they may not converge pointwise to g; see Stein and Shakarchi [94] (Chapter 3, Subsection 2.2), and Rudin [79] (Chapters 4 and 5). On the other hand Fejér's theorem asserts that the sequence $(A_{n,g})$ of average sums converges uniformly to g (see Stein and Shakarchi [94], Chapter 2, Section 5, Theorem 5.2).

This can be shown to be a consequence of Proposition 8.50 by defining the following regularizing functions D_n and K_n:

$$\begin{aligned} D_n(x) &= \sum_{k=-n}^{n} e^{ikx} \\ K_n(x) &= \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=-m}^{m} e^{ikx} = \frac{1}{n}(D_0(x) + \cdots + D_{n-1}(x)). \end{aligned}$$

We leave it as an exercise to prove that

$$\begin{aligned} D_n(x) &= \frac{\sin((2n+1)x/2)}{\sin(x/2)} \\ K_n(x) &= \frac{1}{n} \left(\frac{\sin(nx/2)}{\sin(x/2)} \right)^2. \end{aligned}$$

The functions D_n are known as *Dirichlet kernels*, and the functions K_n are *Fejér kernels*; see Stein and Shakarchi [94] (Chapter 2). The graphs of various $D_n(x)$ and $K_n(x)$ were shown in Figures 6.4, and 6.5 respectively.

The Dirichlet kernels (D_n) do not form a regularizing sequence because they fail to satisfy Property (1) of Proposition 8.50. Indeed, we leave it as an exercise to prove that there is a constant $c > 0$ such that

$$\int_{-\pi}^{\pi} |D_n(x)| \, dx \geq c \log n, \quad \text{as } n \to \infty.$$

However, it is easy to check that (K_n) is a regularizing sequence, that $D_n * g = S_{n,g}$, and that $K_n * g = A_{n,g}$ (see immediately after Proposition 6.1, and Stein and Shakarchi [94], Chapter 2, Section 5, Lemma 5.1). By Proposition 8.50, the sequence $(A_{n,g}) = (K_n * g)$ of averages of the partial sums of the Fourier series of g converge uniformly to g, which is Fejér's theorem (see Stein and Shakarchi [94], Chapter 2, Section 5, Theorem 5.2).

Again, we have a very good example of regularization. After convolving a continuous periodic function g, (which could be much more complicated than a sum of complex exponentials), with D_n, we obtain a function $D_n * g$ which is a sum of complex exponentials.

It is possible to generalize regularizing sequences to families of functions parametrized by a continuous parameter, often called *kernels*.

Example 8.11. As in the previous example, we use the normalized Haar measure $dx/2\pi$ on $\mathbb{T}$ so that $\mathbb{T}$ has measure 1. The *Poisson kernel* on the unit disk is the family of functions $P_r(\theta)$, parametrized by $r \in [0,1)$, and given by

$$P_r(\theta) = \sum_{n=-\infty}^{n=\infty} r^{|n|} e^{in\theta}.$$

To sum this series, using the formula for the sum of a geometric series, observe that

$$\begin{aligned}\sum_{n=-\infty}^{n=\infty} r^{|n|} e^{in\theta} &= \sum_{n=0}^{n=\infty} (re^{i\theta})^n + \sum_{p=1}^{p=\infty} (re^{-i\theta})^p \\ &= \frac{1}{1-re^{i\theta}} + \frac{re^{-i\theta}}{1-re^{-i\theta}} \\ &= \frac{1-r^2}{1-2r\cos\theta + r^2}.\end{aligned}$$

Thus

$$P_r(\theta) = \frac{1-r^2}{1-2r\cos\theta + r^2}.$$

Graphical interpretations of $P_r(\theta)$ were shown in Figures 6.2 and 6.3. Instead of being a sequence of functions indexed by natural numbers, the family (P_r) is a family of functions indexed by the continuous parameter $r \in [0, r)$, but it possesses properties analogous to the properties of regularizing functions, and Proposition 8.50 can be adapted to show that for a bounded periodic continuous function g on $\mathbb{R}$, the functions $(P_r * g)(\theta)$ converge to g uniformly as r tends to 1.

The functions P_r are harmonic for the Laplacian given in polar coordinates by

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2},$$

which means that

$$\Delta P_r = 0.$$

If we write $u(r, \theta) = (P_r * g)(\theta)$, then it can be shown that

$$\Delta u(r, \theta) = 0,$$

that is, $u(r, \theta)$ is harmonic. As a consequence of Proposition 8.50 (suitably generalized), when r tends to 1, we have $u(1, \theta) = g(\theta)$. This shows that $u(r, \theta)$ is a harmonic function solution of a boundary value problem, namely that

$$\Delta u(r, \theta) = 0,$$

with $u(1, \theta) = g(\theta)$ on the boundary, for a prescribed periodic function g.

Other kernels exist for solving partial differential equations, such as the heat equation for the Laplace operator; see Section 6.8. We refer the interested reader to Lang [62] (Chapter VIII, Section 3), Stein and Shakarchi [94], and Folland [32], for more on this topic.

8.16 Regularization of Complex Measures

Regularization can also be used to prove various approximation results involving complex measures. Because we are dealing with locally compact spaces that may not be metrizable, we need the general machinery of filters to define convergence; see Section A.6. We begin with the following general result from Bourbaki [7] (Chapter VIII, Section 2, No. 7, Lemma 4).

Proposition 8.51. *Let X be a locally compact space, $a \in X$ a given point, M a subset of $\mathcal{M}^1(X, \mathcal{A})$, and $\mathcal{F}$ a filter on M. Suppose the following properties hold:*

(1) For every compact subset K of X, the set of numbers $\{|\mu|(K) \mid \mu \in M\}$ is bounded.

(2) For every compact subset K of $X - \{a\}$, we have $\lim_{\mu,\mathcal{F}} |\mu|(K) = 0$.

(3) There is some compact neighborhood V of a such that $\lim_{\mu,\mathcal{F}} |\mu|(V) = 1$.

Then the filter $\mathcal{F}$ converges to the Dirac measure δ_a in $\mathcal{M}^1(X, \mathcal{A})$ (in the norm topology).

Observe that the conditions of Proposition 8.51 are abstact versions of the conditions of Proposition 8.50.

Corollary 8.52. *Let X be a locally compact space, $a \in X$ a given point, M a subset of $\mathcal{M}^1(X, \mathcal{A})$, and $\mathcal{F}$ a filter on M. Suppose the properties of Proposition 8.51 hold and that there is a compact subset K_0 of X such that the complex measures in M have support in K_0. Then the filter $\mathcal{F}$ converges to the Dirac measure δ_a in the space $\mathcal{M}^1_c(X, \mathcal{A})$ of complex measures with compact support (in the norm topology)*

Corollary 8.52 is a more abstract version of Proposition 8.50(i).

Proposition 8.51 can be used to prove the following regularization result for complex measures from Bourbaki [7] (Chapter VIII, Section 4, No. 7, Proposition 19).

Proposition 8.53. *Let G be a locally compact group equipped with a left Haar measure λ. Let $\mathcal{B}$ be a filter basis of neighborhoods of the identity e, consisting of compact neighborhoods, and for every $V \in \mathcal{B}$, let f_V be a positive continuous functions with compact support contained in V, such that $\int f_V \, d\lambda = 1$. For any complex regular Borel measure $\mu \in \mathcal{M}^1(G)$, we have the family of measures $(\mu * f_V)d\lambda$, which has the structure of a filter base on $\mathcal{M}^1(G)$, by considering the subsets $S_V = \{(\mu * f_W)d\lambda \mid W \in \mathcal{B},\ W \subseteq V\}$ (corresponding to the filter of sections of $\mathcal{B}$), and the filter base of subsets S_V converges to μ (in the norm topology).*

Since the measures $(\mu * f_V)d\lambda$ have compact support, Proposition 8.53 shows how the complex regular Borel measure μ can be approximated (in a suitable sense) by measures with compact support (and continuous density). Proposition 8.53 also has the following useful corollary from Bourbaki [7] (Chapter VIII, Section 4, No. 7, Corollary of Proposition 19). A slightly different version of this proposition is given in Folland [33] (Chapter 2, Proposition 2.42).

Proposition 8.54. *Let G be a locally compact group equipped with a left Haar measure λ. Let $\mathcal{B}$ be a filter basis of neighborhoods of the identity e, consisting of compact neighborhoods, and for every $V \in \mathcal{B}$, let f_V be a positive continuous functions with compact support contained in V, such that $\int f_V \, d\lambda = 1$. For any function $g \in \mathrm{L}^p(G)$, $p = 1, 2$, we have the family of functions $(g * f_V)$, which has the structure of a filter base on $\mathrm{L}^p(G)$, by considering the subsets $S_V = \{(g * f_W) \mid W \in \mathcal{B},\ W \subseteq V\}$ (corresponding to the filter of sections of $\mathcal{B}$), and the filter base of subsets S_V converges to g (in the norm topology $\| \, \|_p$).*

Proposition 8.54 is a more abstract version of Proposition 8.50(ii).

In particular, Proposition 8.54 implies that if μ and ν are two complex regular Borel measures, the identity of Definition 8.21 characterzing the convolution $\mu * \nu$ of μ and ν,

$$\int f \, d(\mu * \nu) = \int\!\!\int f(st) \, d\mu(s) \, d\nu(t)$$

holds not only for all functions $f \in \mathcal{C}_0(G, \mathbb{C})$, but also for all *bounded continuous* functions $f \in \mathcal{C}_b(G; \mathbb{C})$. This fact will be needed in Chapter 10.

Families of functions f_V as in Propositions 8.53 and 8.54 are easily constructed using continuous bump functions (since G is locally compact; see Proposition A.39).

Chapter 9

Normed Algebras and Spectral Theory

Let G be a locally compact abelian group. In order to define the notion of Fourier transform on $\mathrm{L}^1(G)$, one needs to figure out what is its domain. The answer is that the domain of the Fourier transform on $\mathrm{L}^1(G)$ is the group $\widehat{G}$ of (unitary) characters of G, the homomorphisms $\chi\colon G \to \mathbb{C}$ such that $|\chi(g)| = 1$ for all $g \in G$. Then one has to give $\widehat{G}$ a topology that makes it into a locally compact group. Doing this is not obvious, but it turns out that as a topological space, $\widehat{G}$ is homeomorphic to the space $\mathsf{X}(\mathrm{L}^1(G))$ of characters of the algebra $\mathrm{L}^1(G)$, the set of algebra homomorphisms $\chi\colon \mathrm{L}^1(G) \to \mathbb{C}$. Here $\mathrm{L}^1(G)$ is the commutative algebra whose multiplication operation is convolution, and it can be shown that $\mathsf{X}(\mathrm{L}^1(G))$ is locally compact.

There is an even deeper connection between the space $\mathsf{X}(\mathrm{L}^1(G))$ of algebra characters and the Fourier transform. Indeed the Fourier cotransform on the commutative algebra $\mathrm{L}^1(G)$ is the Gelfand transform on $\mathrm{L}^1(G)$. For any $f \in \mathrm{L}^1(G)$, the *Gelfand transform* $\mathcal{G}_f$ is a function defined on the set $\mathsf{X}(\mathrm{L}^1(G))$ of characters of $\mathrm{L}^1(G)$ by

$$\mathcal{G}_f(\zeta) = \zeta(f), \qquad \zeta \in \mathsf{X}(\mathrm{L}^1(G)).$$

In general, the algebra $\mathrm{L}^1(G)$ is a complete normed algebra (a Banach algebra), but it does not have a multiplicative unit. However, the space $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(G)$ of complex regular Borel measures on G, simply denoted $\mathcal{M}^1(G)$, is a unital Banach algebra, with the norm $\|\mu\| = |\mu|(G)$ defined in Definition 7.10, with the convolution of measures as multiplication (see Definition 8.21), and with the Dirac measure δ_1 as multiplicative unit. Furthermore, $\mathrm{L}^1(G)$ can be identified with a subalgebra of $\mathcal{M}^1(G)$, using the embedding $f \mapsto f d\lambda$ given by Proposition 7.32.

Therefore we are led to the study of algebras and normed algebras, in particular to complete normed algebras, called Banach algebras. If an algebra A is commutative, then Gelfand had the idea to realize A as a set of complex-valued functions on the set of characters

$\mathsf{X}(A)$ of A. Let $\mathbb{C}^{\mathsf{X}(A)}$ be the set of functions from $\mathsf{X}(A)$ to $\mathbb{C}$. The map $\mathcal{G}\colon A \to \mathbb{C}^{\mathsf{X}(A)}$, called *Gelfand transform*, is defined as follows: for every $a \in A$,

$$\mathcal{G}_a(\chi) = \chi(a), \qquad \chi \in \mathsf{X}(A).$$

If A is a commutative Banach algebra with multiplicative identity element e, then the range of the Gelfand transform $\mathcal{G}_a$, the set $\{\chi(a) \mid \chi \in \mathsf{X}(A)\}$, is equal to the spectrum $\sigma(a)$ of a, namely the set of complex numbers λ such that $\lambda e - a$ is not invertible in A. The spectrum $\sigma(a)$ of a is a generalization of the notion of eigenvalue of a linear map.

The study of algebras and normed algebras focuses on three concepts:

(1) The notion of *spectrum* $\sigma(a)$ of an element a of an algebra A.

(2) If A is a commutative algebra, the notion of *character*, and the space $\mathsf{X}(A)$ of characters of A.

(3) If A is a commutative algebra, the notion of *Gelfand transform*, $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A);\mathbb{C})$.

In Section 9.1 we define algebras and algebras with a multiplicative unit, called unital algebras. We also define normed algebras (without a multiplicative unit), and Cauchy-complete algebras, called Banach algebras. In Section 9.2 we show that every nonunital K-algebra A can be embedded into a unital K-algebra $\widetilde{A}$. We also define the notion of quotient of a normed algebra by an ideal.

If E is an infinite-dimensional vector space, and if $f\colon E \to E$ is a linear map, the definition of an eigenvalue λ of a linear map f used in finite-dimension in terms of the existence of nonzero vector u such that $f(u) = \lambda u$ no longer works because a non-invertible linear map may still be injective. Consequently, there may be some complex number λ such that $\lambda \mathrm{id} - f$ is not invertible, yet there is no nonzero vector u such that $f(u) = \lambda u$. This suggests defining a spectral value as a complex number λ such that $\lambda \mathrm{id} - f$ is not invertible. Then it is an easy step to generalize this definition to any unital algebra A with multiplicative identity e. Given any $a \in A$, viewed as a sort of generalized linear map, a number $\lambda \in \mathbb{C}$ is a *spectral value* for a if $\lambda e - a$ is not invertible.

The notion of *spectrum* is defined and investigated in Section 9.3. The complement $\mathbb{C} - \sigma(a)$ of $\sigma(a)$ is called the *resolvent set* of a, and for any fixed $a \in A$, the function $R(a, \lambda)$ (defined on the set $\mathbb{C} - \sigma(a)$) given by $R(a, \lambda) = (\lambda e - a)^{-1}$ is called the *resolvent* of a. In general there is no guarantee that $\sigma(a)$ is nonempty, but if A is a unital Banach algebra, then $\sigma(a) \neq \emptyset$.

Next we define the notion of character of a commutative unital algebra A. This is a homomorphism $\chi\colon A \to \mathbb{C}$. The space of characters on A is denoted by $\mathsf{X}(A)$. A first connection between the notion of spectrum and the notion of character is that for any $a \in A$ and any $\chi \in \mathsf{X}(A)$, we have $\chi(a) \in \sigma(a)$.

The third key notion, the Gelfand transform, is then defined. The map $\mathcal{G}\colon A \to \mathbb{C}^{\mathsf{X}(A)}$, called *Gelfand transform*, is defined as follows: for every $a \in A$,

$$\mathcal{G}_a(\chi) = \chi(a), \qquad \chi \in \mathsf{X}(A).$$

The function $\mathcal{G}_a$ (or $\mathcal{G}(a)$) is called the *Gelfand transform* of a. If we give $\mathsf{X}(A)$ the topology of pointwise convergence, then $\mathcal{G}$ becomes a continuous map from A to the space $\mathcal{C}(\mathsf{X}(A); \mathbb{C})$ of continuous functions on $\mathsf{X}(A)$.

In order to obtain sharper results about spectra and characters, we consider unital Banach algebras in Section 9.5.

Theorem 9.13 states that if A is a unital Banach algebra, then for any $a \in A$, the spectrum $\sigma(a)$ is nonempty, compact, and contained in the closed ball of radius $\|a\|$. Furthermore, the map $\lambda \mapsto R(a, \lambda)$ is holomorphic.

We prove the Gelfand–Mazur theorem (Theorem 9.14), which says that if a unital Banach algebra A is a (possibly noncommutative) field, then A is isometrically isomorphic to $\mathbb{C}$.

Certain notions defined for complex matrices can be generalized to algebras and normed algebras. In particular, if A is a normed algebra, for any $a \in A$, the number $\rho(a) = \inf_n \|a^n\|^{1/n}$ converges and is called the *spectral radius* of a. Then if A is a unital Banach algebra, we have

$$\rho(a) = \sup\{|\lambda| \mid \lambda \in \sigma(a)\}.$$

If A is a unital Banach algebra, then $\mathsf{X}(A)$ is compact and Hausdorff. If A is a nonunital Banach algebra, then $\mathsf{X}(A)$ is locally compact.

Properties of the Gelfand transform holding for unital Banach algebras are proven in Section 9.7.

Let A be a commutative unital Banach algebra.

(1) For every $a \in A$, the range of $\mathcal{G}_a$ is equal to the spectrum $\sigma(a)$ of a; that is,

$$\mathcal{G}_a(\mathsf{X}(A)) = \{\chi(a) \mid \chi \in \mathsf{X}(A)\} = \sigma(a).$$

(2) The Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A); \mathbb{C})$ is a continuous homomorphism such that $\|\mathcal{G}_a\|_\infty = \rho(a) \leq \|a\|$, and $\mathcal{G}_e$ is the constant function 1.

(3) An element $a \in A$ is invertible iff $\mathcal{G}_a$ does not vanish on $\mathsf{X}(A)$.

If A is a commutative nonunital Banach algebra, then the Gelfand transform is a homomorphism $\mathcal{G}\colon A \to \mathcal{C}_0(\mathsf{X}(A); \mathbb{C})$.

We can characterize when the Gelfand transform is an isometry and when it is injective in terms of its radical. Given a commutative unital algebra A, the *radical* of A, $\operatorname{rad} A$, is the intersection of all maximal ideals in A.

Let A be a commutative unital Banach algebra. We have $\mathrm{Ker}\,\mathcal{G} = \mathrm{rad}\,A$, and the Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A);\mathbb{C})$ is injective iff the radical of A is trivial; that is, $\mathrm{rad}\,A = (0)$.

In Section 9.8, we consider special algebras equipped with an involution, in particular, C^*-algebras.

If A is an algebra an involution is a bijection $a \mapsto a^*$ that satisfies the equations of the conjugate-transpose $A^* = \overline{(A^\top)}$ on complex matrices.

If A is a normed algebra, then an *involutive normed algebra* is an algebra with an involution $a \mapsto a^*$ satisfying the extra axiom

$$\|a\| = \|a^*\|, \qquad \text{for all } a \in A. \tag{i}$$

A C^**-algebra* is a Banach algebra satisfying the axiom

$$\|a\|^2 = \|a^*a\|, \qquad \text{for all } a \in A. \tag{C^*}$$

A C^*-algebra automatically satisfies Axiom (i). The normed algebras $\mathcal{M}^1(G)$ and $\mathrm{L}^1(G)$ are involutive algebras, but in general, not C^*-algebras. The main example of a C^*-algebra is the algebra $\mathcal{L}(H)$ of continuous linear maps on a complex Hilbert space H.

If A is an involutive algebra, an element $a \in A$ is *hermitian* if $a = a^*$, *normal* if $aa^* = a^*a$, *unitary* if $aa^* = a^*a = e$. As in the case of matrices, if A is a unital C^*-algebra, for every $a \in A$, if a is hermitian, then $\sigma(a) \subseteq \mathbb{R}$, and if a is unitary then $\sigma(a) \subseteq \mathbb{T} = \mathbf{U}(1)$.

In Section 9.9 we consider characters and the Gelfand transform in a C^*-algebra. Let A be a commutative unital C^*-algebra. Then for any character $\chi \in \mathsf{X}(A)$, we have

$$\chi(a^*) = \overline{\chi(a)}, \qquad \text{for all } a \in A,$$

or equivalenty $\chi(a) = \overline{\chi(a^*)}$. We say that the characters of A are *hermitian*.

The main theorem of the theory of commutative unital C^*-algebras, due to Gelfand and Naimark, states that every commutative unital C^*-algebra can be viewed as the algebra of continuous functions on a compact space, namely its space of characters $\mathsf{X}(A)$.

More precisely, let A be a commutative unital C^*-algebra. Then the Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A);\mathbb{C})$ is an isometric isomorphism between A and $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$ (and so $\|\mathcal{G}_a\|_\infty = \|a\| = \rho(a)$ for all $a \in A$). Furthermore the Gelfand maps $\mathcal{G}_a$ are hermitian.

The Gelfand–Naimark theorem is used to prove the Plancherel–Godement theorem (see Section 11.8, Theorem 11.41), and some representation theory results in harmonic analysis; see Dieudonné [22].

The spectral theory of C^*-algebras is the key machinery used to develop generalizations of the spectral theorems for normal matrices to bounded (and unbounded) operators of various kinds on a Hilbert space. A condensed presentation of these spectral theorems is given in

Folland [33] (Chapter 1, Section 1.4). An extensive treatment of these spectral theorems is given in Rudin [80], and in Lax [64].

In Section 9.10, given an involutive Banach algebra A, we construct a C^*-algebra $\mathrm{St}(A)$ and an involutive homomorphism $j\colon A \to \mathrm{St}(A)$ (see Definition 9.19) that satisfies a universal mapping condition with respect to homomorphisms of A into a C^*-algebra. For every involutive homomorphism $\varphi\colon A \to B$ of A into a C^*-algebra B, there is a unique involutive homomorphism $\overline{\varphi}\colon \mathrm{St}(A) \to B$ such that

$$\varphi = \overline{\varphi} \circ j,$$

as shown in the following commutative diagram:

$$\begin{tikzcd} A \arrow[r, "j"] \arrow[dr, "\varphi"'] & \mathrm{St}(A) \arrow[d, "\overline{\varphi}"] \\ & B. \end{tikzcd}$$

It is also possible to characterize the set of characters $\mathsf{X}(\mathrm{St}(A))$ of $\mathrm{St}(A)$. Let A be an involutive Banach algebra. Then the map $\mathsf{X}(j)\colon \mathsf{X}(\mathrm{St}(A)) \to \mathsf{X}(A)$ is a homeomorphism of the set of characters $\mathsf{X}(\mathrm{St}(A))$ onto the subspace H of hermitian characters in $\mathsf{X}(A)$; that is, the characters $\chi\colon A \to \mathbb{C}$ such that $\chi(a) = \overline{\chi(a^*)}$ for all $a \in A$.

The above result applies to the involutive Banach algebra $\mathrm{L}^1(G)$ associated with a locally compact group G. In general, $\mathrm{L}^1(G)$ is not a C^*-algebra. Thus we can form the enveloping C^*-algebra $\mathrm{St}(\mathrm{L}^1(G))$ of $\mathrm{L}^1(G)$, denoted $\mathrm{St}(G)$. Remarkably, the canonical map j is injective. As a consequence, it can be shown that there is a homeomorphism between $\mathsf{X}(\mathrm{St}(G))$ and $\mathsf{X}(\mathrm{L}^1(G))$.

More generally, the material of this chapter (spectra, algebra characters, the Gelfand transform), is covered in Bourbaki [9], Dieudonné [24], Folland [33], Lang [62], Lax [64], Rudin [79, 80], and Schwartz [84], the most complete presentations being Rudin [79, 80] and Bourbaki [9].

9.1 Normed Algebras, Banach Algebras

Before defining normed algebras, let us recall the definition of an algebra over a field K. Since the algebras that we will be dealing with *do not always have a multiplicative unit*, for example $\mathrm{L}^1(G)$ with the convolution as product (where G is a locally compact group equipped with a left Haar measure), we use the following definition. A good reference for unital normed algebras is Rudin [79] (Chapter 18).

Definition 9.1. Given a field K, a K*-algebra* is a K-vector space A together with a bilinear operation $\star\colon A \times A \to A$, called *multiplication*, which is associative; that is,

$$(a \star b) \star c = a \star (b \star c) \qquad \text{for all } a, b, c \in A.$$

A K-algebra A is *unital* if there is a multiplicative identity element $\mathbf{1} \neq 0$ so that

$$\mathbf{1} \star a = a \star \mathbf{1} = a, \qquad \text{for all } a \in A.$$

A K-algebra A is *commutative* if

$$a \star b = b \star a, \qquad \text{for all } a, b \in A.$$

If A is a unital algebra, an element $a \in A$ is *invertible* if there is some $b \in A$ such that $a \star b = b \star a = \mathbf{1}$.

Given two K-algebras A and B, a *K-algebra homomorphism* $h \colon A \to B$ is a linear map such that

$$h(a \star_A b) = h(a) \star_B h(b) \qquad \text{for all } a, b \in A.$$

If A and B are unital, we also require that $h(\mathbf{1}_A) = \mathbf{1}_B$.

For example, the ring $\mathrm{M}_n(K)$ of all $n \times n$ matrices over a field K is a unital K-algebra with multiplicative identity element $\mathbf{1} = I_n$.

There are obvious notions of *subalgebra* and *ideal* of a K-algebra:

Definition 9.2. A *subalgebra* B of a K-algebra A is linear subspace closed under multiplication; that is, $x \star y \in B$ for all $x, y \in B$. If A is unital, we require that $\mathbf{1} \in B$. A *left ideal* $\mathfrak{A} \subseteq A$ is a linear subspace of A such that $x \star a \in \mathfrak{A}$ for all $a \in \mathfrak{A}$ and all $x \in A$. A *right ideal* $\mathfrak{A} \subseteq A$ is a linear subspace of A such that $a \star y \in \mathfrak{A}$ for all $a \in \mathfrak{A}$ and all $y \in A$. An *ideal* (or *two-sided ideal*) $\mathfrak{A} \subseteq A$ is a linear subspace of A that is both a left ideal and a right ideal. A left ideal (right ideal) $\mathfrak{A}$ is a *proper left ideal* (a *proper right ideal*) if $\mathfrak{A} \neq A$. The same definition applies to a two-sided ideal. A left ideal (right ideal, two-sided ideal) $\mathfrak{A}$ is a *maximal left ideal* (*maximal right ideal*, *maximal two-sided ideal*) if it is proper and if there is no proper left ideal (proper right ideal, proper two-sided ideal) $\mathfrak{B}$ such that $\mathfrak{A} \subseteq \mathfrak{B}$ and $\mathfrak{A} \neq \mathfrak{B}$.

If A is nonunital, then an ideal is a subalgebra, but the converse is false in general. If A is unital with identity element $\mathbf{1}$, then an ideal $\mathfrak{A}$ is proper iff $\mathbf{1} \notin \mathfrak{A}$, iff $\mathfrak{A}$ contains no invertible element.

If $\mathfrak{A}$ is a two-sided ideal in a K-algebra A (not necessarily unital or commutative), not only is the quotient $A/\mathfrak{A}$ a K-vector space but it is also a K-algebra. Recall the construction of the multiplication operation. The quotient $A/\mathfrak{A}$ consists of the equivalence classes of the equivalence relation $\equiv$ on A defined by

$$x \equiv y \quad \text{iff} \quad x - y \in \mathfrak{A}, \qquad x, y \in A.$$

Let us denote the equivalence class $x + \mathfrak{A}$ of $x \in A$ by $[x]$. We define a multiplication operation on $A/\mathfrak{A}$ by

$$[x] \star [y] = [x \star y], \qquad x, y \in A.$$

This operation is well defined because if $x' \equiv x$ and $y' \equiv y$, then $x' = x + a_1$ and $y' = y + a_2$ for some $a_1, a_2 \in \mathfrak{A}$, so

$$x' \star y' = (x + a_1) \star (y + a_2) = x \star y + x \star a_2 + a_1 \star y + a_1 \star a_2,$$

and since $\mathfrak{A}$ is a two-sided ideal and $a_1, a_2 \in \mathfrak{A}$, we also have $a = x \star a_2 + a_1 \star y + a_1 \star a_2 \in \mathfrak{A}$, an thus $x' \star y' = x \star y + a$ with $a \in \mathfrak{A}$, that is, $[x' \star y'] = [x \star y]$. The verification that the multiplication axioms of an algebra are satisfied is left as an exercise.

In order to generalize certain results to nonunital algebra, in particular to define the notion of radical, we need the notion of regular ideal, as defined in the Appendix of Bourbaki [10].

Definition 9.3. Let A be an algebra, possibly nonunital. A left ideal $\mathfrak{A}$ is *regular* if there is some $u \in A$ such that $x \star u - x \in \mathfrak{A}$ for all $x \in A$. A right ideal $\mathfrak{A}$ is *regular* if there is some $u \in A$ such that $u \star x - x \in \mathfrak{A}$ for all $x \in A$.

If A is unital, then every left (right) ideal is regular (let $u = \mathbf{1}$). Observe that the element u used in Definition 9.3 is a right identity in $A/\mathfrak{A}$ if $\mathfrak{A}$ is a regular left ideal, and a left identity in $A/\mathfrak{A}$ if $\mathfrak{A}$ is a regular right ideal. The following result can be proven; see Bourbaki [10] (Appendix, Proposition 3).

Proposition 9.1. *Let A be a commutative not necessarily unital algebra. An ideal $\mathfrak{A}$ in A is a maximal regular ideal iff $A/\mathfrak{A}$ is a field.*

Using Zorn's lemma we obtain a generalization of another standard result.

Proposition 9.2. *Let A be an algebra, not necessarily unital or commutative. A regular left ideal (regular right ideal) $\mathfrak{A}$ in A distinct from A is contained in a maximal regular left ideal (maximal regular right ideal).*

From now on we assume that the field K is the field $\mathbb{C}$ of complex numbers.

Definition 9.4. A *normed algebra* is a $\mathbb{C}$-algebra A endowed with a norm $\| \; \| \colon A \to \mathbb{R}_+$ satisfying the inequality

$$\|x \star y\| \leq \|x\| \, \|y\| \qquad \text{for all } x, y \in A.$$

If A is unital with identity $\mathbf{1}$, then we require that

$$\|\mathbf{1}\| = 1.$$

If the underlying normed vector space of A is complete, then we say that A is a *Banach algebra*.

The inequality $\|x \star y\| \leq \|x\| \, \|y\|$ shows that the multiplication operation $\star$ is *continuous* (see Proposition A.68). It is a generalization of the inequality characterizing matrix norms. Intuitively, a normed algebra can be viewed as a sort of generalized space of matrices.

For simplicity of notation we write xy instead of $x \star y$, unless confusion arises. *We also use the notation e for the multiplicative identity of A instead of* $\mathbf{1}$.

Example 9.1.

(1) If E is a normed vector space, then the space $\mathcal{L}(E)$ of continuous linear maps $f\colon E \to E$ is a unital algebra under composition, with id_E as identity element. It is a normed algebra under the operator norm,

$$\|f\| = \sup\{\|f(x)\| \mid \|x\| = 1\}.$$

If E is a Banach space (that is, a complete normed vector space), then $\mathcal{L}(E)$ is a Banach algebra.

(2) Let X be a topological space. Then the space $\mathcal{C}_b(X;\mathbb{C})$ of bounded continuous functions on X is a commutative unital Banach algebra, with the norm $\| \, \|_\infty$; functions are multiplied pointwise; that is, $(fg)(x) = f(x)g(x)$ of all $x \in X$. The multiplicative identity is the constant function 1. If X is a (Hausdorff) locally compact space, then the space $\mathcal{C}_0(X;\mathbb{C})$ is a commutative Banach algebra and an ideal in $\mathcal{C}_b(X;\mathbb{C})$, but it not unital unless X is compact (recall that Definition 2.16 implies that $\mathcal{C}_0(X;\mathbb{C}) = \mathcal{K}_{\mathbb{C}}(X)$ if X is compact). The space of $\mathcal{K}_{\mathbb{C}}(X)$ of continuous functions with compact support is a commutative normed algebra and an ideal in $\mathcal{C}_b(X;\mathbb{C})$, but it is not complete (and not unital unless X is compact).

(3) Let G be a locally compact group. The space $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(G)$ of complex regular Borel measures on G, simply denoted $\mathcal{M}^1(G)$, is a unital Banach algebra, with the norm $\|\mu\| = |\mu|(G)$ defined in Definition 7.10, with the convolution of measures as multiplication (see Definition 8.21), and with the Dirac measure δ_1 as multiplicative unit (see Section 8.11). This is the most important example of this book.

(4) Let G be a locally compact group equipped with a left Haar measure λ. The space $\mathrm{L}^1(G)$ (with the L^1-norm) can be identified with a subspace of $\mathcal{M}^1(G)$, using the norm-preserving embeding $f \mapsto f d\lambda$ given by Proposition 7.32. The space $\mathrm{L}^1(G)$ is a Banach algebra with the convolution of functions as multiplication, but it is not unital unless G is discrete.

(5) As a special case of (4), let $G = \mathbb{Z}$, in which case $\mathrm{L}^1(G)$ is the set of all sequences $x = (x_m)_{m\in\mathbb{Z}}$ with $x_m \in \mathbb{C}$, such that $\sum_{m\in\mathbb{Z}} |x_m| < \infty$. This space is also denoted $l^1(\mathbb{Z})$. The convolution product $x * y$ of $x = (x_m)$ and $y = (y_m)$ is given by

$$(x * y)_m = \sum_{p\in\mathbb{Z}} x_p y_{m-p},$$

and the norm by $\|x\| = \sum_{m\in\mathbb{Z}} |x_m|$. This is a commutative unital Banach algebra with identity element e_0 such that $e_0(0) = 1$ and $e_0(m) = 0$ for all $m \neq 0$.

Define δ^m by $\delta^m(m) = 1$ and $\delta^m(k) = 0$ if $k \neq m$. It is easy to see that $\delta^0 = e_0$, and $\delta^m * \delta^n = \delta^{m+n}$, so $(\delta^m)^{-1} = \delta^{-m}$. Then we see that for any $x = (x_m)_{m\in\mathbb{Z}} \in l^1(\mathbb{Z})$, we have $x = \sum_{m\in\mathbb{Z}} x_m \delta^m$, which shows that $l^1(\mathbb{Z})$ is generated by δ^1 and $\delta^{-1} = (\delta^1)^{-1}$.

(6) For any $c = (c_m) \in l^1(\mathbb{Z})$, let $\varphi_c \colon \mathbb{T} \to \mathbb{C}$ be the function given by

$$\varphi_c(e^{i\theta}) = \sum_{m\in\mathbb{Z}} c_m e^{im\theta}, \qquad \theta \in \mathbb{R}/(2\pi\mathbb{Z}).$$

The series defining φ_c is absolutely and uniformly convergent. Thus we deduce that

$$c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_c(e^{i\theta}) e^{-im\theta} d\theta,$$

the mth Fourier coefficient of φ_c; so the map $\varphi \colon c \mapsto \varphi_c$ is an injection from $l^1(\mathbb{Z})$ into the commutative unital Banach algebra of absolutely and uniformly convergent Fourier series (under pointwise multiplication of functions in $\mathrm{L}^1(\mathbb{T})$). The norm on this space is the norm induced by φ from the norm on $l^1(\mathbb{Z})$. The identity is the constant function 1.

(7) Let G be a compact group equipped with a Haar measure. In this case $\mathrm{L}^2(G) \subseteq \mathrm{L}^1(G)$, and both are Banach algebras under convolution ($\mathrm{L}^2(G)$ with the L^2-norm and $\mathrm{L}^1(G)$ with the L^1-norm). These algebras are nonunital unless G is finite. The Banach algebra $\mathrm{L}^2(G)$ is also a Hilbert space. This yields even more structure on $\mathrm{L}^2(G)$ which turns out to be a *Hilbert algebra*, as discussed later. The significance of this fact is that there is a structure theorem about separable Hilbert algebras (they can be expressed as the Hilbert sum of special kinds of ideals), and this structure theorem is a key ingredient in the proof of the Peter–Weyl theorem, the fundamental theorem about the structure of $\mathrm{L}^2(G)$ and of the irreducible unitary representations of the compact group G.

(8) Let $\mathcal{C}^n([0,1])$ be the algebra of functions $f \colon [0,1] \to \mathbb{C}$ having a continuous derivative $f^{(k)}$ for $k = 1, \ldots, n$, under pointwise addition and multiplication. If we let

$$\|f\| = \sum_{k=0}^{n} \frac{1}{k!} \sup_{0\leq t\leq 1} |f^{(k)}(t)|,$$

then we can check that this is indeed a norm, and that $\|fg\| \leq \|f\| \, \|g\|$. With this norm, $\mathcal{C}^n([0,1])$ is a commutative unital Banach algebra. The identity is the constant function 1.

The following simple proposition is left as an exercise.

Proposition 9.3. *Let A be a normed algebra. The closure of a subalgebra of A is a subalgebra of A, and the closure of any ideal of A is an ideal in A.*

Part (2) of the following proposition implies that the set of invertible elements in a unital Banach algebra is open.

Proposition 9.4. *Let A be a unital Banach algebra.*

(1) For any element $a \in A$, if $\|a\| < 1$, then $e - a$ is invertible, the series $\sum_{n=0} a^n$ converges absolutely,

$$(e - a)^{-1} = \sum_{n=0} a^n = e + a + a^2 + \cdots + a^n + \cdots ,$$

and

$$\left\|(e - a)^{-1} - e - a\right\| \leq \frac{\|a\|^2}{1 - \|a\|}.$$

(2) Let $a \in A$ be an invertible element. Then for any $h \in A$, if $\|h\| < (1/2) \left\|a^{-1}\right\|^{-1}$, then $a + h$ is invertible, and

$$\left\|(a + h)^{-1} - a^{-1} + a^{-1}ha^{-1}\right\| \leq 2 \left\|a^{-1}\right\|^3 \|h\|^2 .$$

Proof. (1) Since $\|xy\| \leq \|x\| \|y\|$, we have $\|a^n\| \leq \|a\|^n$, and so

$$1 + \|a\| + \cdots + \|a^n\| \leq 1 + \|a\| + \cdots + \|a\|^n = \frac{1 - \|a\|^{n+1}}{1 - \|a\|}.$$

Since $\|a\| < 1$, the geometric series converges, and so does the series $(\sum_n \|a^n\|)$. Let

$$S_n = e + a + \cdots + a^n.$$

For all $n \geq m$, we have

$$\|S_n - S_m\| = \left\|a^{m+1} + \cdots + a^n\right\| \leq \left\|a^{m+1}\right\| + \cdots + \|a^n\| .$$

Since the sequence $\sum_{k=0}^{m} \|a^m\|$ converges, it is a Cauchy sequence, thus for every $\epsilon > 0$, we can find $p > 0$ so that for all $m \geq n \geq p$ we have

$$\left\|a^{m+1}\right\| + \cdots + \|a^n\| \leq \epsilon,$$

thus the sequence (S_m) is also a Cauchy sequence. Since A is complete, this sequence has a limit, say b. Since

$$S_n(e - a) = (e - a)S_n = e - a^{n+1},$$

and since $\lim_{n \mapsto 0} a^{n+1} = 0$ (because $\|a^{n+1}\| \leq \|a\|^{n+1}$ and $\|a\| < 1$), by continuity of multiplication, we get

$$b(e - a) = (e - a)b = e,$$

which means that b is the inverse of $e - a$.

Since $b = \sum_{n=0}^{\infty} a^n$, we have

$$\|b - e - a\| = \left\|a^2 + \cdots + a^n + \cdots\right\| \leq \sum_{n=2}^{\infty} \|a\|^n = \frac{\|a\|^2}{1 - \|a\|}.$$

(2) Since a is invertible,

$$a + h = a(e + a^{-1}h),$$

and since $\|h\| < (1/2)\left\|a^{-1}\right\|^{-1}$, we get

$$\left\|a^{-1}h\right\| \leq \left\|a^{-1}\right\| \|h\| \leq \left\|a^{-1}\right\| (1/2) \left\|a^{-1}\right\|^{-1} = 1/2.$$

Applying (1) to $-a^{-1}h$, since $\|-a^{-1}h\| = \|a^{-1}h\| < 1/2$, we deduce that $e+a^{-1}h$ is invertible. Therefore $a + h = a(e + a^{-1}h)$ is invertible. We also have $(a + h)^{-1} = (a(e + a^{-1}h))^{-1} = (e + a^{-1}h)^{-1}a^{-1}$, and so

$$(a+h)^{-1}-a^{-1}+a^{-1}ha^{-1} = ((e+a^{-1}h)^{-1}-e+a^{-1}h)a^{-1} = ((e-(-a^{-1}h))^{-1}-e-(-a^{-1}h))a^{-1},$$

which by (1) and the fact that $\|-a^{-1}h\| < 1/2$ yields

$$\begin{aligned}\left\|(a + h)^{-1} - a^{-1} + a^{-1}ha^{-1}\right\| &= \left\|((e - (-a^{-1}h))^{-1} - e - (-a^{-1}h))a^{-1}\right\| \\ &\leq \frac{\|-a^{-1}h\|^2 \left\|a^{-1}\right\|}{1 - \|-a^{-1}h\|} \\ &\leq 2\left\|-a^{-1}h\right\|^2 \left\|a^{-1}\right\| \\ &\leq 2\left\|a^{-1}\right\|^3 \|h\|^2,\end{aligned}$$

as claimed. □

As a corollary of Proposition 9.4, we obtain the following result.

Proposition 9.5. *Let A be a unital Banach algebra. The set $G(A)$ of invertible elements of A is open and contains the open ball of center e given by $\{a \in A \mid \|a - e\| < 1\}$. The inversion map $\iota\colon G(A) \to G(A)$ given by $\iota(a) = -a^{-1}$ is differentiable and its derivative is given by $d\iota(a)(h) = -a^{-1}ha^{-1}$ for all $a \in G(A)$ and all $h \in A$. Consequently, ι is continuous.*

9.2 Two Algebra Constructions

Let K be an arbitrary field. Every nonunital K-algebra A can be embedded into a unital K-algebra $\widetilde{A}$ as follows.

Definition 9.5. Consider the vector space $\widetilde{A} = K \times A$, and define a multiplication by

$$(\lambda, a)(\mu, b) = (\lambda\mu, \lambda b + \mu a + ab). \qquad (*)$$

It is easily verified that $\widetilde{A}$ with this multiplication is a K-algebra, and that $e = (1, 0)$ is a multiplicative unit. The algebra A is embedded into $\widetilde{A}$ using the map $a \mapsto (0, a)$, and it is immediately verified that A is a left ideal in $\widetilde{A}$ (in fact, a maximal ideal). Since $(\lambda, a)(0, b) = (0, \lambda b + ab)$, and $e = (1, 0)$, no element in A is invertible.

Since $(\lambda, a) = \lambda(1, 0) + (0, a)$, we can write $(\lambda, a) = \lambda e + a$. With this notation the multiplication of (λ, a) and (μ, b) becomes

$$(\lambda e + a)(\mu e + b) = \lambda\mu e + \lambda b + \mu a + ab,$$

which shows that the formula $(*)$ defining multiplication is not so strange after all.

If A already has a multiplicative identity ϵ, we immediately check that $e - \epsilon = (1, -\epsilon)$ has the following properties:

$$(e - \epsilon)^2 = e - \epsilon, \quad (e - \epsilon)A = A(e - \epsilon) = \{(0, 0)\}.$$

Then $K(e-\epsilon)$ is a unital algebra with identity $e-\epsilon$ (in fact, a field). Let $(K(e-\epsilon)) \times A$ be the product algebra of $K(e - \epsilon)$ and A under componentwise addition and multiplication. It is a unital algebra with identity $(e - \epsilon, \epsilon)$. To avoid confusion between the two multiplications, for the rest of this paragraph, let use denote multiplication in $\widetilde{A}$ by $\star$. Define the map $\varphi \colon \widetilde{A} \to (K(e - \epsilon)) \times A$ by

$$\varphi(\lambda, a) = (\lambda(e - \epsilon), \lambda\epsilon + a).$$

We check immediately that φ is a linear map. We have

$$\begin{aligned}
\varphi(\lambda, a)\varphi(\mu, b) &= (\lambda(e - \epsilon), \lambda\epsilon + a)(\mu(e - \epsilon), \mu\epsilon + b) \\
&= (\lambda\mu(e - \epsilon), (\lambda\epsilon + a)(\mu\epsilon + b)) \\
&= (\lambda\mu(e - \epsilon), \lambda\mu\epsilon + \lambda b + \mu a + ab),
\end{aligned}$$

and

$$\begin{aligned}
\varphi((\lambda, a) \star (\mu, b)) &= \varphi(\lambda\mu, \lambda b + \mu a + ab) \\
&= (\lambda\mu(e - \epsilon), \lambda\mu\epsilon + \lambda b + \mu a + ab),
\end{aligned}$$

so

$$\varphi((\lambda, a) \star (\mu, b)) = \varphi(\lambda, a)\varphi(\mu, b),$$

which shows that φ is an algebra homomorphism. The map φ is obviously injective, and it is surjective because

$$\varphi(\lambda, a - \lambda\epsilon) = (\lambda(e - \epsilon), \lambda\epsilon + a - \lambda\epsilon) = (\lambda(e - \epsilon), a).$$

In summary we have the following result.

Proposition 9.6. *If A is a unital K-algebra, then we have an algebra isomorphism $\varphi \colon \widetilde{A} \to (K(e - \epsilon)) \times A$.*

If A is a normed algebra, it is also easy to check that the map

$$\|(\lambda, a)\| = |\lambda| + \|a\|$$

makes $\widetilde{A}$ into a unital normed algebra, and if A is a Banach algebra, then so is $\widetilde{A}$.

Later on we will encounter special kinds of Banach algebras called C^*-algebras. In this case, in order to make $\widetilde{A}$ into a C^*-algebra, a different norm is needed. It turns out that

$$\|(\lambda, a)\| = \sup\{\|\lambda b + ab\| \mid\in A,\ \|b\| \leq 1\}$$

makes $\widetilde{A}$ into a C^*-algebra and agrees with the original norm on A. It is the only norm that does so; see Folland [33] (Chapter 1, Section 4), or Bourbaki [9] (Chapter 1, Section 6, No. 3).

In many concrete cases the general construction of $\widetilde{A}$ is not needed. For example, since $\mathrm{L}^1(G)$ is an algebra embedded in $\mathcal{M}^1(G)$, we can use $\mathrm{L}^1(G) \oplus \mathbb{C}\delta_1$ as the completion of $\mathrm{L}^1(G)$ into a unital algebra.

The following construction will be used in Section 9.5. If A is a normed algebra and if $\mathfrak{A}$ is a (topologically) closed ideal in A, then the quotient $A/\mathfrak{A}$ is an algebra which can be made into a normed algebra as follows. If $\pi\colon A \to A/\mathfrak{A}$ is the quotient map, then for any $a \in A$, let

$$\|\pi(a)\| = \inf\{\|a + z\| \mid z \in \mathfrak{A}\}.$$

Proposition 9.7. *Let A be a normed algebra and let $\mathfrak{A}$ be a closed ideal in A. The map $\| \,\| : A/\mathfrak{A} \to \mathbb{R}_+$ given by*

$$\|\pi(a)\| = \inf\{\|a + z\| \mid z \in \mathfrak{A}\}, \qquad a \in A,$$

is a norm on $A/\mathfrak{A}$. If A is a Banach algebra, then so is $A/\mathfrak{A}$, and if A is unital, then so is $A/\mathfrak{A}$.

Proof. Following Schwartz [84] (Chapter II, Sec. 2) we will prove that the map $\| \,\| : A/\mathfrak{A} \to \mathbb{R}_+$ is a norm if and only if $\mathfrak{A}$ is closed.

First observe that by definition of $\|\pi(a)\|$, we have $\|\pi(a)\| \leq \|a\|$, so π is continuous. Note that $\pi(a) = [a]$, so by definition of the multiplication on $A/\mathfrak{A}$ it is immediately verified that π is a homomorphism. Let us check the triangle inequality and the fact that if $\|\pi(a)\| = 0$, then $a \in \mathfrak{A}$, which means that $\pi(a) = 0$.

By definition of a greatest lower bound, given any $\epsilon > 0$, for any two elements $x, y \in A$, there exist some $u, v \in \mathfrak{A}$ such that

$$\|x + u\| \leq \|\pi(x)\| + \epsilon/2$$
$$\|y + v\| \leq \|\pi(y)\| + \epsilon/2.$$

But $\pi(x+y+u+v) = \pi(x+y)$, so

$$\|\pi(x+y)\| \leq \|x+y+u+v\| \leq \|x+u\| + \|y+v\| \leq \|\pi(x)\| + \|\pi(y)\| + \epsilon.$$

Since ϵ is arbitrary, we get

$$\|\pi(x+y)\| = \|\pi(x) + \pi(y)\| \leq \|\pi(x)\| + \|\pi(y)\|.$$

We leave the verification that

$$\|\lambda\pi(x)\| = |\lambda| \, \|\pi(x)\|$$

as an exercise.

Assume that $\|\pi(a)\| = 0$. We want to prove that $a \in \mathfrak{A}$. For every $n > 0$, there is some $u_n \in \mathfrak{A}$ such that $\|a + u_n\| < \|\pi(a)\| + 1/n = 1/n$. Consequently the sequence $(a + u_n)$ converges to 0, which implies that the sequence (u_n) converges to $-a$. Since $u_n \in \mathfrak{A}$ for all $n \geq 1$ and since $\mathfrak{A}$ is closed, $-a \in \mathfrak{A}$. But $\mathfrak{A}$ is a vector space so $a \in \mathfrak{A}$, that is, $\pi(a) = 0$, as claimed.

We also have

$$\|(x+u)(y+v)\| \leq \|x+u\| \, \|y+v\| \leq (\|\pi(x)\| + \epsilon/2)(\|\pi(y)\| + \epsilon/2).$$

Since $(x+u)(y+v) = xy+xv+uy+uv$ and $u, v \in \mathfrak{A}$, by definition $\|\pi(xy)\| \leq \|(x+u)(y+v)\|$. Since ϵ is arbitrary, since by definition $\|\pi(x)\pi(y)\| = \|\pi(xy)\|$, we deduce that

$$\|\pi(x)\pi(y)\| \leq \|\pi(x)\| \, \|\pi(y)\|.$$

If e is the unit of A, then $\pi(e)$ is the unit of $A/\mathfrak{A}$. By definition $\|\pi(e)\| \leq \|e\| = 1$. We also have

$$\|\pi(e)\| = \left\|(\pi(e))^2\right\| \leq (\|\pi(e)\|)^2,$$

which implies that $\|\pi(e)\| \geq 1$, so $\|\pi(e)\| = 1$.

It remains to prove that if A is complete, then $A/\mathfrak{A}$ is also complete. Let $(\pi(a_n))$ be a Cauchy sequence in $A/\mathfrak{A}$. Taking a subsequence we may assume that

$$\|\pi(a_n) - \pi(a_{n-1})\| < \frac{1}{2^n} \quad \text{for all } n \geq 1.$$

We construct inductively a sequence (x_n) in A such that $\pi(x_n) = \pi(a_n)$ and

$$\|x_n - x_{n-1}\| < \frac{1}{2^n} \quad \text{for all } n \geq 1.$$

Intitially $x_1 = a_1$, and if $x_1, \ldots, x_n$ have been defined, since

$$\|\pi(a_{n+1}) - \pi(a_n)\| = \|\pi(a_{n+1} - a_n)\| < \frac{1}{2^{n+1}}$$

there is some $y \in \mathfrak{A}$ such that

$$\|a_{n+1} - a_n + y\| < \frac{1}{2^{n+1}}.$$

Then we let $x_{n+1} = x_n + a_{n+1} - a_n + y$, which works since

$$\pi(x_n + a_{n+1} - a_n + y) = \pi(x_n) + \pi(a_{n+1}) - \pi(a_n) + \pi(y) = \pi(a_n) + \pi(a_{n+1}) - \pi(a_n) = \pi(a_{n+1}).$$

Using the triangle inequality and the fact that

$$\|x_n - x_{n-1}\| < \frac{1}{2^n} \quad \text{for all } n \geq 1,$$

for all $n, p \geq 1$, we have

$$\begin{aligned}
\|x_{n+p} - x_n\| &\leq \|x_{n+p} - x_{n+p-1}\| + \|x_{n+p-1} - x_{n+p-2}\| + \cdots + \|x_{n+1} - x_n\| \\
&< \frac{1}{2^{n+p}} + \frac{1}{2^{n+p-1}} + \cdots + \frac{1}{2^{n+1}} \\
&= \frac{1}{2^{n+1}} \left(1 + \frac{1}{2} + \cdots \frac{1}{2^{n+p-1}}\right) \\
&< \frac{1}{2^n}.
\end{aligned}$$

Consequently, the sequence (x_n) is a Cauchy sequence in A, and since A is complete, this sequence converges to a limit $a \in A$. But π is continuous so the sequence $(\pi(a_n)) = (\pi(x_n))$ converges to $\pi(a)$, as desired. $\square$

9.3 Spectrum, I; For an Algebra

Let E be a finite-dimensional vector space over the field $\mathbb{C}$, and let $f \colon E \to E$ be a linear map. Recall that some $\lambda \in \mathbb{C}$ is an *eigenvalue* of f if there is some *nonzero* vector $u \in E$, called an *eigenvector* associated with λ, such that

$$f(u) = \lambda u. \tag{$*$}$$

Equation $(*)$ holds iff

$$(\lambda \mathrm{id} - f)(u) = 0$$

iff $\lambda \mathrm{id} - f$ is *not injective*. But since E is finite-dimensional, a linear map is injective iff it is invertible, thus $\lambda \mathrm{id} - f$ is not injective iff $\lambda \mathrm{id} - f$ is not invertible. Therefore, *$\lambda \in \mathbb{C}$ is an eigenvalue for f iff $\lambda \mathrm{id} - f$ is not invertible*, which we call the second definition of an eigenvalue. In turn, $\lambda \mathrm{id} - f$ is not invertible iff $\det(\lambda \mathrm{id} - f) = 0$, thus the eigenvalues of f are precisely the zeros of the polynomial $\det(\lambda \mathrm{id} - f) = 0$, viewed as a polynomial in λ.

If E is infinite-dimensional, the situation is more complicated because an injective linear map may not be surjective, so $\lambda \mathrm{id} - f$ could be noninvertible, yet injective for some λ, in which case there are no nonzero eigenvectors associated with λ. Here is an example illustrating this situation.

Example 9.2. Let H be the Hilbert space $\mathrm{L}^2([0,1])$, and let $E = \mathcal{L}(H)$, the Banach algebra of continuous linear maps from H to H. Let $T\colon H \to H$ be the operator in E given by

$$T(h)(x) = xh(x), \quad \text{for all } h \in H \text{ and all } x \in [0,1].$$

For any $\lambda \in [0,1]$, observe that

$$(\lambda \mathrm{id}_E - T)(h)(x) = \lambda h(x) - xh(x) = (\lambda - x)h(x),$$

so $(\lambda \mathrm{id}_E - T)(h)(\lambda) = 0$. This shows that for every $h \in H$, the function $(\lambda \mathrm{id}_E - T)(h)$ vanishes at λ. Thus $\lambda \mathrm{id}_E - T$ is not surjective, because the constant function 1 belongs to H, but it is not in the range of $\lambda \mathrm{id}_E - T$ since it does not vanish anywhere. Therefore $\lambda \mathrm{id}_E - T$ is not invertible, so λ satisfies the second definition for being an eigenvalue of T. However, there is no nonzero eigenvector $h \in H$ associated with λ. Indeed, such a function $h \in H$ would satisfy the equation

$$(\lambda \mathrm{id}_E - T)(h)(x) = 0 \quad \text{for all } x \in [0,1],$$

that is,

$$(\lambda - x)h(x) = 0 \quad \text{for all } x \in [0,1],$$

which implies that $h(x) = 0$ for all $x \in [0,1]$, except for $x = \lambda$. This function is equal to the zero function almost everywhere, so in $H = \mathrm{L}^2([0,1])$, it is the zero function. This shows that $\lambda \mathrm{id}_E - T$ is injective for all $\lambda \in [0,1]$, but we also showed earlier that $\lambda \mathrm{id}_E - T$ is not surjective.

In summary, the definition of an eigenvalue λ of a linear map f used in finite-dimension in terms of the existence of nonzero vector u such that $f(u) = \lambda u$ no longer works in infinite dimension. However, if we redefine an eigenvalue of T to be a complex number λ such that $\lambda \mathrm{id}_E - T$ is not invertible, then every $\lambda \in [0,1]$ is an eigenvalue of T, even though T has *no eigenvectors*.

Example 9.2 suggests a definition of the notion of eigenvalue for a linear map f defined on an infinite-dimensional space E: it is a number $\lambda \in \mathbb{C}$ such that $\lambda \mathrm{id} - f$ is *not* invertible. From this, it is an easy step to generalize this definition to any unital algebra A. Given any $a \in A$, viewed as a sort of generalized linear map, a number $\lambda \in \mathbb{C}$ is a *spectral value* for a if $\lambda e - a$ is not invertible.

Definition 9.6. Let A be complex unital algebra with multiplicative unit e ($e \neq 0$). For any $a \in A$, the *spectrum* $\sigma(a)$ of a is the set of all $\lambda \in \mathbb{C}$ such that $\lambda e - a$ is not invertible. The complement $\mathbb{C} - \sigma(a)$ of $\sigma(a)$ is called the *resolvent set* of a. For any fixed $a \in A$, the function $R(a, \lambda)$ with values in A defined on the set $\mathbb{C} - \sigma(a)$ and given by

$$R(a, \lambda) = (\lambda e - a)^{-1}$$

is called the *resolvant* of a.

If different algebras are involved, to avoid confusion, for any $a \in A$ we write $\sigma_A(a)$ for $\sigma(a)$. Note that there is no guarantee that the spectrum is nonempty for any $a \in A$. However, if A is a unital Banach algebra, we will see that $\sigma(a)$ is nonempty for all $a \in A$. More is true: each $\sigma(a)$ is compact.

Let us mention some simple properties of the spectrum, most of which are proven in Bourbaki [9] (Chapter 1, Section 1, No. 2). See also Rudin [79] (Chapter 18).

Proposition 9.8. *The following properties of the spectrum hold.*

1. *For all $\lambda \in \mathbb{C}$, we have $\sigma(\lambda e) = \{\lambda\}$, and $\sigma(a + \lambda e) = \sigma(a) + \lambda$.*

2. *An element $a \in A$ is invertible iff $0 \notin \sigma(a)$.*

3. *Let $P(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$ be a polynomial of degree $n \geq 1$, with $\alpha_n \neq 0$. For any $a \in A$, let*
$$P(a) = \alpha_0 e + \alpha_1 a + \cdots + \alpha_n a^n.$$
Then we have
$$P(\sigma(a)) = \sigma(P(a)),$$
with $P(\sigma(a)) = \{P(\lambda) \mid \lambda \in \sigma(a)\}$.

4. *If $a \in A$ is nilpotent (that is, $a^n = 0$ for some $n \in \mathbb{N}$), then $\sigma(a) = \{0\}$.*

5. *Let A and B be two unital algebras, and let $\varphi \colon A \to B$ be a homomorphism between them; recall that $\varphi(e_A) = e_B$. Then for any $a \in A$, we have $\sigma_B(\varphi(a)) \subseteq \sigma_A(a)$.*

Definition 9.7. If A is not a unital algebra, for any $a \in A$ we define the spectrum $\sigma'(a)$ of a as the spectrum of a in $\widetilde{A}$.

Since no element of A is invertible, we always have $0 \in \sigma'(a)$. If A is already a unital algebra with multiplicative identity ϵ, then we saw in the previous section that $\widetilde{A}$ is isomorphic to the product algebra $(K(e - \epsilon)) \times A$. If (A_1, e_1) and (A_2, e_2) are unital algebras, then for any $(a_1, a_2) \in A_1 \times A_2$ we have

$$\sigma((a_1, a_2)) = \sigma(a_1) \cup \sigma(a_2),$$

because $\lambda(e_1, e_2) - (a_1, a_2) = (\lambda e_1 - a_1, \lambda e_2 - a_2)$ is not invertible iff either $\lambda e_1 - a_1$ is not invertible or $\lambda e_2 - a_2$ is not invertible. Therefore, by letting $A_1 = K(e - \epsilon)$ and $A_2 = A$, if A is a unital algebra, we obtain

$$\sigma'(a) = \sigma(a) \cup \{0\}.$$

In general, if A is a unital algebra, then as the following example demonstrates, $\sigma(ab) \neq \sigma(ba)$.

Example 9.3. Let H be a Hilbert space with a countable orthonormal basis $(e_1, e_2, \ldots, e_n, \ldots)$. Let $f\colon H \to H$ and $g\colon H \to H$ be the continuous linear maps defined by

$$\begin{aligned} f(e_n) &= e_{n+1}, && n \geq 1, \\ g(e_{n+1}) &= e_n, && n \geq 1, && g(e_1) = 0. \end{aligned}$$

Then $g \circ f = \mathrm{id}_H$, but $(f \circ g)(e_{n+1}) = e_{n+1}$ and $(f \circ g)(e_1) = 0$, so $\sigma(g \circ f) = \{1\}$, and $\sigma(f \circ g) = \{0, 1\}$.

However, for any algebra (unital or not),

$$\sigma'(ab) = \sigma'(ba).$$

The above equation follows from the following proposition, which generalizes a well-known property of matrices; namely that if A is a $m \times n$ matrix and B is a $n \times m$ matrix, then AB and BA have the same nonzero eigenvalues.

Proposition 9.9. *Let A be a unital algebra. For any two elements $a, b \in A$ and any nonzero scalar $\lambda \in \mathbb{C}$) (actually, any field K), if $ab - \lambda e$ is invertible, then $ba - \lambda e$ is invertible.*

Proof. Let u be the inverse of $ab - \lambda e$. We have

$$e = (ab - \lambda e)u = abu - \lambda u,$$

so $abu = \lambda u + e$, and then

$$\begin{aligned} (ba - \lambda e)(bua - e) &= b(abu)a - ba - \lambda bua + \lambda e \\ &= b(\lambda u + e)a - ba - \lambda bua + \lambda e \\ &= \lambda e. \end{aligned}$$

We also have

$$e = u(ab - \lambda e) = uab - \lambda u,$$

so $uab = \lambda u + e$, and then

$$\begin{aligned} (bua - e)(ba - \lambda e) &= b(uab)a - ba - \lambda bua + \lambda e \\ &= b(\lambda u + e)a - ba - \lambda bua + \lambda e \\ &= \lambda e. \end{aligned}$$

Since $\lambda \neq 0$, the above shows that $\lambda^{-1}(bua - e)$ is the inverse of $ba - \lambda e$. ∎

9.4 Characters, Gelfand Transform, I; For an Algebra

The notion of character of an algebra plays a crucial role in harmonic analysis, as a technical tool to generalize the Fourier transform. Thus we introduce it right away.

Definition 9.8. Let A be a complex, *commutative*, unital algebra with multiplicative identity e. A *character* of A is any algebra homomorphism $\chi\colon A \to \mathbb{C}$. Thus it is a linear form such that

$$\chi(ab) = \chi(a)\chi(b), \qquad \text{for all } a, b \in A,$$

and

$$\chi(e) = 1.$$

The set of characters of A is denoted by $\mathsf{X}(A)$.

Note that even though A is commutative, we do not denote its multiplication by $+$, to avoid confusion with addition in A.

Remark: Definition 9.8 still makes sense if A if a noncommutative unital algebra. In fact, Rudin discusses properties of characters of noncommutative unital algebras in Chapter 10 of Rudin [80] under the name of *complex homomorphisms*. However, certain results no longer hold. The main problem is that if A is noncommutative, although the Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A);\mathbb{C})$ (see Definition 9.11) is an algebra homomorphism, since the algebra $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$ (under pointwise multiplication) is commutative, the Gelfand transform can't be injective.

As in all other sources known to us, we always assume when discussing the Gelfand transform that our algebras are commutative and unital.

Proposition 9.10. *Let A be a (commutative) unital algebra with multiplicative identity e. For any character χ, the condition $\chi(e) = 1$ is equivalent to the condition that χ is not identically 0. If so, χ is surjective.*

Proof. Obviously, if $\chi(e) = 1$, then χ is not identically 0. Conversely, if $\chi(a) \neq 0$ for some $a \in A$ then

$$\chi(a) = \chi(ae) = \chi(a)\chi(e),$$

so $\chi(a)(1 - \chi(e)) = 0$, and since $\chi(a) \neq 0$, then $\chi(e) = 1$. Since χ is linear, $\chi(\lambda e) = \lambda\chi(e) = \lambda$, so χ is surjective. ☐

Proposition 9.11. *Let A be a unital (commutative) algebra. For any character χ, if $a \in A$ is invertible, then $\chi(a) \neq 0$.*

Proof. We have

$$1 = \chi(e) = \chi(aa^{-1}) = \chi(a)\chi(a^{-1}),$$

which implies that $\chi(a) \neq 0$. ☐

Definition 9.9. Let $\varphi\colon A \to B$ be a homomorphism of (commutative) unital algebras. Then φ induces a map $\mathsf{X}(\varphi)\colon \mathsf{X}(B) \to \mathsf{X}(A)$ given by

$$\mathsf{X}(\varphi)(\chi) = \chi \circ \varphi, \quad \chi \in \mathsf{X}(B).$$

It is immediately verified that $\mathsf{X}(\psi \circ \varphi) = \mathsf{X}(\varphi) \circ \mathsf{X}(\psi)$ and $\mathsf{X}(\mathrm{id}_A) = \mathrm{id}_{\mathsf{X}(A)}$.

It is easy to show that if $\varphi\colon A \to B$ is surjective, then $\mathsf{X}(\varphi)$ is a bijection of $\mathsf{X}(B)$ onto the set of characters of A that vanish on $\mathrm{Ker}\,\varphi$.

The following proposition characterizes $\mathsf{X}(A)$ in terms of ideals of A and shows a connection between spectra and characters.

Proposition 9.12. *Let A be a unital commutative algebra.*

(1) If $\mathcal{Y}$ denotes the set of ideals of codimension 1 in A, then the map $\chi \mapsto \mathrm{Ker}\,\chi$ is a bijection between $\mathsf{X}(A)$ and $\mathcal{Y}$.

(2) If $a \in A$ and if $\chi \in \mathsf{X}(A)$, then $\chi(a) \in \sigma(a)$. This property also holds is A is noncommutative.

Proof. (1) Since $\mathbb{C}$ has dimension 1 and a character χ is surjective onto $\mathbb{C}$, the kernel of χ has codimension 1. If $\mathfrak{A} \in \mathcal{Y}$ is an ideal of codimension 1, then $A/\mathfrak{A}$ is an algebra of dimension 1, so it is isomorphic to $\mathbb{C}$. But since $A/\mathfrak{A}$ has an identity element ϵ (see Proposition 9.7), a homomorphism φ from $A/\mathfrak{A}$ to $\mathbb{C}$ is uniquely determined by its value on ϵ, so there is a unique isomorphism from $A/\mathfrak{A}$ to $\mathbb{C}$. If $\pi\colon A \to A/\mathfrak{A}$ is the canonical projection, then the homomorphism $\varphi \circ \pi$ from A to $\mathbb{C}$ is the unique character with kernel $\mathfrak{A}$.

(2) Since χ is a homomorphism, we have

$$\chi(\chi(a)e - a) = \chi(a)\chi(e) - \chi(a) = \chi(a)1 - \chi(a) = 0.$$

This implies that $\chi(a)e - a$ is not invertible, since otherwise by Proposition 9.11 we would have $\chi(\chi(a)e - a) \neq 0$. ∎

Remark: When A is a (commutative) nonunital algebra, a character is still a homomorphism $\chi\colon A \to \mathbb{C}$, but this time, the trivial map with constant value 0 is a character.

Definition 9.10. We denote the set of characters of a nonunital algebra A by $\mathsf{X}'(A)$, and we let $\mathsf{X}(A) = \mathsf{X}'(A) - \{0\}$.

If A is already unital, then a character $\chi \in \mathsf{X}'(A)$ (which is a homomorphism not restricted to map e to 1) is nonzero iff $\chi(e) = 1$, which shows that $\mathsf{X}(A) = \mathsf{X}'(A) - \{0\}$ is equal to the set of characters of the unital algebra A. It is also easy to see that if A is nonunital, then there is a bijection between the set of characters $\mathsf{X}(\widetilde{A})$ and the set of characters $\mathsf{X}'(A)$.

The third fundamental concept in the theory of algebras is due to Gelfand. The idea is to realize a commutative algebra A as a set of complex-valued functions on the set of characters $\mathsf{X}(A)$ of A. Let $\mathbb{C}^{\mathsf{X}(A)}$ be the set of functions from $\mathsf{X}(A)$ to $\mathbb{C}$.

Definition 9.11. Let A be a commutative unital algebra. The map $\mathcal{G}\colon A \to \mathbb{C}^{\mathsf{X}(A)}$, called *Gelfand transform*, is defined as follows: for every $a \in A$,

$$\mathcal{G}_a(\chi) = \chi(a), \qquad \chi \in \mathsf{X}(A).$$

The function $\mathcal{G}_a$ (or $\mathcal{G}(a)$) is called the *Gelfand transform* of a.

If necessary to avoid ambiguities, we write $\mathcal{G}_A$ instead of $\mathcal{G}$. Note that $\mathcal{G}_a(\chi)$ is just *evaluation of* χ *on* a.

The set $\mathbb{C}^{\mathsf{X}(A)}$ of functions from $\mathsf{X}(A)$ to $\mathbb{C}$ is a commutative unital algebra under pointwise multiplication. Observe that map $\mathcal{G}\colon A \to \mathbb{C}^{\mathsf{X}(A)}$ is a homomorphism. Indeed we have

$$\mathcal{G}_{ab}(\chi) = \chi(ab) = \chi(a)\chi(b) = \mathcal{G}_a(\chi)\mathcal{G}_b(\chi),$$

so $\mathcal{G}_{ab} = \mathcal{G}_a\mathcal{G}_b$. We also have $\mathcal{G}_e(\chi) = \chi(e) = 1$, so $\mathcal{G}_e$ is the multiplicative unit in $\mathbb{C}^{\mathsf{X}(A)}$.

Observe that the Gelfand transform is not necessarily injective. For example, if $a \in A$ is a nonzero nilpotent, then by Proposition 9.8(4), $\sigma(a) = \{0\}$, and since by Proposition 9.12(2), $\chi(a) \in \sigma(a)$, we obtain so $\chi(a) = 0$ for all characters χ, which means that $\mathcal{G}_a = 0$, even though $a \neq 0$.

Since $\mathsf{X}(A)$ consists of functions from A to $\mathbb{C}$, we can give it the topology of pointwise convergence; see Definition 2.2. Recall that in this topology, a subset of functions $f\colon A \to \mathbb{C}$ is open if it is the union of subsets U_Ω of functions for which there is a finite subset Ω of A and some open intervals $(-r_a, r_a)$ (with $r_a > 0$) for all $a \in \Omega$, such that $f(a) \in (-r_a, r_a)$ for all $a \in \Omega$ (and $f(a)$ is arbitrary for all $a \in A - \Omega$). The topology of pointwise convergence is also defined in terms of semi-norms; see Example 2.4 in Section 2.7. Because $\mathbb{C}$ is Hausdorff, the topology of pointwise convergence is Hausdorff. It is also easy to see that a sequence (f_n) of functions $f_n\colon A \to \mathbb{C}$ converges to a function f in this topology iff it converges pointwise to f (that is, for every $a \in A$, the sequence $(f_n(a))$ converges to $f(a)$); see Section 2.1 or Folland [34] (Chapter 4, Proposition 4.12). Then, by definition of the topology of pointwise convergence, each Gelfand map $\mathcal{G}_a\colon \chi \mapsto \chi(a)$ is continuous (for every open interval $(-r, r)$ of $\mathbb{C}$, we have $\mathcal{G}_a^{-1}((-r, r)) = \{\chi \in \mathsf{X}(A) \mid \mathcal{G}_a(\chi) = \chi(a) \in (-r, r)\}$, which is open in $\mathsf{X}(A)$, with $\Omega = \{a\}$, by definition of the pointwise topology on $\mathsf{X}(A)$). In fact, we leave it as an exercise to prove that the topology of pointwise convergence on $\mathsf{X}(A)$ is the weakest (coarsest) topology for which the Gelfand maps $\mathcal{G}_a$ are continuous. This topology is also called the *weak topology* on $\mathsf{X}(A)$. In summary, the Gelfand transform $\mathcal{G}$ maps A to the space $\mathcal{C}(\mathsf{X}(A); \mathbb{C})$ of continuous functions on $\mathsf{X}(A)$.

In order to obtain sharper results about spectra and characters, we now consider unital Banach algebras.

9.5 Spectrum, II; For a Unital Banach Algebra

If A is a unital Banach algebra, then we have a more precise characterization of the spectrum of an element $a \in A$. Part (3) of Theorem 9.13 is a nontrivial and deep fact.

Theorem 9.13. *Let A be a unital Banach algebra. The following properties hold.*

(1) *For every $a \in A$, the spectrum $\sigma(a)$ is a compact subset of $\mathbb{C}$ contained in the closed ball of radius $\|a\|$ (thus $|\lambda| \leq \|a\|$ for all $\lambda \in \sigma(a)$).*

(2) *For any fixed $a \in A$, the resolvent*

$$R(a, \lambda) = (\lambda e - a)^{-1}$$

is a holomorphic function $R(a, \lambda)\colon (\mathbb{C} - \sigma(a)) \to A$ (which means that $(d/d\lambda)(R(a, \lambda))$ exists for all $\lambda \in \mathbb{C} - \sigma(a)$), and tends to zero at infinity.

(3) *For every $a \in A$, the spectrum $\sigma(a)$ is nonempty.*

Proof sketch. Theorem 9.13 is proven in Dieudonné [24] (Chapter XV, Section 2), Rudin [80] (Chapter 10, Theorem 10.13), Bourbaki [9] (Chapter 1, Section 2, No. 5), and Folland [33] (Chapter 1, Section 1).

If $|\lambda| > \|a\|$, then by Proposition 9.4(1) the element $e - \lambda^{-1}a$ is invertible since $\|\lambda^{-1}a\| = |\lambda^{-1}| \, \|a\| < 1$. Since $\lambda \neq 0$, the element $\lambda e - a$ is also invertible, which shows that $\lambda \notin \sigma(a)$. Therefore, if $\lambda \in \sigma(a)$, then $|\lambda| \leq \|a\|$.

Define the map g by $g(\lambda) = \lambda e - a$. Then g is continuous and invertible on $\Omega = \mathbb{C} - \sigma(a)$. By Proposition 9.5, $G(A)$ is open, and since g is continuous, we conclude that $\Omega = g^{-1}(G(A))$ is open, where $G(A)$ is the set of invertible elements in A. Therefore $\sigma(a)$ is closed. As a closed subset of the compact ball of radius $\|a\|$, $\sigma(a)$ is compact.

Replace a by $\lambda e - a$ and h by $(\mu - \lambda)e$ in Proposition 9.4(2). If $\lambda \in \Omega$ and μ is close enough to λ we have

$$\left\| R(a, \mu) - R(a, \lambda) + (\mu - \lambda) R(a, \lambda)^2 \right\| \leq 2 \left\| R(a, \lambda) \right\|^3 |\mu - \lambda|^2,$$

which shows that

$$\frac{d}{d\lambda} R(a, \lambda) = -R(a, \lambda)^2.$$

By induction we obtain

$$\frac{d^k}{d\lambda^k} R(a, \lambda) = (-1)^k k! R(a, \lambda)^{k+1}.$$

Therefore $R(a, \lambda)$ is holomorphic on Ω.

For λ such that $|\lambda| > \|a\|$ we know that $e - \lambda^{-1}a$ is invertible, and by Proposition 9.4(1) we get

$$R(a, \lambda) = (\lambda e - a)^{-1} = \lambda^{-1}(e - \lambda^{-1}a)^{-1} = \sum_{n=0}^{\infty} \lambda^{-(n+1)} a^n,$$

and when $|\lambda|$ tends to infinity, since $|\lambda^{-1}|$ tends to zero, the above series tends to 0.

As explained in Rudin [80] (Chapter 3, pages 82-85), Cauchy's theorem and Liouville's theorem generalize to holomorphic functions from an open subset of $\mathbb{C}$ to a Banach space (even a Fréchet space). If $\sigma(a)$ was empty, then $R(a, \lambda)$ would be a holomorphic entire function which is bounded, so by Liouville's theorem it would be constant. Since $R(a, \lambda)$ tends to zero at infinity, this constant would be zero, which is absurd (see also Rudin [80], Chapter 10, Theorem 10.13). $\square$

It is shown in Bourbaki [9] (Chapter 1, Section 2, Corollary 1) that if A is *any* unital normed algebra ($A \neq (0)$), not necessarily complete, then the spectrum $\sigma(a)$ is nonempty for every $a \in A$.

Remark: It is easy to show that

$$R(a, \mu) - R(a, \lambda) = (\lambda - \mu) R(a, \lambda) R(a, \mu)$$

for all $(\lambda, \mu) \in (\mathbb{C} - \sigma(a)) \times (\mathbb{C} - \sigma(a))$. This shows that $R(a, \lambda)$ and $R(a, \mu)$ commute.

As a corollary of Theorem 9.13 we have the following theorem.

Theorem 9.14. *(Gelfand–Mazur) Let A be a unital Banach algebra. If A is a (possibly noncommutative) field, then A is isometrically isomorphic to $\mathbb{C}$.*

Proof. For any $a \in A$, since $\sigma(a)$ is nonempty there is some λ such that $\lambda e - a$ is not invertible. Since A is a field, every nonzero element is invertible, and we must have $\lambda e - a = 0$, so $a = \lambda e$. But if $\lambda_1 \neq \lambda_2$, then at most one of $\lambda_1 e - a$ and $\lambda_2 e - a$ is zero, so there is a unique $\lambda(a)$ such that $a = \lambda(a) e$, and so the map $a \mapsto \lambda(a)$ is an isomorphism. We have $|\lambda(a)| = \|\lambda(a) e\| = \lambda(a)$, so this map is an isometry. $\square$

Proposition 9.15. *Let A be a unital Banach algebra. For any invertible element $a \in A$, if $\|a\| = \|a^{-1}\| = 1$, then $\sigma(a) \subseteq \mathbf{U}(1)$.*

Proof. We know by Theorem 9.13(1) that $\sigma(a) \subseteq \{z \in \mathbb{C} \mid |z| \leq 1\}$, and similarly $\sigma(a^{-1}) \subseteq \{z \in \mathbb{C} \mid |z| \leq 1\}$. However if a is invertible it is easy to show that $\sigma(a^{-1}) = (\sigma(a))^{-1} = \{\lambda^{-1} \mid \lambda \in \sigma(a)\}$ (or see Dieudonné [24] (Chapter XV, Section 2), so $\sigma(a) \subseteq \mathbf{U}(1)$. $\square$

Proposition 9.15 generalizes the fact that the eigenvalues of a unitary matrix belong to $\mathbf{U}(1)$.

We can improve the bound on the radius of the smallest closed disc containing $\sigma(a)$ by introducing the spectral radius of a.

Proposition 9.16. *Let A be a normed algebra. For any $a \in A$, the sequence $(\|a^n\|^{1/n})$ converges and its limit is $\inf_n \|a^n\|^{1/n}$.*

Proposition 9.16 is proven in Dieudonné [24] (Chapter XV, Section 2) and Rudin [80] (Chapter 10, Theorem 10.13).

Definition 9.12. Let A be a normed algebra. For any $a \in A$, the number $\rho(a) = \inf_n \|a^n\|^{1/n}$ is called the *spectral radius* of a.

By definition, $\rho(a) \leq \|a\|$.

Proposition 9.17. *Let A be a unital Banach algebra. For any $a \in A$, the spectral radius $\rho(a)$ of a is equal to the radius of the smallest closed disc containing the spectrum $\sigma(a)$ of a, that is, $\rho(a) = \sup\{|\lambda| \mid \lambda \in \sigma(a)\}$.*

A proof of Proposition 9.17 is given in in Dieudonné [24] (Chapter XV, Section 2), Rudin [80] (Chapter 10, Theorem 10.13), and Folland [33] (Chapter 1, Section 1.1). Proposition 9.17 is a generalization of a well-known fact about the spectral radius of a matrix A, which is the largest modulus of the eigenvalues of A.

Proposition 9.8(3) implies that $\sigma(a^n) = (\sigma(a))^n = \{\lambda^n \mid \lambda \in \sigma(a)\}$, so by Proposition 9.17, we have

$$\rho(a^n) = (\rho(a))^n.$$

If A and B are two unital Banach algebras, and if A is a subalgebra of B, for any element $a \in A$, if a has an inverse a^{-1} in A, then $a^{-1} \in B$ so a is also invertible in B, but a could have an inverse $a^{-1} \in B$ such that $a^{-1} \notin A$. The following proposition addresses this situation.

Proposition 9.18. *Let A and B be two unital Banach algebras, with A a closed subalgebra of B. For any $a \in A$, we have $\sigma_B(a) \subseteq \sigma_A(a)$. Every boundary point of $\sigma_A(a)$ belongs to $\sigma_B(a)$. Hence if $\sigma_A(a)$ has empty interior, then $\sigma_B(a) = \sigma_A(a)$.*

Proof. Since for any $a \in A$, if the element $\lambda e - a$ is not invertible in B then it is not invertible in A, we have $\sigma_B(a) \subseteq \sigma_A(a)$.

For the second statement, we need to show that if λ_0 belongs to the boundary of $\sigma_A(a)$, then $\lambda_0 e - a$ is not invertible in B. Since λ_0 belongs to the boundary of $\sigma_A(a)$, there is a sequence (λ_n) with $\lambda_n \in \mathbb{C} - \sigma_A(a)$ converging to λ_0. For every n, the inverse $(\lambda_n e - a)^{-1}$ exists in A, and thus $(\lambda_n e - a)^{-1} \in B$. If $\lambda_0 \notin \sigma_B(a)$, then $(\lambda_0 e - a)^{-1} \in B$, and by Theorem 9.13, since the resolvent $R(a, \lambda)$ is continuous, the sequence $((\lambda_n e - a)^{-1})$ would converge to $(\lambda_0 e - a)^{-1}$. Since A is closed in B, and since $(\lambda_n e - a)^{-1} \in A$, we would have $(\lambda_0 e - a)^{-1} \in A$, contradicting the hypothesis that $\lambda_0 \in \sigma_A(a)$. $\square$

Let us now turn to the characters of a commutative unital Banach algebra.

9.6 Characters, II; Commutative Unital Banach Algebras

We will show in the next theorem that $\mathsf{X}(A)$ is contained in the unit ball B of the dual A' of A (the space of continuous linear forms on A under the operator norm induced by the

norm on A). Unfortunately, the unit ball B in A' is generally not compact in A' (with the topology induced by the operator norm). However, if we consider a weaker topology on A', namely the topology of pointwise convergence on A', then by the Banach–Alaoglu theorem, B is compact in this topology (see Rudin [80] (Chapter 3, Theorem 3.15, or Folland [34] (Chapter 5, Theorem 5.18)). Since $\mathbb{C}$ is Hausdorff, the topology of pointwise convergence on A' is Hausdorff (see the end of Section 2.1).

It turns out that $\mathsf{X}(A)$ is closed in B for the topology of pointwise convergence, and so $\mathsf{X}(A)$ is compact for the topology of pointwise convergence. This is the reason for dropping the norm topology on $\mathsf{X}(A)$ and adopting the topology of pointwise convergence. For historical reasons this topology is also known under another name.

Definition 9.13. We define the *weak*-topology* on A' as the topology of pointwise convergence on A'.

Theorem 9.19. *Let A be a commutative unital Banach algebra.*

(1) Every character $\chi \in \mathsf{X}(A)$ is a continuous map of norm ≤ 1 (by norm, we mean operator norm).

(2) The space $\mathsf{X}(A)$ is compact (and thus Hausdorff) in the topology of the pointwise convergence on A' restricted to $\mathsf{X}(A)$.

Proof sketch. Part (1) of Theorem 9.19 is easy. From Proposition 9.12, we have $\chi(a) \in \sigma(a)$ for all $\in A$, and by Proposition 9.17, $|\chi(a)| \leq \rho(a) \leq \|a\|$, which implies (1).

Part (2) is proven in Bourbaki [9] (Chapter 1, Section 3, No. 1) and Rudin [80] (Chapter 11, Theorem 11.9). By Part (1), $\mathsf{X}(A)$ is a subset of the closed unit ball B in A'. As we explained earlier, the unit ball B in A' is compact in the weak*-topology on A'. It is not hard to show that $\mathsf{X}(A)$ is closed in A' in the weak*-topology on A', and the restriction of this topology to $\mathsf{X}(A)$ is the topology of pointwise convergence. Therefore, $\mathsf{X}(A)$ being closed in a compact subset is compact in the topology of pointwise convergence. $\square$

Remark: If the commutative unital Banach algebra A is also separable, then $\mathsf{X}(A)$ is metrizable; see Dieudonné [24] (Chapter XV, Section 3, Theorem 15.3.2).

Proposition 9.20. *If A is a nonunital commutative Banach algebra, then $\mathsf{X}'(A)$ is compact, and $\mathsf{X}(A)$ is locally compact (in the topology of the pointwise convergence on A').*

For a proof, see Bourbaki [9] (Chapter 1, Section 3, No. 1) or Folland [33] (Chapter 1, Section 1.3, Theorem 1.30). Actually, the proof is almost the same as before, except that we prove that $\mathsf{X}'(A)$ is closed in B in the weak*-topology on A'. Since $\mathsf{X}(A) = \mathsf{X}'(A) - \{0\}$, the result follows.

Theorem 9.21. *Let A be a nonunital commutative Banach algebra. The map $\chi \mapsto \mathrm{Ker}\,\chi$ is a bijection from $\mathsf{X}(A)$ to the set of maximal regular ideals in A. If A is a unital commutative Banach algebra, then the map $\chi \mapsto \mathrm{Ker}\,\chi$ is a bijection from $\mathsf{X}(A)$ to the set of maximal ideals in A.*

Proof. We only prove the second statement in which A is unital. The case where A is nonunital is dealt with in Bourbaki [9] (Chapter 1, Section 3, Theorem 2). Let $\mathfrak{A}$ be a maximal ideal in A. By Proposition 9.3, this ideal must be closed, since otherwise the closure of $\mathfrak{A}$ would be an ideal strictly containing $\mathfrak{A}$, and such an ideal is proper because $\mathfrak{A}$ is contained in the set of noninvertible elements of A (if $\mathfrak{A}$ contains any invertible element, then $\mathfrak{A} = A$, but a maximal ideal is properly contained in A), which is closed, contradicting the maximality of $\mathfrak{A}$. By Proposition 9.7, the quotient algebra $A/\mathfrak{A}$ is a unital Banach algebra. But since $\mathfrak{A}$ is a maximal ideal, $A/\mathfrak{A}$ is a field. By the Gelfand–Mazur theorem (Theorem 9.14), the Banach algebra $A/\mathfrak{A}$ is isomorphic to $\mathbb{C}$, which implies that $\mathfrak{A}$ has codimension 1. Obviously an ideal of codimension 1 is maximal. By Proposition 9.12, the map $\chi \mapsto \operatorname{Ker}\chi$ is a bijection from $\mathsf{X}(A)$ to the set of maximal ideals in A. ☐

As a corollary of Theorem 9.21 we have the following result proven in Bourbaki [9] (Chapter 1, Section 3, No. 2) and Dieudonné [24] (Chapter 15, Example 15.3.7) showing that every compact space E is realized as the set of characters of some commutative unital Banach algebra of functions. Recall that for every $a \in E$, we have the linear functional (evaluation at a) given by $\delta_a(f) = f(a)$ for all $f \in \mathcal{K}_{\mathbb{C}}(E)$, called (with an abuse of language) a Dirac measure (see Example 7.1).

Proposition 9.22. *Let E be a compact space. The map $a \mapsto \delta_a$ is a homeomorphism from E to the set of characters $\mathsf{X}(\mathcal{C}_{\mathbb{C}}(E))$ of the unital Banach algebra $\mathcal{C}_{\mathbb{C}}(E)$. If E is a locally compact space, then the map $a \mapsto \delta_a$ is a homeomorphism from E to the set of characters $\mathsf{X}(\mathcal{C}_0(E;\mathbb{C}))$ of the unital Banach algebra $\mathcal{C}_0(E;\mathbb{C})$ of continuous functions that tend to zero at infinity.*

Proposition 9.22 implies that if E is compact then the set of characters of $\mathcal{C}_{\mathbb{C}}(E)$ is the set of Dirac measures. Similarly, if E is locally compact then the set of characters of $\mathcal{C}_0(E;\mathbb{C})$ is also the set of Dirac measures. In both cases we have a bijection between the set of characters and E itself. Since the the characters are of the form δ_a for any $a \in E$, the Gelfand transform $\mathcal{G}_f$ of any function $f \in \mathcal{C}_0(E;\mathbb{C})$ is given by

$$\mathcal{G}_f(\delta_a) = \delta_a(f) = f(a).$$

Therefore, if we identify E and $\mathsf{X}(\mathcal{C}_0(E;\mathbb{C}))$ under the homeomorphism $a \mapsto \delta_a$, we see that the Gelfand map becomes the identity.

Using the fact that the unital commutative Banach algebra $l^1(\mathbb{Z})$ of Example 9.1(5) is generated by δ^1 and $(\delta^1)^{-1}$, it is shown in Proposition 9.24 that $\mathsf{X}(l^1(\mathbb{Z})) = \sigma(\delta^1)$.

Example 9.4. Let us show that $\sigma(\delta^1) = \mathbb{T}$. We follow Folland [33] (Chapter 1, Section 1.2). We need to figure out for which $\lambda \in \mathbb{C}$ is $\lambda\delta^0 - \delta^1$ is invertible. Suppose that $a \in l^1(\mathbb{Z})$ is an inverse of $\lambda\delta^0 - \delta^1$, so that

$$(\lambda\delta^0 - \delta^1) * a = \delta^0.$$

We leave it as an exercise to show that

$$[(\lambda\delta^0 - \delta^1) * a]_n = \lambda a_n - a_{n-1}.$$

Consequently, $(\lambda\delta^0 - \delta^1) * a = \delta^0$ iff

$$\lambda a_0 - a_{-1} = 1 \quad \text{and} \quad \lambda a_n - a_{n-1} = 0 \quad \text{for all } n \neq 0.$$

It is easy to solve these equations recursively and we obtain

$$\begin{aligned} a_{-1} &= \lambda a_0 - 1 \\ a_n &= \lambda^{-n} a_0, \quad n \geq 0 \\ a_{-n} &= \lambda^{n-1} a_{-1}, \quad n \geq 2. \end{aligned}$$

In order for a to belong to $l^1(\mathbb{Z})$ the condition $\sum_{n=-\infty}^{\infty} |a_n| < \infty$ must hold, which forces $a_0 = 0$ if $|\lambda| \leq 1$ and $a_{-1} = 0$ if $|\lambda| \geq 1$. These conditions imply that there is a unique inverse $a \in l^1(\mathbb{Z})$ iff $|\lambda| \neq 1$, namely

$$a = \begin{cases} -\sum_{n=1}^{\infty} \lambda^{n-1}\delta^{-n} & \text{if } |\lambda| < 1 \\ \sum_{n=0}^{\infty} \lambda^{-n-1}\delta^{n} & \text{if } |\lambda| > 1. \end{cases}$$

Therefore, $\lambda\delta^0 - \delta^1$ is not invertible iff $|\lambda| = 1$, which shows that $\sigma(\delta^1) = \mathbf{U}(1) = \mathbb{T}$.

If A is the unital commutative Banach algebra of absolutely convergent Fourier series of Example 9.1(6), it can be shown that again $\mathsf{X}(A) = \mathbb{T}$; see Folland [33] (Chapter 1, Section 1.2).

9.7 Gelfand Transform, II; For a Commutative Unital Banach Algebra

If A is a commutative unital algebra, we already know that for each $a \in A$ the Gelfand transform $\mathcal{G}_a$ is a continuous map $\mathcal{G}_a \colon \mathsf{X}(A) \to \mathbb{C}$ (where $\mathsf{X}(A)$ is given the topology of pointwise convergence). If A is a Banach algebra, then we have the following sharper result.

Theorem 9.23. *Let A be a commutative unital Banach algebra.*

(1) For every $a \in A$, the range of $\mathcal{G}_a$ is equal to the spectrum $\sigma(a)$ of a; that is,

$$\mathcal{G}_a(\mathsf{X}(A)) = \{\chi(a) \mid \chi \in \mathsf{X}(A)\} = \sigma(a).$$

(2) The Gelfand transform $\mathcal{G} \colon A \to \mathcal{C}(\mathsf{X}(A); \mathbb{C})$ is a continuous homomorphism such that $\|\mathcal{G}_a\|_\infty = \rho(a) \leq \|a\|$, and $\mathcal{G}_e$ is the constant function 1.

(3) An element $a \in A$ is invertible iff $\mathcal{G}_a$ does not vanish on $\mathsf{X}(A)$.

Proof. (1) We already know from Proposition 9.12 that $\chi(a) \in \sigma(a)$ for every $a \in A$, so we just have to prove that for every $\lambda \in \sigma(a)$, there is some $\chi \in \mathsf{X}(A)$ such that $\chi(a) = \lambda$. If $\lambda \in \sigma(a)$, then $\lambda e - a$ is not invertible. Since $\lambda e - a$ is not invertible, the ideal $A(\lambda e - a)$ generated by $\lambda e - a$ is distinct from A. Using Zorn's lemma, it is a standard argument to show that this ideal is contained in a maximal ideal $\mathfrak{A}$. By Theorem 9.21, there is some character χ such that $\mathfrak{A} = \operatorname{Ker} \chi$, so χ vanishes on $A(\lambda e - a)$, and in particular on $\lambda e - a$, which shows that $\chi(a) = \lambda$.

(2) We already showed that $\mathcal{G}$ is a homomorphism and that $\mathcal{G}_e = 1$. Since by (1) we have $\mathcal{G}_a(\mathsf{X}(A)) = \sigma(a)$ and since by Proposition 9.17, we have $\rho(a) = \sup\{|\lambda| \mid \lambda \in \sigma(a)\}$, we get

$$\|\mathcal{G}_a\|_\infty = \sup |\sigma(a)| = \rho(a).$$

We already know that $\rho(a) \leq \|a\|$, so $\|\mathcal{G}_a\| \leq \|a\|$, which shows that $\mathcal{G}$ is continuous.

(3) We know that a is invertible iff $0 \notin \sigma(a)$, and by (1), this is equivalent to the fact that $\mathcal{G}_a$ does not vanish on $\mathsf{X}(A)$. □

Remark: If A is a commutative nonunital Banach algebra, then the Gelfand transform is a homomorphism $\mathcal{G} \colon A \to \mathcal{C}_0(\mathsf{X}(A); \mathbb{C})$; see Bourbaki [9] (Chapter 1, Section 3, No. 3) or Folland [33] (Chapter 1, Section 1.3, Theorem 1.30).

As a corollary of Theorem 9.23 we have the following result.

Proposition 9.24. *Let A be a commutative unital Banach algebra. For any fixed $a \in A$, the Gelfand transform $\mathcal{G}_a$ is a homeomorphism from $\mathsf{X}(A)$ to $\sigma(a)$ in the following two cases:*

(1) The algebra A is generated by a and e.

(2) The algebra A is generated by a and a^{-1} (assuming that a is invertible).

Proof. By Theorem 9.23, the map $\mathcal{G}_a$ is continuous and surjective onto $\sigma(a)$. Since $\mathsf{X}(A)$ and $\sigma(a)$ are compact Hausdorff spaces, by the corollary to Proposition A.33, it suffices to show that this map is injective. But any character $\chi \colon A \to \mathbb{C}$ is uniquely determined by $\chi(a)$, which is trivial in (1), and in (2) follows from the fact that $\chi(a^{-1}) = (\chi(a))^{-1}$. If $\mathcal{G}_a(\chi_1) = \mathcal{G}_a(\chi_2)$, then $\chi_1(a) = \chi_2(a)$, and since χ_1 is completely determined by $\chi_1(a)$ and similarly χ_2 is completely determined by $\chi_2(a)$, we have $\chi_1 = \chi_2$, and $\mathcal{G}_a$ is injective. □

In particular, since $l^1(\mathbb{Z})$ is generated by δ^1 and $\delta^{-1} = (\delta^1)^{-1}$, Proposition 9.24 shows that $\mathsf{X}(l^1(\mathbb{Z}))$ is homeomorphic to the spectrum of δ^1 (recall Example 9.1(5)). As we said earlier, this spectrum is equal to $\mathbb{T}$, so $\mathsf{X}(l^1(\mathbb{Z})) \cong \mathbb{T}$.

Example 9.5. Proposition 9.24 allows us to figure out what the Gelfand transform on $l^1(\mathbb{Z})$ is. Indeed, for every spectral value $e^{i\theta} \in \mathbb{T}$, there is a unique character χ_θ such that $\chi_\theta(\delta^1) = e^{i\theta}$. Since every $c \in l^1(\mathbb{Z})$ is written uniquely as $c = \sum_{m\in\mathbb{Z}} c_m(\delta^1)^m$, we have

$$\chi_\theta(c) = \sum_{m\in\mathbb{Z}} c_i \chi_\theta((\delta^1)^m) = \sum_{m\in\mathbb{Z}} c_i(\chi_\theta(\delta^1))^m = \sum_{m\in\mathbb{Z}} c_i e^{im\theta},$$

the Fourier series associated with $c = (c_m)_{m\in\mathbb{Z}} \in l^1(\mathbb{Z})$. Since there is a bijection between $\mathsf{X}(l^1(\mathbb{Z}))$ and $\mathbb{T}$, we can identify χ_θ and $e^{i\theta}$, and we see that the Gelfand transform from $l^1(\mathbb{Z})$ to $\mathcal{C}(\mathsf{X}(l^1(\mathbb{Z})))$ is given by

$$\mathcal{G}_c(e^{i\theta}) = \chi_\theta(c) = \sum_{m\in\mathbb{Z}} c_i e^{im\theta}.$$

As a corollary of the above characterization of the Gelfand transform on $l^1(\mathbb{Z})$ and Theorem 9.23 we obtain the following nontrivial theorem of Wiener.

Proposition 9.25. *For any $c = (c_m)_{m\in\mathbb{Z}} \in l^1(\mathbb{Z})$, if the Fourier series $f(e^{i\theta}) = \sum_{m\in\mathbb{Z}} c_i e^{im\theta}$ does not vanish (does not take the value 0 for any $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$), then $1/f$ is given by the absolutely convergent Fourier series $\sum_{m\in\mathbb{Z}} b_m e^{im\theta}$, with $b = c^{-1}$.*

Proof. By Theorem 9.23(3), the element $c \in l^1(\mathbb{Z})$ is invertible iff the Fourier series $f = \mathcal{G}_c$ does not vanish. In this case, $\mathcal{G}_{c^{-1}}$ is the inverse $1/f$ of f. $\square$

The Gelfand transform on $l^1(\mathbb{Z})$ turns out to be the Fourier cotransform on $l^1(\mathbb{Z})$. More generally, if G is a commutative locally compact group equipped with a Haar measure λ, the Gelfand transform can be viewed as the Fourier cotransform on the commutative Banach algebra $\mathrm{L}^1(G)$. For any $f \in \mathrm{L}^1(G)$, the Gelfand transform $\mathcal{G}_f$ is a function defined on the set $\mathsf{X}(\mathrm{L}^1(G))$ of characters of $\mathrm{L}^1(G)$ by

$$\mathcal{G}_f(\zeta) = \zeta(f), \qquad \zeta \in \mathsf{X}(\mathrm{L}^1(G)).$$

However, it turns out that there is a homeomorphism between $\mathsf{X}(\mathrm{L}^1(G))$ and the dual group $\widehat{G}$ of G, which is the group of continuous homomorphisms $\chi\colon G \to \mathbf{U}(1)$; for example, see Theorem 10.6 or Folland [33] (Chapter 4, Section 1, Theorem 4.2). This homeomorphism from the dual group $\widehat{G}$ to $\mathsf{X}(\mathrm{L}^1(G))$ is given by the map $\chi \mapsto \zeta_\chi$, with

$$\mathcal{G}_f(\zeta_\chi) = \zeta_\chi(f) = \int \chi(s) f(s)\, d\lambda(s), \qquad \chi \in \widehat{G}.$$

Consequently we can view the Gelfand transform $\mathcal{G}_f$ of f as a map $\overline{\mathcal{F}}(f)$ defined on $\widehat{G}$ instead of $\mathsf{X}(\mathrm{L}^1(G))$, namely

$$\overline{\mathcal{F}}(f)(\chi) = \int \chi(s) f(s)\, d\lambda(s), \qquad \chi \in \widehat{G}.$$

The map $\overline{\mathcal{F}}(f)$ is the *Fourier cotransform* of f. For technical reasons this map is denoted as $\overline{\mathcal{F}}(f)$ instead of $\mathcal{F}(f)$. The *Fourier transform* of f is the map $\mathcal{F}(f)$ defined on $\widehat{G}$ by

$$\mathcal{F}(f)(\chi) = \int \overline{\chi(s)} f(s)\, d\lambda(s), \qquad \chi \in \widehat{G}.$$

Most authors define the Fourier transform with the conjugate term $\overline{\chi(s)}$ under the integral, but this convention is not universally adopted. As in Folland and Bourbaki, this is the convention that we adopt. The theory of Fourier transforms on a commutative locally compact group will be discussed thoroughly in Chapter 10.

Finally we can characterize when the Gelfand transform is an isometry and when it is injective.

Proposition 9.26. *Let A be a commutative unital Banach algebra. The Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A); \mathbb{C})$ is an isometry iff $\|a^2\| = \|a\|^2$ for all $a \in A$.*

Proof. First we prove that $\|\mathcal{G}_a\|_\infty = \|a\|$ iff $\left\|a^{2^k}\right\| = \|a\|^{2^k}$ for all $k \geq 1$.

Since by Theorem 9.23 we have $\|\mathcal{G}_a\|_\infty = \rho(a)$, and since

$$\rho(a) = \lim_{k \mapsto \infty} \left\|a^{2^k}\right\|^{1/2^k},$$

if $\left\|a^{2^k}\right\| = \|a\|^{2^k}$ for all $k \geq 1$, then $\|\mathcal{G}_a\|_\infty = \|a\|$.

Conversely, assume that $\|\mathcal{G}_a\|_\infty = \|a\|$. Then

$$\left\|a^{2^k}\right\| \leq \|a\|^{2^k} = \|\mathcal{G}_a\|_\infty^{2^k} = \rho(a)^{2^k} = \rho(a^{2^k}) \leq \left\|a^{2^k}\right\|,$$

which shows that $\left\|a^{2^k}\right\| = \|a\|^{2^k}$ for all $k \geq 1$. Now if $\|a^2\| = \|a\|^2$, then by induction $\left\|a^{2^k}\right\| = \|a\|^{2^k}$ for all $k \geq 1$, so $\|\mathcal{G}_a\|_\infty = \|a\|$. Conversely, we already proved that if $\|\mathcal{G}_a\|_\infty = \|a\|$, then $\left\|a^{2^k}\right\| = \|a\|^{2^k}$ for all $k \geq 1$; in particular, $\|a^2\| = \|a\|^2$. ☐

Definition 9.14. Given a commutative unital algebra A, the *radical* of A, $\mathrm{rad}\, A$, is the intersection of all maximal ideals in A.

Proposition 9.27. *Let A be a commutative unital Banach algebra. We have $\mathrm{Ker}\, \mathcal{G} = \mathrm{rad}\, A$, so the following statements are equivalent.*

(1) The Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A); \mathbb{C})$ is injective.

(2) The radical of A is trivial; that is, $\mathrm{rad}\, A = (0)$.

(3) The set $\{a \in A \mid \sigma(a) = \{0\}\}$ is reduced to $\{0\}$.

(4) The set $\{a \in A \mid \rho(a) = 0\}$ *is reduced to* $\{0\}$.

Proof. The Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A); \mathbb{C})$ is injective iff its kernel is (0). We have $a \in \operatorname{Ker} \mathcal{G}$ iff $\mathcal{G}_a = 0$, which means that $\chi(a) = 0$ for all $\chi \in \mathsf{X}(A)$, that is, $a \in \operatorname{Ker} \chi$ for all $\chi \in \mathsf{X}(A)$. Since $\operatorname{Ker} \chi$ is a maximal ideal this shows that $a \in \operatorname{rad} A$. Thus $\operatorname{Ker} \mathcal{G} \subseteq \operatorname{rad} A$. Since by Theorem 9.21, every maximal ideal is the kernel of some character, $\operatorname{rad} A \subseteq \operatorname{Ker} \mathcal{G}$, so we have $\operatorname{Ker} \mathcal{G} = \operatorname{rad} A$.

$\operatorname{Ker} \mathcal{G} = \operatorname{rad} A$ implies that (1) and (2) are equivalent.

Since $a \in \operatorname{Ker} \mathcal{G}$ iff $\sigma(a) = \mathcal{G}_a(\mathsf{X}(A)) = \{0\}$, we see that $\mathcal{G}$ is not injective iff there is some nonzero $a \in A$ such that $\sigma(a) = \{0\}$, so (1) and (3) are indeed equivalent.

Since $\sigma(a) = \{0\}$ iff $\rho(a) = 0$, (3) and (4) are equivalent. $\square$

The notion of radical can also be defined for nonunital, noncommutative algebras using the notion of regular ideal (see Definition 9.3); see Bourbaki [10], Appendix, Section 3.

Definition 9.15. Let A be a nonunital, possibly noncommutative algebra. The *radical* $\operatorname{rad} A$, is the intersection of all maximal regular left ideals in A.

It is shown in Bourbaki [10] (Appendix, Section 3, Proposition 5) that this more general notion of a radical $\operatorname{rad} A$ is a two-sided ideal which is isomorphic to the intersection of all maximal left ideals of $\widetilde{A}$, namely the radical of $\widetilde{A}$ (see Section 9.2 for the definition of $\widetilde{A}$). The following generalization of Proposition 9.27 holds; see Bourbaki [9] (Chapter 1, Section 3, Proposition 5).

Proposition 9.28. *If A be a commutative nonunital Banach algebra, then* $\operatorname{Ker} \mathcal{G} = \operatorname{rad} A$.

We will show later on that if G is a locally compact group, then $\mathrm{L}^1(G)$ has a trivial radical. Recall that in general $\mathrm{L}^1(G)$ is a Banach algebra which is neither commutative nor unital so Definition 9.15 is needed to define its radical.

9.8 Banach Algebras with Involution; C^*-Algebras

If A is a complex matrix, then we have its conjugate-transpose $A^* = \overline{(A^\top)} = (\overline{A})^\top$, and we know that it satisfies various identities such as

$$(A+B)^* = A^* + B^*, \quad (AB)^* = B^* A^*, \quad (A^*)^* = A,$$

and so on. It turns out to be fruitful to define algebras and normed algebras having an operation $a \mapsto a^*$ satisfying the most useful laws of conjugate-transposition.

Definition 9.16. Let A be an algebra over $\mathbb{C}$ (not necessarily unital). An *involution* on A is a bijection $a \mapsto a^*$ satisfying the following axioms:

$$\begin{aligned} (a^*)^* &= a & \qquad (a+b)^* &= a^* + b^* \\ (\lambda a)^* &= \overline{\lambda} a^* & (ab)^* &= b^* a^*, \end{aligned}$$

for all $a, b \in A$ and all $\lambda \in \mathbb{C}$. The element a^* is called the *adjoint* of a. If $a = a^*$, then a is called *hermitian* (or *self-adjoint*). An algebra with an involution is called an *involutive algebra*.

If A is a normed algebra, then an *involutive normed algebra* is an algebra with an involution $a \mapsto a^*$ satisfying the extra axiom

$$\|a\| = \|a^*\|, \qquad \text{for all } a \in A. \tag{i}$$

A C^**-algebra* is a Banach algebra with an involution $a \mapsto a^*$ satisfying the axiom

$$\|a\|^2 = \|a^* a\|, \qquad \text{for all } a \in A. \tag{C^*}$$

Remark: Oddly, Rudin uses the term B^**-algebra* instead of C^*-algebra; see Rudin [80], Definition 11.17. The term C^**-algebra* seems to be used predominantly.

Here are a few immediate consequences of the axioms.

1. We have $0^* = 0$.

   This is because $0^* = (0+0)^* = 0^* + 0^*$, so $0^* = 0$.

2. If A is unital with identity e, then $e^* = e$.

   This is because using the axioms, for all $a \in A$, we have

   $$ae^* = (a^*)^* e^* = (ea^*)^* = (a^*)^* = a,$$

   and similarly

   $$e^* a = e^* (a^*)^* = (a^* e)^* = (a^*)^* = a,$$

   so e^* is a multiplicative identity element, and by uniqueness of such an element, $e^* = e$.

3. If $a \in A$ is invertible, then so is a^* and $(a^*)^{-1} = (a^{-1})^*$.

   We have

   $$(a^{-1})^* a^* = (aa^{-1})^* = e^* = e,$$

   and

   $$a^* (a^{-1})^* = (a^{-1} a)^* = e^* = e,$$

   so $(a^*)^{-1} = (a^{-1})^*$.

4. For any $a \in A$, a is invertible iff a^* is invertible.

   Since $(a^*)^* = a$, by applying (3) to a^* we see that if a^* is invertible, then a is invertible, and (3) gives the converse.

5. For all $a \in A$ and all $\lambda \in \mathbb{C}$, we have $\lambda \in \sigma(a)$ iff $\overline{\lambda} \in \sigma(a^*)$.

   This is because, by (4), $\lambda e - a$ is invertible iff $\overline{\lambda} e^* - a^* = \overline{\lambda} e - a^*$ is invertible.

6. If A is a C^*-algebra, then Equation (i) holds. Therefore a C^*-algebra is an involutive algebra.

   We already know that $0^* = 0$. For any $a \neq 0$,

$$\|a\|^2 = \|a^*a\| \leq \|a^*\| \, \|a\| \, ,$$

   so $\|a\| \leq \|a^*\|$. Since $(a^*)^* = a$, we also get $\|a^*\| \leq \|a\|$, so $\|a^*\| = \|a\|$.

7. If A is a C^*-algebra, then $\|aa^*\| = \|a\|^2$.

   Using the fact that $(a^*)^* = a$, and $\|a\| = \|a^*\|$, substituting a^* for a in $\|a^*a\| = \|a\|^2$, we get $\|aa^*\| = \|a^*\|^2 = \|a\|^2$.

8. If A is a unital C^*-algebra, then $\|e\| = 1$.

   We have

$$\|e\|^2 = \|e^*e\| = \|e\| \, .$$

   This implies that $\|e\| = 0, 1$, but since $e \neq 0$, we must have $\|e\| = 1$.

Example 9.6. The examples below are among the examples listed in Example 9.1.

(1) If E is a complex Hilbert space, then the space $\mathcal{L}(E)$ of continuous linear maps $f\colon E \to E$ is a C^*-algebra, with involution $h \mapsto h^*$.

(2) Let X be a topological space. Then the space $\mathcal{C}_b(X; \mathbb{C})$ of bounded continuous functions on X is a commutative C^* unital Banach algebra with involution $f \mapsto \overline{f}$. Let X be a compact topological space. Then the space $\mathcal{C}(X; \mathbb{C})$ of continuous functions on X is a commutative C^* unital Banach algebra with involution $f \mapsto \overline{f}$. If X is a locally compact space, then $\mathcal{C}_0(X; \mathbb{C})$ is a nonunital C^*-algebra with involution $f \mapsto \overline{f}$.

(3) Let G be a locally compact group. The space $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(G)$ of complex regular Borel measures on G, simply denoted $\mathcal{M}^1(G)$, is a unital Banach algebra, with the norm $\|\mu\| = |\mu|(G)$ defined in Definition 7.10, with the convolution as multiplication, and with the Dirac measure δ_1 as multiplicative unit. By Proposition 8.46, the map $\mu \mapsto \mu^* = \overline{\check{\mu}}$ is an involution that makes $\mathcal{M}^1(G)$ into an involutive algebra. In general it is not a C^*-algebra.

(4) Let G be a locally compact group equipped with a left Haar measure λ. The space $\mathrm{L}^1(G)$ (with the L^1-norm) can be identified with a subspace of $\mathcal{M}^1(G)$, using the embedding $f \mapsto f d\lambda$ given by Proposition 7.32. The space $\mathrm{L}^1(G)$ is a Banach algebra, but it is not unital unless G is discrete. If G is unimodular, then the map $f \mapsto f^*$ where $f^*(s) = \overline{f(s^{-1})}$ is an involution. If G is not unimodular, then we define f^* by $f^*(s) = \Delta(s^{-1})\overline{f(s^{-1})}$, and then the map $f \mapsto f^*$ is an involution such that the map $f \mapsto f\, d\lambda$ is an embedding of the involutive Banach algebra $\mathrm{L}^1(G)$ into the unital involutive Banach algebra $\mathcal{M}^1(G)$. In general $\mathrm{L}^1(G)$ is not a C^*-algebra.

(5) As a special case of (4), let $G = \mathbb{Z}$, in which case $\mathrm{L}^1(G)$ is the set of all sequences $x = (x_m)_{m\in\mathbb{Z}}$ with $x_m \in \mathbb{C}$, such that $\sum_{m\in\mathbb{Z}} |x_m| < \infty$. This space is also denoted $l^1(\mathbb{Z})$. The convolution product $x * y$ of $x = (x_m)$ and $y = (y_m)$ is given by

$$(x * y)_m = \sum_{p\in\mathbb{Z}} x_p y_{m-p},$$

and the norm by $\|x\| = \sum_{m\in\mathbb{Z}} |x_m|$. This is a commutative unital Banach algebra with identity element e_0 such that $e_0(0) = 1$ and $e_0(m) = 0$ for all $m \neq 0$. The map $x \mapsto x^*$ where

$$x^*_m = \overline{x}_{-m}.$$

is an involution. The involutive algebra $l^1(\mathbb{Z})$ is not a C^*-algebra.

Definition 9.17. Let A be an involutive algebra. An element $a \in A$ is *hermitian* (or *self-adjoint*) if $a^* = a$.

Observe that for any $a \in A$, the elements $x_1 = (a + a^*)/2$ and $x_2 = (a - a^*)/(2i)$ are also hermitian, and so are aa^* and a^*a.

Proposition 9.29. *Let A be an involutive algebra. Every $a \in A$ can be written as $a = x_1 + ix_2$ for two unique hermitian elements x_1, x_2.*

Proof. The elements $x_1 = (a + a^*)/2$ and $x_2 = (a - a^*)/(2i)$ are hermitian, and we have $a = x_1 + ix_2$. Conversely, if $a = x_1 + ix_2$ with x_1, x_2 hermitian, then $a^* = x_1^* - ix_2^* = x_1 - ix_2$, so $x_1 = (a + a^*)/2$ and $x_2 = (a - a^*)/(2i)$ are uniquely determined hermitian elements. $\square$

Definition 9.18. Let A be an involutive algebra. An element $a \in A$ is *normal* if $a^*a = aa^*$. If A is unital with identity element e, then $a \in A$ is *unitary* if $aa^* = a^*a = e$, that is, if a is invertible and if $a^{-1} = a^*$.

If A is a unital involutive algebra, then the unitary elements form a subgroup of A. If a is unitary, namely $a^{-1} = a^*$, then

$$(a^{-1})^* = (a^*)^* = a = (a^{-1})^{-1}$$

so a^{-1} is unitary. And if $a^{-1} = a^*$ and $b^{-1} = b^*$, then

$$(ab)^* = b^* a^* = b^{-1} a^{-1} = (ab)^{-1},$$

so ab is unitary.

As in the case of matrices, we have the following result about spectra.

Proposition 9.30. *Let A be a unital C^*-algebra. For every $a \in A$, if a is hermitian, then $\sigma(a) \subseteq \mathbb{R}$, and if a is unitary, then $\sigma(a) \subseteq \mathbb{T} = \mathbf{U}(1)$.*

Proof. Assume that $a^* = a$. If $\alpha + i\beta \in \sigma(a)$, with $\alpha, \beta \in \mathbb{R}$, then for every real number λ, we have $\alpha + i(\beta + \lambda) \in \sigma(a + i\lambda e)$, since $(\alpha + i(\beta + \lambda))e - (a + i\lambda e) = (\alpha + i\beta)e - a$. By Theorem 9.13, we have $|\mu| \leq \|b\|$ for every $\mu \in \sigma(b)$, and since λ is real $\overline{\lambda} = \lambda$, so we have

$$\begin{aligned}
\alpha^2 + (\beta + \lambda)^2 &\leq \|a + i\lambda e\|^2 \\
&= \|(a + i\lambda e)^*(a + i\lambda e)\| \\
&= \left\|a^* a + i\lambda a^* - i\overline{\lambda} a + \lambda^2 e\right\| \\
&= \left\|a^* a + \lambda^2 e\right\| \\
&\leq \|a^* a\| + \lambda^2 \\
&= \|a\|^2 + \lambda^2,
\end{aligned}$$

which yields

$$2\beta\lambda \leq \|a\|^2 - \alpha^2 - \beta^2.$$

Since the above holds for all $\lambda \in \mathbb{R}$, by picking λ of the same sign as β and $|\lambda|$ large enough we would violate the above inequality, so we must have $\beta = 0$.

If $aa^* = a^* a = e$, then

$$\|a\|^2 = \|a^* a\| = \|e\| = 1,$$

hence $\|a\| = 1$. Similarly, since a^{-1} is also unitary, we have $\|a^{-1}\| = 1$. By Proposition 9.15, we conclude that $\sigma(a) \subseteq \mathbf{U}(1)$. $\square$

Remark: The first part of Proposition 9.30 also holds for a nonunital C^*-algebra A. For every $a \in A$, if a is hermitian then $\sigma'(a) \subseteq \mathbb{R}$; see Bourbaki [9] (Chapter I, §6, No. 3, Proposition 3).

Definition 9.19. Let A and B be two involutive algebras. A map $\varphi\colon A \to B$ is an *involutive homomorphism* if it is a homomorphism of algebras such that $\varphi(a^*) = \varphi(a)^*$ for all $a \in A$. As usual, if A and B are unital we require that $\varphi(e_A) = e_B$. We say that A is an *involutive subalgebra* of B if it is a subalgebra of B and if A is closed under the involution $a \mapsto a^*$; that is, if $a \in A$, then $a^* \in A$.

The following result is proven in L. Schwartz [84] (Chapter II, Section 14, page 374).

Proposition 9.31. *Let A be a C^*-algebra (not necessarily commutative). For every $a \in A$, if a is normal, then we have $\|a^2\| = \|a\|^2$. As a consequence, $\rho(a) = \|a\|$. In particular, if A is commutative then the above facts hold.*

Proof. Using the fact that $\|bb^*\| = \|b\|^2$, we get

$$\|aa^*(aa^*)^*\| = \|aa^*\|^2 = \|a\|^4 .$$

We also have $\|aa^*(aa^*)^*\| = \|aa^*(a^*)^*a^*\| = \|aa^*aa^*\|$, and since a is normal, $aa^* = a^*a$, so $\|aa^*aa^*\| = \|a^2(a^*)^2\|$. But $(a^2)^* = (aa)^* = a^*a^* = (a^*)^2$, so

$$\left\|a^2(a^*)^2\right\| = \left\|a^2(a^2)^*\right\| = \left\|a^2\right\|^2 .$$

Consequently $\|a^2\|^2 = \|a\|^4$, and so $\|a^2\| = \|a\|^2$. By induction, we get $\left\|a^{2^k}\right\| = \|a\|^{2^k}$. Since

$$\rho(a) = \lim_{k \mapsto \infty} \left\|a^{2^k}\right\|^{1/2^k},$$

we conclude that $\rho(a) = \|a\|$. $\square$

Proposition 9.32. *Let A be a unital involutive Banach algebra and let B be a unital C^*-algebra. If $\varphi\colon A \to B$ is an involutive homomorphism, then $\|\varphi(a)\| \leq \|a\|$ for all $a \in A$. Thus φ is continuous.*

Proof. We know from Proposition 9.31 that $\rho(b) = \|b\|$ for every hermitian $b \in B$ (since a hermitian element is obviously normal). It can easily be shown that $\sigma_B(\varphi(a)) \subseteq \sigma_A(a)$, so by definition of $\rho(a)$ (see Definition 9.12), we have

$$\rho(\varphi(a)) \leq \rho(a) \leq \|a\| .$$

Since $\rho(b) = \|b\|$ for every hermitian $b \in B$, and since $\varphi(a^*a)$ is hermitian (because $\varphi(a^*a)^* = \varphi((a^*a)^*) = \varphi(a^*a)$), we get

$$\|\varphi(a)\|^2 = \|\varphi(a)^*\varphi(a)\| = \|\varphi(a^*)\varphi(a)\| = \|\varphi(a^*a)\| = \rho(\varphi(a^*a)) \leq \|a^*a\| = \|a\|^2 ,$$

which implies $\|\varphi(a)\| \leq \|a\|$. $\square$

Proposition 9.32 also holds if A and B are not unital. This is because $\sigma'_B(\varphi(a)) \subseteq \sigma'_A(a)$; see Bourbaki [9] (Chapter I, §6, No. 3, Proposition 1).

Proposition 9.18 is sharpened as follows.

Proposition 9.33. *Let A and B be two unital C^*-algebra with A a closed involutive subalgebra of B. We have $\sigma_B(a) = \sigma_A(a)$ for all $a \in A$. If $a \in A$ is invertible in B, then $a^{-1} \in A$.*

Proof. If $a \in A$ is hermitian, we know from Proposition 9.30 that $\sigma_A(a) \subseteq \mathbb{R}$. Thus all points of $\sigma_A(a)$ are boundary points, so by Proposition 9.18 we have $\sigma_B(a) = \sigma_A(a)$.

Consider any $a \in A$ such that $a^{-1} \in B$. Then a^* is also invertible in B, so aa^* is invertible in B (similarly a^*a is invertible in B). Since $a \in A$ and A is an involutive subalgebra, $a^* \in A$, then $aa^* \in A$, and since aa^* is hermitian, by the fact we just proved above, $(aa^*)^{-1} \in A$. This implies that $aa^*(aa^*)^{-1} = e_A$, so a has a right inverse in A. A similar argument applied to a^*a shows that $(a^*a)^{-1} \in A$, so $(a^*a)^{-1}a^*a = e_A$ and a has a left inverse in A. Therefore $a^{-1} \in A$. This argument applied to $\lambda e - a$ (with $a \in A$) shows that $\sigma_B(a) = \sigma_A(a)$. ☐

Let A be an involutive algebra. For any linear form $f\colon A \to \mathbb{C}$, let f^* be the map given by

$$f^*(a) = \overline{f(a^*)}, \qquad a \in A.$$

We have

$$f^*(a+b) = \overline{f((a+b)^*)} = \overline{f(a^* + b^*)} = \overline{f(a^*)} + \overline{f(b^*)} = f^*(a) + f^*(b),$$

and

$$f^*(\lambda a) = \overline{f((\lambda a)^*)} = \overline{f(\overline{\lambda} a^*)} = \overline{\overline{\lambda} f(a^*)} = \lambda \overline{f(a^*)} = \lambda f^*(a),$$

so f^* is also a linear form. We verify immediately that

$$(f^*)^* = f, \qquad (f+g)^* = f^* + g^*, \qquad (\lambda f)^* = \overline{\lambda} f^*.$$

Definition 9.20. Let A be an involutive algebra. For any linear form $f\colon A \to \mathbb{C}$, the linear form f^* given by

$$f^*(a) = \overline{f(a^*)}, \qquad a \in A$$

is called the *adjoint* of f. We say that f is *hermitian* (or *self-adjoint*) if $f^* = f$.

9.9 Characters and Gelfand Transform in a C^*-Algebra

Interestingly, the characters of a commutative unital C^*-algebra are hermitian.

Proposition 9.34. *Let A be a commutative unital C^*-algebra. Then for any character $\chi \in \mathsf{X}(A)$, we have*

$$\chi(a^*) = \overline{\chi(a)}, \qquad \text{for all } a \in A,$$

or equivalently $\chi(a) = \overline{\chi(a^*)}$, which shows that the characters are hermitian. Consequently $\mathcal{G}_{a^*} = \overline{\mathcal{G}_a}$ for all $a \in A$, where $\mathcal{G}_a$ is the Gelfand transform of a.

Proof. First assume that $a \in A$ is hermitian. By Proposition 9.30 we have $\sigma(a) \subseteq \mathbb{R}$, and since by Proposition 9.12 we have $\chi(a) \in \sigma(a)$ for any $\chi \in \mathsf{X}(A)$, we have $\chi(a) \in \mathbb{R}$, and since a is hermitian $a^* = a$, so

$$\chi(a^*) = \chi(a) = \overline{\chi(a)}.$$

Any arbitrary $a \in A$ can be written as $a = x_1 + ix_2$ for two unique hermitian elements $x_1, x_2 \in A$ (see Proposition 9.29), and by the above fact

$$\chi(x_1) = \overline{\chi(x_1)}, \qquad \chi(x_2) = \overline{\chi(x_2)}$$

and since $a = x_1 + ix_2$ and $a^* = x_1^* - ix_2^* = x_1 - ix_2$ (because x_1 and x_2 are hermitian),

$$\chi(a^*) = \chi(x_1 - ix_2) = \chi(x_1) - i\chi(x_2) = \overline{\chi(x_1)} - i\overline{\chi(x_1)} = \overline{\chi(x_1) + i\chi(x_2)} = \overline{\chi(a)}.$$

Since by definition $\mathcal{G}_a(\chi) = \chi(a)$, we have

$$\mathcal{G}_{a^*}(\chi) = \chi(a^*) = \overline{\chi(a)} = \overline{\mathcal{G}_a(\chi)},$$

which means that $\mathcal{G}_{a^*} = \overline{\mathcal{G}_a}$, as claimed. □

Remark: Proposition 9.34 also holds if A is a noncommutative and nonunital C^*-algebra; see Bourbaki [9] (Chapter I, §6, No. 4, Theorem 1).

If A is a commutative unital C^*-algebra, we can make the following addition to Proposition 9.24.

Proposition 9.35. *Let A be a commutative unital C^*-algebra. For any fixed $a \in A$, if the algebra A is generated by a, a^*, and e, then the Gelfand transform $\mathcal{G}_a$ is a homeomorphism from $\mathsf{X}(A)$ to $\sigma(a)$.*

Proof. By Theorem 9.23, the map $\mathcal{G}_a$ is continuous and surjective onto $\sigma(a)$. Since $\mathsf{X}(A)$ and $\sigma(a)$ are compact Hausdorff spaces (by Theorem 9.19(2) and Theorem 9.13(1)), by the corollary to Proposition A.33, it suffices to show that this map is injective. But any character $\chi\colon A \to \mathbb{C}$ is uniquely determined by $\chi(a)$, since by Proposition 9.34, we have $\chi(a^*) = \overline{\chi(a)}$. If $\mathcal{G}_a(\chi_1) = \mathcal{G}_a(\chi_2)$, then $\chi_1(a) = \chi_2(a)$, and since χ_1 is completely determined by $\chi_1(a)$ and similarly χ_2 is completely determined by $\chi_2(a)$, we have $\chi_1 = \chi_2$, and $\mathcal{G}_a$ is injective. □

As an application of Proposition 9.35, let H be a Hilbert space, and let T be a bounded normal operator on H (that is, $TT^* = T^*T$). Let $\mathcal{A}_T$ be the subalgebra of $\mathcal{L}(H)$ generated by T, T^* and I. Since T and T^* commute, $\mathcal{A}_T$ is a commutative unital C^*-algebra, and $\mathsf{X}(\mathcal{A}_T)$ is homeomorphic to the spectrum $\sigma(T)$ of $T \in \mathcal{L}(H)$. This is the first step in obtaining a spectral theorem for a bounded normal operator on a Hilbert space; see Folland [33] (Chapter 1, Section 1.4) and Dieudonné [24] (Chapter XV, Section 11).

We are now ready to prove the main theorem of the theory of commutative unital C^*-algebras due to Gelfand and Naimark, namely that every commutative unital C^*-algebra can be viewed as the algebra of continuous functions on a compact space, namely its space of characters $\mathsf{X}(A)$. The proof makes use of the version of the Stone–Weierstrass theorem for complex-valued functions that we now recall.

Theorem 9.36. *(Stone–Weierstrass) Let X be a compact space, and let $\mathcal{C}(X;\mathbb{C})$ be the algebra of continuous functions on X. Let B be a subalgebra of $\mathcal{C}(X;\mathbb{C})$ satisfying the following properties:*

(1) The algebra B contains the constant functions.

(2) The algebra B separates the points of X, which means that for any points $x, y \in X$, if $x \neq y$ then there is some function $f \in B$ such that $f(x) \neq f(y)$.

(3) The algebra B is stable under conjugation; that is, for any $f \in B$, the function $\overline{f}$ also belongs to B (where $\overline{f}(x) = \overline{f(x)}$ for all $x \in X$).

Then B is dense in $\mathcal{C}(X;\mathbb{C})$ (with respect to the $\| \|_\infty$ norm); that is, for every function $f \in \mathcal{C}(X;\mathbb{C})$, there is a sequence (f_n) with $f_n \in B$ that converges uniformly to f.

Theorem 9.36 is a cornerstone of analysis. Its proof can be found in many books, including Schwartz [84], Folland [34] (Chapter 4, Theorem 4.15), and Rudin [80] (Chapter 5, Theorem 5.7).

Theorem 9.37. *(Gelfand–Naimark) Let A be a commutative unital C^*-algebra. Then the Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A);\mathbb{C})$ is an isometric isomorphism between A and $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$ (and so $\|\mathcal{G}_a\|_\infty = \|a\| = \rho(a)$ for all $a \in A$). Furthermore the Gelfand maps $\mathcal{G}_a$ are hermitian, which means that $\mathcal{G}_{a^*} = \overline{\mathcal{G}_a}$, for all $a \in A$.*

Proof. Since A is commutative, by Proposition 9.31, we have $\|a^2\| = \|a\|^2$ for all $a \in A$. Since A is a unital Banach algebra, by Proposition 9.26 the Gelfand transform $\mathcal{G}\colon A \to \mathcal{C}(\mathsf{X}(A);\mathbb{C})$ is an isometry. In particular it is injective. This implies that the image $\mathcal{G}(A)$ of A is closed in $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$. Indeed, for any Cauchy sequence $(\mathcal{G}_{a_n})$ in $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$, since $\|a_m - a_n\| = \|\mathcal{G}_{a_m} - \mathcal{G}_{a_n}\|_\infty$, the sequence (a_n) is a Cauchy sequence in A, and since A is a Banach space the sequence (a_n) has a limit $a \in A$. Since the Gelfand transform is continuous, $\mathcal{G}_a$ is the limit of the sequence $(\mathcal{G}_{a_n})$. Therefore $\mathcal{G}(A)$ is closed in $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$. It remains to prove that $\mathcal{G}(A) = \mathcal{C}(\mathsf{X}(A);\mathbb{C})$.

For this we check that the hypotheses of the Stone–Weierstrass theorem (Theorem 9.36) are satisfied. Since A is algebra and $\mathcal{G}$ is a homomorphism, $B = \mathcal{G}(A)$ is a subalgebra of $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$.

(1) The algebra B contains all the constant functions, since $\mathcal{G}_{\lambda e}$ is the constant function λ.

(2) The algebra B separates points. Indeed, if χ_1 and χ_2 are two distinct characters, then $\chi_1(a) \neq \chi_2(a)$ for some $a \in A$, and then $\mathcal{G}_a(\chi_1) = \chi_1(a) \neq \chi_2(a) = \mathcal{G}_a(\chi_2)$, so $\mathcal{G}_a$ separates χ_1 and χ_2

(3) By Proposition 9.34, we have $\mathcal{G}_{a^*} = \overline{\mathcal{G}_a}$ for all $a \in A$, so B is stable under conjugation.

By the Stone–Weierstrass theorem, B is dense in $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$. But B is closed in $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$, so $\mathcal{G}(A) = B = \mathcal{C}(\mathsf{X}(A);\mathbb{C})$, proving that $\mathcal{G}$ is an isomorphism. ☐

The Gelfand–Naimark theorem is used to prove the Plancherel–Godement theorem (see Section 11.8, Theorem 11.41), and some representation theory results in harmonic analysis; see Dieudonné [22].

If A is a nonunital commutative C^*-algebra, then there is a version of the Gelfand–Naimark theorem in which the algebra $\mathcal{C}(\mathsf{X}(A);\mathbb{C})$ is replaced by the algebra $\mathcal{C}_0(\mathsf{X}(A);\mathbb{C})$ of continuous functions that tend to zero at infinity. Thus there is an isometric isomorphism $\mathcal{G}\colon A \to \mathcal{C}_0(\mathsf{X}(A);\mathbb{C})$; see Bourbaki [9] (Chapter 1, Section 6, No. 4) and Folland [33] (Chapter 1, Section 1.3).

There is also a version of the Gelfand–Naimark theorem for noncommutative C^*-algebras. Roughly speaking, a C^*-algebra is isometrically isomorphic to a C^*-subalgebra of the algebra of bounded operators on some Hilbert space; see Rudin [80].

The spectral theory of C^*-algebras is the key machinery used to develop generalizations of the spectral theorems for normal matrices to bounded (and unbounded) operators of various kinds on a Hilbert space. A condensed presentation of these spectral theorems is given in Folland [33] (Chapter 1, Section 1.4) and in Dieudonné [24] (Chapter XV, Section 11). An extensive treatment of spectral theorems is given in Rudin [80], and in Lax [64].

Since the main goals of this book are to discuss harmonic analysis and representation theory, spectral theorems for families of operators on a Hilbert space are not a prime topic of interest. However, Theorem 9.37 and Proposition 9.35 yield an interesting preliminary version of the spectral theorem for bounded normal operators.

Let H be a Hilbert space, and let T be a bounded normal operator on H, which means that $TT^* = T^*T$. Let $\mathcal{A}_T$ be the subalgebra of $\mathcal{L}(H)$ generated by T, T^* and I. Since T and T^* commute, $\mathcal{A}_T$ is a commutative unital C^*-algebra. By Proposition 9.35, the Gelfand transform $\mathcal{G}_T\colon \mathsf{X}(\mathcal{A}_T) \to \sigma(T)$ (given by $\mathcal{G}_T(\chi) = \chi(T), \chi \in \mathsf{X}(\mathcal{A}_T)$) is a homeomorphism between $\mathsf{X}(\mathcal{A}_T)$ and the spectrum $\sigma(T)$ of $T \in \mathcal{A}_T$.

Actually, it is important to note that if $\mathcal{A}$ is any unital (not necessarily commutative) C^*-subalgebra of $\mathcal{L}(H)$, by Proposition 9.33, the spectrum of $T \in \mathcal{A}$ with respect to the algebra $\mathcal{A}$ is equal to the spectrum of T with repect to the algebra $\mathcal{L}(H)$. Thus from now on we will always assume that the spectrum $\sigma(T)$ of a map T in $\mathcal{A} \subseteq \mathcal{L}(H)$ is defined with respect to $\mathcal{L}(H)$.

Theorem 9.38. *Let H be a Hilbert space and let T be a bounded normal operator on H. There is an isometric isomorphism $G\colon \mathcal{A}_T \to \mathcal{C}(\sigma(T);\mathbb{C})$ such that*

$$G(T) = \mathrm{id}_{\sigma(T)}.$$

Proof. By Gelfand–Naimark (Theorem 9.37), the Gelfand transform $\mathcal{G}\colon \mathcal{A}_T \to \mathcal{C}(\mathsf{X}(\mathcal{A}_T),\mathbb{C})$ is an isometric isomorphism. The homeomorphism $\mathcal{G}_T\colon \mathsf{X}(\mathcal{A}_T) \to \sigma(T)$ has an inverse

$\mathcal{G}_T^{-1}\colon \sigma(T) \to \mathsf{X}(\mathcal{A}_T)$, and the map $\mathcal{G}_T^{-1}$ induces a map $\theta\colon \mathcal{C}(\mathsf{X}(\mathcal{A}_T);\mathbb{C}) \to \mathcal{C}(\sigma(T);\mathbb{C})$ given by

$$\theta(f) = f \circ \mathcal{G}_T^{-1}, \quad f \in \mathcal{C}(\mathsf{X}(\mathcal{A}_T);\mathbb{C}).$$

We leave it as an exercise to check that θ is an isometric isomorphism. Let $G = \theta \circ \mathcal{G}$. Since both $\mathcal{G}$ and θ are isometric isomorphisms, G is an isometric isomorphism between $\mathcal{A}_T$ and $\mathcal{C}(\sigma(T);\mathbb{C})$, so it remains to prove that $G(T) = \mathrm{id}_{\sigma(T)}$. For $f = \mathcal{G}_T$, for every $\lambda \in \sigma(T)$, we have

$$\begin{aligned} G(T)(\lambda) &= (\theta(\mathcal{G}(T)))(\lambda) \\ &= (\theta(\mathcal{G}_T))(\lambda) \\ &= (\mathcal{G}_T \circ \mathcal{G}_T^{-1})(\lambda) = \lambda, \end{aligned}$$

so $G(T) = \mathrm{id}_{\sigma(T)}$, as claimed. $\square$

It can be shown that the map G of Theorem 9.38 satisfying the property $G(T) = \mathrm{id}_{\sigma(T)}$ is unique; see Schwartz [84].

Observe that the inverse $G^{-1}\colon \mathcal{C}(\sigma(T);\mathbb{C}) \to \mathcal{A}_T$ of the isomorphism $G\colon \mathcal{A}_T \to \mathcal{C}(\sigma(T);\mathbb{C})$ is a bounded linear map on $\mathcal{C}(\sigma(T);\mathbb{C})$ taking its values in the Banach space $\mathcal{A}_T \subseteq \mathcal{L}(H)$. Thus G^{-1} is a vector-valued continuous Radon functional and we should expect that it can be defined as an integral with respect to some kind of measure. This can be indeed be done using H-projection-valued measures, as explained in Folland [33] (Chapter 1, Section 1.4) and in a slightly different formalism in Dieudonné [24] (Chapter XV, Section 11). The key point is that for any two vectors $u, v \in H$, the map $\Phi_{u,v}$ defined on $\mathcal{C}(\sigma(T), \mathbb{C})$ by

$$\Phi_{u,v}(f) = \langle G^{-1}(f)(u), v\rangle$$

is a bounded Radon functional, so by Radon–Riesz III, it corresponds to a unique regular complex Borel measure $\mu_{u,v}$ such that

$$\langle G^{-1}(f)(u), v\rangle = \int_{\sigma(T)} f\, d\mu_{u,v} \quad \text{for all } f \in \mathcal{C}(\sigma(T);\mathbb{C}).$$

The next step is to define the projection-valued measures, but we will not do this here; we refer the reader to Folland [33] (Chapter 1, Section 1.4) and Dieudonné [24] (Chapter XV, Section 11). Folland actually deals with the more general situation of an arbitary commutative C^*-subalgebra of $\mathcal{L}(H)$ containing I, whereas Dieudonné restricts his attention to the C^*-algebra $\mathcal{A}_T$. They both make the crucial observation that $G^{-1}\colon \mathcal{C}(\sigma(T);\mathbb{C}) \to \mathcal{A}_T$ is a C^*-algebra homomorphism whose range is a subspace of $\mathcal{L}(H)$, so it is a *representation* of the algebra $\mathcal{C}(\sigma(T);\mathbb{C})$ in the Hilbert space H, in the sense of Definition 11.1 (in fact, it is a faithful representation). If H is a separable Hilbert space, representations of $\mathcal{C}(K;\mathbb{C})$ in H where K is a compact metrizable space can be completely classified, which is the approach followed by Dieudonné (recall that $\sigma(T)$ is compact). For more on this topic, see Section 11.9.

9.10 Enveloping C^*-Algebra of an Involutive Banach Algebra

If A is an involutive Banach algebra, there is a C^*-algebra $\mathrm{St}(A)$ and an involutive homomorphism $j\colon A \to \mathrm{St}(A)$ that satisfy a universal mapping condition with respect to homomorphisms of A into a C^*-algebra. To construct $\mathrm{St}(A)$, first we establish the following result.

Proposition 9.39. *Let A be an involutive algebra (over $\mathbb{C}$) and let p be a semi-norm on A. The following conditions are equivalent:*

(1) We have $p(ab) \leq p(a)p(b)$, $p(a^*) = p(a)$, $(p(a))^2 = p(a^*a)$, for all $a, b \in A$.

(2) The set $\mathfrak{N}$ of all $a \in A$ such that $p(a) = 0$ is a self-adjoint ($\mathfrak{N}^* = \mathfrak{N}$) two-sided ideal of A, and the norm induced on $A/\mathfrak{N}$ (in Proposition 9.7) makes $A/\mathfrak{N}$ an involutive normed algebra whose completion is a C^*-algebra.

(3) There is a homomorphism φ of the involutive algebra A into a C^*-algebra such that $p(a) = \|\varphi(a)\|$ for all $a \in A$.

The proof of Proposition 9.39 is given in Bourbaki [9] (Chapter I, §6, No. 6, Lemma 1).

Definition 9.21. Let A be an involutive algebra (over $\mathbb{C}$). A semi-norm p satisfying the conditions of Proposition 9.39 is called a *stellar semi-norm* on A.

Let us now assume that A is an involutive Banach algebra. Let S be the set of stellar semi-norms on A. By Proposition 9.32, we have $p(a) \leq \|a\|$ for all $a \in A$ and all $p \in S$. Consequently the function $a \mapsto \|a\|_*$ given by

$$\|a\|_* = \sup_{p \in S} p(a)$$

is the largest stellar semi-norm in A.

Definition 9.22. Let A be an involutive Banach algebra, and let $\mathfrak{N}$ be the set of all $a \in A$ such that $\|a\|_* = 0$. The C^*-algebra obtained by completing the normed involutive algebra $A/\mathfrak{N}$ for the norm induced by $\| \, \|_*$ is called the *enveloping C^*-algebra* of A, and is denoted $\mathrm{St}(A)$. The map from A to the completion of $A/\mathfrak{N}$ induced by the canonical map from A to $A/\mathfrak{N}$ is denoted by j.

The enveloping C^*-algebra $\mathrm{St}(A)$ has the following universal mapping property.

Theorem 9.40. *Let A be an involutive Banach algebra. For every involutive homomorphism $\varphi\colon A \to B$ of A into a C^*-algebra B, there is a unique involutive homomorphism $\overline{\varphi}\colon \mathrm{St}(A) \to B$ such that*

$$\varphi = \overline{\varphi} \circ j,$$

as shown in the following commutative diagram:

$$\begin{tikzcd} A \arrow[r, "j"] \arrow[dr, "\varphi"'] & \mathrm{St}(A) \arrow[d, "\overline{\varphi}"] \\ & B. \end{tikzcd}$$

If A is commutative, then $\mathrm{St}(A)$ *is commutative, and if A is unital, then* $\mathrm{St}(A)$ *is unital.*

Proof sketch. The map $a \mapsto \|\varphi(a)\|$ is a stellar semi-norm on A. Thus $\|\varphi(a)\| \leq \|a\|_*$ for all $a \in A$, and it is easy to check that the homomorphism obtained by the quotient operation is continuous from $A/\mathfrak{N}$ to B, and extends uniquely to $\mathrm{St}(A)$. The uniqueness of $\overline{\varphi}$ is standard. ∎

If A is abelian it is also possible to characterize the set of characters $\mathsf{X}(\mathrm{St}(A))$ of $\mathrm{St}(A)$. Recall that given the homomorphism $j\colon A \to \mathrm{St}(A)$, the homomorphism $\mathsf{X}(j)\colon \mathsf{X}(\mathrm{St}(A)) \to \mathsf{X}(A)$ is given by

$$\mathsf{X}(j)(\chi) = \chi \circ j, \qquad \chi \in \mathsf{X}(\mathrm{St}(A)).$$

Proposition 9.41. *Let A be a commutative involutive Banach algebra. The map $\mathsf{X}(j)$ is a homeomorphism of the set of characters $\mathsf{X}(\mathrm{St}(A))$ onto the subspace H of hermitian characters in $\mathsf{X}(A)$; that is, the characters $\chi\colon A \to \mathbb{C}$ such that $\chi(a) = \overline{\chi(a^*)}$ for all $a \in A$.*

Proposition 9.41 is proven in Bourbaki [9] (Chapter I, §6, No. 6, corollary).

It is easy to show that if $a \in \mathrm{rad}\, A$ (the radical of A), then $j(a) = 0$.

If G is a locally compact group, then $\mathrm{L}^1(G)$ is an involutive Banach algebra, but in general it is not a C^*-algebra. Thus we can form the enveloping C^*-algebra $\mathrm{St}(\mathrm{L}^1(G))$ of $\mathrm{L}^1(G)$, denoted $\mathrm{St}(G)$.

Definition 9.23. If G is a locally compact group, then the enveloping C^*-algebra $\mathrm{St}(\mathrm{L}^1(G))$ of $\mathrm{L}^1(G)$ is denoted $\mathrm{St}(G)$.

Remarkably, the canonical map j is injective.

Proposition 9.42. *Let G be a locally compact group. The canonical map j from $\mathrm{L}^1(G)$ to its enveloping C^*-algebra $\mathrm{St}(G)$ is injective.*

Proposition 9.42 is proven in Bourbaki [9] (Chapter I, §6, No. 7, Proposition 12).

As a corollary, the following result is obtained; see Bourbaki [9] (Chapter 1, Section 7, No. 7).

Proposition 9.43. *If G is a locally compact group, then the radical of $\mathrm{L}^1(G)$ is the zero ideal.*

We will see in Section 10.1 that if G is an abelian locally compact group, then every character of $\mathrm{L}^1(G)$ is hermitian. By Proposition 9.41, there is a homeomorphism between $\mathsf{X}(\mathrm{St}(G))$ and $\mathsf{X}(\mathrm{L}^1(G))$ (see Proposition 10.8).

Chapter 10

Harmonic Analysis on Locally Compact Abelian Groups

In this chapter we generalize the various Fourier transforms and cotransforms defined in Chapter 6 to an arbitrary locally compact abelian group (often abbreviated as LCA groups).

The fact that the Fourier transform $\widehat{f}\colon \mathbb{R} \to \mathbb{C}$ of a function $f \in \mathrm{L}^1(\mathbb{R})$, given by

$$\widehat{f}(x) = \int e^{-iyx} f(y) \, \frac{dx(y)}{\sqrt{2\pi}}$$

is also a function defined on $\mathbb{R}$, is an accident (perhaps a convenient accident).

On the other hand, given a function $f \in \mathrm{L}^1(\mathbb{T})$ (a periodic function of period 2π), its Fourier transform $\widehat{f} = \mathcal{F}(f)$ is the *sequence* $\widehat{f} = (c_m)_{\in \mathbb{Z}}$ of *Fourier coefficients*

$$c_m = \int_{-\pi}^{\pi} e^{-imx} f(x) \frac{dx}{2\pi}.$$

We can view $\widehat{f} = (c_m)_{\in \mathbb{Z}}$ as a function $\widehat{f}\colon \mathbb{Z} \to \mathbb{C}$. The domain of the Fourier transform $\widehat{f}$ is $\mathbb{Z}$, which is completely different from $\mathbb{T}$.

If we now consider functions in $l^1(\mathbb{Z})$, which are sequences $c = (c_m)_{\in \mathbb{Z}}$ with $c_m \in \mathbb{C}$ such that $\sum_{m \in \mathbb{Z}} |c_m| < \infty$, then we can define the Fourier transform $\widehat{c} = \mathcal{F}(c)$ of c as the function $\mathcal{F}(c)\colon \mathbb{T} \to \mathbb{C}$ defined on $\mathbb{T}$ given by

$$\mathcal{F}(c)(e^{i\theta}) = \sum_{m \in \mathbb{Z}} c_m e^{-im\theta}.$$

The domain of this Fourier transform is $\mathbb{T}$, which is completely different from $\mathbb{Z}$. Because $\sum_{m \in \mathbb{Z}} |c_m| < \infty$, the above series is absolutely convergent, so $\mathcal{F}(c) \in \mathrm{L}^1(\mathbb{T})$.

Observe an asymmetry. If $f \in \mathrm{L}^1(\mathbb{T})$, then the Fourier transform $\widehat{f} = \mathcal{F}(f) = (c_m)_{\in \mathbb{Z}}$ may not belong to $l^1(\mathbb{Z})$. The same problem arises for the Fourier transform on $\mathrm{L}^1(\mathbb{R})$. In

general, $\widehat{f} \notin \mathrm{L}^1(\mathbb{R})$. This is the problem of Fourier inversion. We would like to know when it is possible to recover a function f from its Fourier transform $\widehat{f}$. This is a difficult problem.

Remarkably, the Fourier transform on L^2 is better behaved. This is the content of the Plancherel theorem which shows that Fourier inversion is possible.

The question remains, what should be the domain of a Fourier transform?

Observe that the examples that we considered involve the Fourier transform of the L^1-functions on the abelian groups, $\mathbb{R}, \mathbb{T}$, and $\mathbb{Z}$. These groups are locally compact.

In the case of a commutative locally compact group G (equipped with a Haar measure λ), it turns out that a good solution is to define the domain of the Fourier transform as the dual group $\widehat{G}$ of G, which is a certain group of homomorphisms $\chi\colon G \to \mathbb{C}$, namely the continuous unitary homomorphisms $\chi\colon G \to \mathbf{U}(1)$.

The group $\widehat{G}$ of characters of G is defined in Section 10.1. Having defined a group structure on $\widehat{G}$, the next goal is to make $\widehat{G}$ into a topological group which is locally compact. Since $\widehat{G}$ consists of continuous functions from G to $\mathbf{U}(1)$, we can give $\widehat{G}$ the compact-open topology, but proving that the resulting space is locally compact is nontrivial. This can be done by proving that the spaces $\widehat{G}$ and $\mathsf{X}(\mathrm{L}^1(G))$ are homeomorphic; see Theorem 10.6. Since by Proposition 9.20, the space $\mathsf{X}(\mathrm{L}^1(G))$ is locally compact, we obtain the fact that $\widehat{G}$ is locally compact.

Actually, in Proposition 10.5, we prove that if G is a locally compact abelian group with a left Haar measure λ, then for every character $\chi \in \widehat{G}$, the map ζ_χ given by

$$\zeta_\chi(\mu) = \int \chi(a)\, d\mu(a) \qquad \text{for all } \mu \in \mathcal{M}^1(G)$$

is a hermitian character (an algebra homomorphism such that $\zeta_\chi(\overline{\mu}) = \overline{\zeta_\chi(\mu)}$) of the algebra $\mathcal{M}^1(G)$. By restriction to $\mathrm{L}^1(G)$, the map ζ_χ given by

$$\zeta_\chi(f) = \int \chi(a) f(a)\, d\lambda(a) \qquad \text{for all } f \in \mathrm{L}^1(G)$$

is a hermitian character of $\mathrm{L}^1(G)$ not equal to the zero function. Then the map $j\colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$ given by

$$j(\chi)(f) = \zeta_\chi(f) = \int \chi(a) f(a)\, d\lambda(a), \qquad \chi \in \widehat{G},\ f \in \mathrm{L}^1(G),$$

is a homeomorphism of $\widehat{G}$ onto $\mathsf{X}(\mathrm{L}^1(G))$.

We also prove that, the spaces $\widehat{G}$, $\mathsf{X}(\mathrm{L}^1(G))$, and $\mathsf{X}(\mathrm{St}(G))$, are homeomorphic.

Next in Section 10.2 we determine the characters of the groups $\mathbb{Z}, \mathbb{T}, \mathbb{Z}/p\mathbb{Z}$, and $\mathbb{R}$. As a corollary, we obtain the isomorphisms

$$\widehat{\mathbb{R}^n} \cong \mathbb{R}^n, \quad \widehat{\mathbb{T}^n} \cong \mathbb{Z}^n, \quad \widehat{\mathbb{Z}^n} \cong \mathbb{T}^n.$$

We prove that if G is a finite locally compact abelian group, then $\widehat{G}$ is isomorphic to G.

If G is a compact abelian group of Haar measure 1, then its characters form an orthonormal set in $\mathrm{L}^2(G)$.

We conclude by showing that there is a natural injection of G into its double dual $\widehat{\widehat{G}}$.

Given any $a \in G$, define the map $\eta_a\colon \widehat{G} \to \mathbb{C}$ by

$$\eta_a(\chi) = \chi(a), \qquad \text{evaluation at } a.$$

The map $\eta\colon G \to \widehat{\widehat{G}}$ given by $\eta(a) = \eta_a$ is a continuous homomorphism from G to its double dual $\widehat{\widehat{G}}$.

Actually, η is an isomorphism, but this is much harder to prove (this is the Pontrjagin duality theorem).

Section 10.3 is devoted to the definition of the Fourier transform and the Fourier cotransform on an arbitrary locally compact abelian group G equipped with a Haar measure λ. For every function $f \in \mathrm{L}^1(G)$,

(1) The *Fourier transform* of f is the function $\mathcal{F}(f)\colon \widehat{G} \to \mathbb{C}$ given by

$$\mathcal{F}(f)(\chi) = \int \overline{\chi(a)} f(a)\, d\lambda(a), \qquad \chi \in \widehat{G}.$$

(2) The *Fourier cotransform* of f is the function $\overline{\mathcal{F}}(f)\colon \widehat{G} \to \mathbb{C}$ given by

$$\overline{\mathcal{F}}(f)(\chi) = \int \chi(a) f(a)\, d\lambda(a), \qquad \chi \in \widehat{G}.$$

These transforms are not independent. In fact, each one can be obtained from the other. For all $f \in \mathrm{L}^1(G)$, and all $\chi \in \widehat{G}$, we have

$$\overline{\mathcal{F}}(f)(\chi) = \mathcal{F}(f)(\chi^{-1}) = \mathcal{F}(\check{f})(\chi) = \overline{\mathcal{F}(\overline{f})(\chi)}.$$

We show that modulo the isomorphism $j\colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$, the Fourier cotransform $\overline{\mathcal{F}}(f)$ is the Gelfand transform $\mathcal{G}_f$ from $\mathrm{L}^1(G)$ to $\mathsf{X}(\mathrm{L}^1(G))$.

Actually, it is possible to define the Fourier transform and the Fourier cotransform on the algebra $\mathcal{M}^1(G)$; see Definition 10.4.

We prove the main properties of the Fourier transform and of the Fourier cotransform. In particular, the Fourier transform $\mathcal{F}$ and the Fourier cotransform $\overline{\mathcal{F}}$ are injective involutive homomorphisms from the involutive Banach algebra $\mathrm{L}^1(G)$ to the involutive Banach algebra $\mathcal{C}_0(\widehat{G};\mathbb{C})$ of continuous functions on $\widehat{G}$ that tend to zero at infinity. In particular for any two functions $f, g \in \mathrm{L}^1(G)$, we have

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g), \qquad \overline{\mathcal{F}}(f * g) = \overline{\mathcal{F}}(f)\overline{\mathcal{F}}(g),$$

and

$$\mathcal{F}(f^*) = (\mathcal{F}(f))^*, \qquad \overline{\mathcal{F}}(f^*) = (\overline{\mathcal{F}}(f))^*.$$

In Section 10.4 we discuss thoroughly the Fourier transform $\mathcal{F}$ on $\mathrm{L}^2(G)$ and the Fourier cotransform $\overline{\mathcal{F}}$ on $\mathrm{L}^2(\widehat{G})$ for a *finite* abelian group. In this case it is possible to work out directly Fourier inversion, the Plancherel theorem, and the convolution rule.

In Section 10.5 we make a brief excursion into number theory. If we consider the multiplicative group $G = (\mathbb{Z}/m\mathbb{Z})^*$ of units of the group $\mathbb{Z}/m\mathbb{Z}$, then its characters are the *Dirichlet characters*. They can be extended to $\mathbb{Z}$ and are called *Dirichlet characters* modulo m. It turns out the Fourier inversion formula for functions over $(\mathbb{Z}/m\mathbb{Z})^*$ is one of the steps in the proof of Dirichlet's famous theorem on arithmetic progressions of integers $mk + \ell$ with $\gcd(\ell, m) = 1$ and $k \in \mathbb{N}$, that says that such a sequence contains infinitely many primes. We briefly discuss this fascinating result.

If G is a finite abelian group, it is possible to formulate the Fourier transform $\mathcal{F}$ on $\mathrm{L}^2(G)$ and the Fourier cotransform $\overline{\mathcal{F}}$ on $\mathrm{L}^2(\widehat{G})$ in terms of matrices. This is achieved in Section 10.6. The details are a bit technical due to the appearance of various dual spaces. If we denote the vector space of functions from G to $\mathbb{C}$ as $[G \to \mathbb{C}]$ (which is equal to $\mathrm{L}^1(G)$ and $\mathrm{L}^2(G)$), then it turns out that the key is to extend $\mathcal{F}$ to a bilinear form on $[G \to \mathbb{C}]^*$, and to extend $\overline{\mathcal{F}}$ to a bilinear form on $[G \to \mathbb{C}]^{**}$. The matrices associated with these bilinear forms are called *Fourier matrices*, and they are mutual inverses.

In Section 10.7 we consider the special case of the group $G = \mathbb{Z}/n\mathbb{Z}$. We obtain the *discrete Fourier transform* and the *discrete Fourier cotransform* or *inverse discrete Fourier transform*. The Fourier matrix F is particularly interesting, as it is a Vandermonde matrix determined by the primitive nth root of unity $\omega = e^{-2\pi i/n}$ and its powers. Convolution of two sequences f and g in $\mathbb{C}^n$ can be expressed as $H(f)g$, where $H(f)$ is a *circulant matrix*. The matrix $H(f)$ has the remarkable property that its eigenvectors are the columns of the matrix $\overline{F}$, with corresponding eigenvalues the entries in the vector $n\widehat{f}$ (where $\widehat{f}$ is the discrete Fourier transform of f). As a consquence, we obtain another proof of the convolution rule.

Section 10.8 discusses Plancherel's theorem and Fourier inversion.

Let G be a locally compact abelian group equipped with a Haar measure λ. In general, given a function $f \in \mathrm{L}^1(G)$, its Fourier transform $\mathcal{F}(f)$ does not belong to $\mathrm{L}^1(\widehat{G})$.

Plancherel's theorem (Theorem 10.27) asserts that there is a Haar measure $\widehat{\lambda}$ on the dual group $\widehat{G}$ such that the map $f \mapsto \mathcal{F}(f)$ sends $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$ into $\mathrm{L}^2(\widehat{G})$, and has a unique extension which is an isometry from $\mathrm{L}^2(G)$ to $\mathrm{L}^2(\widehat{G})$.

One should realize that Theorem 10.27 does not say that the Fourier transform $\mathcal{F}$ (or the Fourier cotransform $\overline{\mathcal{F}}$) is defined on $\mathrm{L}^2(G)$, because in general the integral will not converge for f outside of $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$. What is happening is more subtle. It is always possible by using a limit process to define the Fourier transform of any $f \in \mathrm{L}^2(G)$, and this extension of $\mathcal{F}$ to $\mathrm{L}^2(G)$ is an isometry.

Plancherel's theorem has an interesting corollary when G is compact and abelian. If G is a compact abelian group endowed with a Haar measure λ normalized so that G has measure $\lambda(G) = 1$, then $\widehat{G}$ is a Hilbert basis for $\mathrm{L}^2(G)$ (it is orthonormal and dense in $\mathrm{L}^2(G)$).

The Pontrjagin duality theorem is presented in Section 10.9. This is the most important and most beautiful theorem in the theory of locally compact abelian groups.

Let G be a locally compact abelian group endowed with a Haar measure λ, let $\widehat{G}$ be its dual group endowed with the associated Haar measure $\widehat{\lambda}$ (see Definition 10.19), and let $\widehat{\widehat{G}}$ be its double dual endowed with the associated measure $\widehat{\widehat{\lambda}}$. The Pontrjagin duality theorem asserts two facts:

(1) The map $\eta\colon G \to \widehat{\widehat{G}}$ is an isomorphism and a homeomorphism between the topological groups G and $\widehat{\widehat{G}}$.

(2) If we identify G and $\widehat{\widehat{G}}$ using the isomorphism η, then the extension $\mathcal{F}\colon \mathrm{L}^2(G) \to \mathrm{L}^2(\widehat{G})$ of the Fourier transform to $\mathrm{L}^2(G)$ and the extension $\overline{\mathcal{F}}\colon \mathrm{L}^2(\widehat{G}) \to \mathrm{L}^2(G)$ of the Fourier cotransform to $\mathrm{L}^2(\widehat{G})$ are mutual inverses. In particular, Fourier inversion holds; that is,
$$f = (\overline{\mathcal{F}} \circ \mathcal{F})(f) \circ \eta, \qquad \text{for all } f \in \mathrm{L}^2(G).$$

As a corollary of the Pontrjagin duality theorem we can show that Fourier inversion holds for an interesting class of functions. We define $B(G)$ as the set of functions
$$B(G) = \{f \in \mathrm{L}^1(G) \mid \mathcal{F}(f) \in \mathrm{L}^1(\widehat{G})\}.$$
The restriction of $\mathcal{F}$ to $B(G)$ is a bijection from $B(G)$ to $B(\widehat{G})$, whose inverse is the restriction of $\overline{\mathcal{F}}$ to $B(\widehat{G})$.

Another corollary is that for any locally compact abelian group G, the group G is discrete if and only if $\widehat{G}$ is compact (and by duality, G is compact if and only if $\widehat{G}$ is discrete).

The dual group $\widehat{G}$ was first defined by Pontrjagin (1934) and van Kampen (1935). Versions of the duality theorem were also first proven by Pontrjagin and van Kampen. The first proof of the general version of Pontrjagin duality appears to have been published by André Weil [105] (Chapter VI, Section 28). The definition of the Fourier transform on an arbitrary locally compact group is due to Weil [105] (Chapter VI, Section 30). In this same section Weil proves versions of Plancherel's theorem and of the Pontrjagin duality theorem.

Our exposition relies heavily on Folland [33] and Bourbaki [9], which is even more abstract than Folland. A more elementary presentation (dealing with σ-compact, metrizable, locally compact, abelian groups) can be found in Deitmar [19], which constitutes a very good warm up for the more general treatment given in this chapter.

10.1 Characters and The Dual Group

The dual of a commutative locally compact group G is defined in terms of certain homomorphisms $\chi\colon G \to \mathbf{U}(1)$ called characters. Even though G is commutative, we use a multiplicative notation for the group operation.

Definition 10.1. Let G be a commutative locally compact group (with identity element e). A *character*[1] is a continuous homomorphism $\chi\colon G \to \mathbf{U}(1)$, that is, we have

$$\begin{aligned} \chi(ab) &= \chi(a)\chi(b), && \text{for all } a, b \in G \\ |\chi(a)| &= 1, && \text{for all } a \in G. \end{aligned}$$

The set of characters of G is denoted by $\widehat{G}$.

The characters of G satisfy the following properties.

Proposition 10.1. *Let G be a commutative locally compact group. The following properties hold:*

(1) For every character $\chi \in \widehat{G}$, we have $\chi(e) = 1$.

(2) For every character $\chi \in \widehat{G}$, for every $a \in G$,

$$\chi(a^{-1}) = (\chi(a))^{-1}.$$

(3) For every character $\chi \in \widehat{G}$, for every $a \in G$,

$$\chi(a^{-1}) = \overline{\chi(a)}.$$

Proof. Since χ is a homomorphism, we have $\chi(e) = 1$, because $\chi(e) = \chi(ee) = \chi(e)\chi(e)$, and since $|\chi(e)| = 1$, we deduce that $\chi(e) = 1$. We also have

$$1 = \chi(aa^{-1}) = \chi(a)\chi(a^{-1})$$

and

$$1 = \chi(a^{-1}a) = \chi(a^{-1})\chi(a),$$

so $\chi(a^{-1}) = (\chi(a))^{-1}$, and since $\chi(a) \in \mathbb{C}$ and $|\chi(a)| = 1$, we have $(\chi(a))^{-1} = \overline{\chi(a)}$, so we get

$$\chi(a^{-1}) = \overline{\chi(a)}, \qquad \text{for all } a \in G. \qquad \square$$

The following fact will be needed in the proof of Theorem 10.6.

[1] Sometimes, to emphasize that their range is $\mathbf{U}(1)$, they are called *unitary characters*.

Proposition 10.2. *If a group homomorphism $\chi\colon G \to \mathbb{C}$ is bounded, which means that there is some $C > 0$ such that $|\chi(g)| \leq C$ for all $g \in G$, then $|\chi(g)| = 1$ for all $g \in G$, that is, $\chi\colon G \to \mathbf{U}(1)$.*

Proof. Suppose that $|\chi(g)| \neq 1$ for some $g \in G$. Since χ is a homomorphism, $\chi(g^{-1}) = (\chi(g))^{-1}$, so either $|\chi(g)| > 1$ or $|\chi(g^{-1})| > 1$, and we may assume that $|\chi(g)| > 1$. Since χ is a homomorphism, $\chi(g^n) = (\chi(g))^n$ for all $n \geq 0$, and since $|\chi(g)| > 1$, for n large enough we obtain $|\chi(g^n)| = |\chi(g)|^n > C$, contradicting the fact that χ is bounded. $\square$

Our next goal is to make the set of characters into a commutative locally compact group. The first step is to define a multiplication operation on characters. We proceed as follows.

Given two characters $\chi_1, \chi_2 \in \widehat{G}$, we define $\chi_1\chi_2$ by

$$(\chi_1\chi_2)(a) = \chi_1(a)\chi_2(a), \qquad a \in G.$$

Since $\mathbb{C}$ is commutative, we have

$$\begin{aligned}(\chi_1\chi_2)(ab) &= \chi_1(ab)\chi_2(ab)\\ &= \chi_1(a)\chi_1(b)\chi_2(a)\chi_2(b)\\ &= \chi_1(a)\chi_2(a)\chi_1(b)\chi_2(b)\\ &= (\chi_1\chi_2)(a)(\chi_1\chi_2)(b)\end{aligned}$$

so $\chi_1\chi_2$ is a homomorphism. Since $|\chi_1(a)| = |\chi_2(a)| = 1$, we also have

$$|(\chi_1\chi_2)(a)| = |\chi_1(a)\chi_2(a)| = |\chi_1(a)||\chi_2(a)| = 1.$$

Since χ_1 and χ_2 are continuous, and multiplication on $\mathbb{C}$ is continuous, the map $\chi_1\chi_2$ is continuous. Therefore $\chi_1\chi_2$ is a character.

If we define $\overline{\chi}$ by

$$\overline{\chi}(a) = \overline{\chi(a)}, \qquad a \in A,$$

then we have

$$\overline{\chi}(ab) = \overline{\chi(ab)} = \overline{\chi(a)\chi(b)} = \overline{\chi(a)}\,\overline{\chi(b)} = \overline{\chi}(a)\overline{\chi}(b),$$

and

$$|\overline{\chi}(a)| = |\overline{\chi(a)}| = |\chi(a)| = 1,$$

and since χ is continuous, and conjugation on $\mathbb{C}$ is continuous, $\overline{\chi}$ is obviously continuous. Thus $\overline{\chi}$ is a character. Finally, since $|\chi(a)| = 1$, Proposition 10.1(3) implies that

$$(\chi\overline{\chi})(a) = \chi(a)\overline{\chi(a)} = 1,$$

and that

$$(\overline{\chi}\chi)(a) = \overline{\chi(a)}\chi(a) = 1.$$

In summary, we proved the following result.

Proposition 10.3. *The set $\widehat{G}$ of characters with the multiplication operation $(\chi_1, \chi_2) \mapsto \chi_1\chi_2$ defined above is a commutative group with the constant function from G to $\mathbf{U}(1)$ with value 1 as identity. The inverse operation is $\chi \mapsto \overline{\chi}$.*

The next step is to give $\widehat{G}$ a topology that will make it a locally compact group. Although it is far from obvious why this works, since $\widehat{G}$ consists of continuous functions from G to $\mathbf{U}(1)$, we give it the compact-open topology (see Definition 2.11). A subbasis for this topology consists of the sets

$$S(K, U) = \{f \mid f \in \widehat{G},\ f(K) \subseteq U\},$$

where K is any compact subset of G and U is any open subset of $\mathbf{U}(1)$; see Definition 2.11. The group operations (multiplication and inversion) are continuous in this topology, although this not immediately obvious.

Since $\widehat{G}$ is a group, it suffices to show that for every open subset U of $\mathbf{U}(1)$ containing 1, for every open subset $S(K, U)$ containing the constant function 1, there is some subset $S(K_1, V_1)$ such that $S(K_1, V_1)S(K_1, V_1) \subseteq S(K, U)$, and that there is some subset $S(K_2, V_2)$ such that $(S(K_2, V_2))^{-1} \subseteq S(K, U)$. Since $\mathbf{U}(1)$ is a topological group, there is some open subset V_1 of U containing 1 such that $V_1V_1 \subseteq U$, and then $1 \in S(K, V_1)$ and $S(K, V_1)S(K, V_1) \subseteq S(K, U)$. Since inversion in G is continuous, there is also an open subset V_2 of $\mathbf{U}(1)$ such that $V_2^{-1} \subseteq U$, and then $1 \in S(K, V_2^{-1})$ and $(S(K, V_2^{-1}))^{-1} \subseteq S(K, U)$.

Since $\mathbf{U}(1)$ is Hausdorff, the compact-open topology is Hausdorff. Indeed, if $\chi_1, \chi_2 \in \widehat{G}$ and if $\chi_1 \neq \chi_2$, then there is some $a \in G$ such that $\chi_1(a) \neq \chi_2(a)$, and since $\mathbf{U}(1)$ is Hausdorff there exists two disjoint open subsets U_1, U_2 with $\chi_1(a) \in U_1$ and $\chi_2(a) \in U_2$. Then $S(\{a\}, U_2) \cap S(\{a\}, U_2) = \emptyset$, with $\chi_1 \in S(\{a\}, U_1)$ and $\chi_2 \in S(\{a\}, U_2)$.[2]

In summary, we proved the following result.

Proposition 10.4. *Let G be a locally compact abelian group. The set $\widehat{G}$ of characters of G with the multiplication $(\chi_1, \chi_2) \mapsto \chi_1\chi_2$ given by*

$$(\chi_1\chi_2)(a) = \chi_1(a)\chi_2(a), \qquad a \in G,$$

and endowed with the compact-open topology, is an abelian topological group.

Definition 10.2. Let G be a locally compact abelian group. The topological abelian group $\widehat{G}$ of characters of G is called the *dual* (or *Pontrjagin dual*) of G.

Proving directly that $\widehat{G}$ is locally compact is not so easy. André Weil gives a clever proof in [105]. Another way to proceed is to use the fact that G is endowed with a Haar measure λ and to prove that $\widehat{G}$ is homeomorphic to $\mathsf{X}(\mathrm{L}^1(G))$, a remarkable result showing that the group characters of G and the algebra characters of $\mathrm{L}^1(G)$ are in some sense equivalent.

[2] Recall that a compact subset K of G is a subset such that every open cover of K by open subsets in G contains a finite subfamily covering K. Obviously every finite subset is compact.

The proof that $\widehat{G}$ and $\mathsf{X}(\mathrm{L}^1(G))$ are homeomorphic is quite technical. Folland [33] (Chapter 4) gives the main idea, but does not prove that the map is injective, nor that its inverse is continuous. Bourbaki [9] (Chapter 2, Section 1) gives a complete, but terse proof. The details of the proof are not illuminating so we will only indicate its main ideas.

The crucial step is to show that every group character $\chi \in \widehat{G}$ induces an algebra character $\zeta_\chi \in \mathsf{X}(\mathrm{L}^1(G))$; namely, for every $f \in \mathrm{L}^1(G)$,

$$\zeta_\chi(f) = \int \chi(a) f(a)\, d\lambda(a).$$

Actually, every group character $\chi \in \widehat{G}$ induces an algebra character $\zeta_\chi \in \mathsf{X}(\mathcal{M}^1(G))$; namely, the map given by

$$\zeta_\chi(\mu) = \int \chi(a)\, d\mu(a)$$

for every complex measure $\mu \in \mathcal{M}^1(G)$ is a character of the algebra $\mathcal{M}^1(G)$. This more general fact will be needed.

Recall from Example 9.6(3) that $\mathcal{M}^1(G)$ is a unital Banach algebra under convolution, with involution given by $\mu^* = \overline{\check{\mu}}$. The identity element of $\mathcal{M}^1(G)$ is the Dirac measure δ_e. The algebra $\mathrm{L}^1(G)$ is also a Banach algebra under convolution, with involution given by $f \mapsto \overline{\check{f}}$, but unless G is discrete, it is nonunital. However, $\mathrm{L}^1(G)$ is embedded in $\mathcal{M}^1(G)$ as a closed Banach involutive subalgebra (via $f \mapsto f\, d\lambda$), so we can consider the unital Banach involutive subalgebra $\mathrm{L}^1(G) \oplus \mathbb{C}\delta_e$. Any character $\chi\colon \mathrm{L}^1(G) \to \mathbb{C}$ extends uniquely to a character $\chi'\colon (\mathrm{L}^1(G) \oplus \mathbb{C}\delta_e) \to \mathbb{C}$ by letting

$$\chi'(f\, d\lambda + \alpha\delta_e) = \chi(f) + \alpha.$$

Proposition 10.5. *Let G be any locally compact abelian group with a left Haar measure λ. For every character $\chi \in \widehat{G}$, the map ζ_χ given by*

$$\zeta_\chi(\mu) = \int \chi(a)\, d\mu(a) \qquad \text{for all } \mu \in \mathcal{M}^1(G)$$

is a hermitian character (an algebra homomorphism such that $\zeta_\chi(\overline{\check{\mu}}) = \overline{\zeta_\chi(\mu)}$) of the algebra $\mathcal{M}^1(G)$. By restriction to $\mathrm{L}^1(G)$, the map ζ_χ given by

$$\zeta_\chi(f) = \int \chi(a) f(a)\, d\lambda(a) \qquad \text{for all } f \in \mathrm{L}^1(G)$$

is a hermitian character of $\mathrm{L}^1(G)$ not equal to the zero function.

Proof. Let μ, ν be any two complex measures in $\mathcal{M}^1(G)$. The function $\chi\colon G \to \mathbb{C}$ is continuous and bounded, so by the corollary of Proposition 8.54, we have

$$\begin{aligned}
\zeta_\chi(\mu * \nu) &= \int \chi(a)\, d(\mu * \nu)(a) \\
&= \int\int \chi(ab)\, d\mu(a)\, d\nu(b) \\
&= \int\int \chi(a)\chi(b)\, d\mu(a)\, d\nu(b) \\
&= \left(\int \chi(a)\, d\mu(a)\right)\left(\int \chi(b)\, d\nu(b)\right) \\
&= \zeta_\chi(\mu)\zeta_\chi(\nu).
\end{aligned}$$

Recall from Proposition 7.24,

$$\int \varphi\, d\overline{\mu} = \overline{\int \overline{\varphi(s)}\, d\mu(s)},$$

and by Proposition 8.45,

$$\int \varphi\, d\check{\mu} = \int \check{\varphi}\, d\mu,$$

with $\check{\varphi}(a) = \varphi(a^{-1})$ for all $a \in G$. Thus we have

$$\begin{aligned}
\zeta_\chi(\overline{\check{\mu}}) &= \int \chi(a)\, d\overline{\check{\mu}} \\
&= \overline{\int \overline{\chi(a)}\, d\check{\mu}}, \qquad \text{by Proposition 7.24} \\
&= \overline{\int \chi(a^{-1})\, d\check{\mu}}, \qquad \text{by Proposition 10.1(3)} \\
&= \overline{\int \chi(a)\, d\mu(a)} = \overline{\zeta_\chi(\mu)}, \qquad \text{by Proposition 8.45.}
\end{aligned}$$

Thus ζ_χ is a Hermitian character.

By restriction to $\mathrm{L}^1(G)$, we obtain an algebra homomorphism, with

$$\zeta_\chi(f) = \int f(a)\chi(a)\, d\lambda(a).$$

We need to prove that ζ_χ is not the zero function. The proof is not trivial. One method is to observe that this is a special case of the fundamental fact that there is a bijection between the set of unitary representations of the group G and the set of nondegenerate representations of the algebra $\mathrm{L}^1(G)$. This connection will be discussed in Section 12.3. It is the method followed by Folland [33] (Chapter 3, Theorem 3.9, and Chapter 4, Section 4.1).

The characters $\chi\colon G \to \mathbf{U}(1)$ are indeed unitary representations of G. The other method used by Bourbaki [9] (Chapter 2, Section 1, No 1) is to use Corollary 8.52. We can choose a filter where the measures in $\mathcal{M}^1(G)$ are of the form $fd\lambda$ with $f \in \mathcal{K}_{\mathbb{C}}(G)$, so that if $fd\lambda$ tends to δ_e, then $\zeta_\chi(f)$, which is equal to $\zeta_\chi(fd\lambda)$, tends to $\zeta_\chi(\delta_e) = 1 \neq 0$. $\square$

The following deep result is obtained.

Theorem 10.6. *Let G be a locally compact abelian group (equipped with a Haar measure λ). The map $j\colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$ given by*

$$j(\chi)(f) = \zeta_\chi(f) = \int \chi(a)f(a)\, d\lambda(a), \qquad \chi \in \widehat{G},\ f \in \mathrm{L}^1(G),$$

is a homeomorphism of $\widehat{G}$ onto $\mathsf{X}(\mathrm{L}^1(G))$.

Proof sketch. We follow Bourbaki [9] (Chapter 2, Section 1, No 1). In this proof we view every function in $\mathrm{L}^1(G)$ as a measure in $\mathcal{M}^1(G)$. Recall from Proposition 8.44(2) that $(\delta_a * f)(s) = (\lambda_a f)(s) = f(a^{-1}s)$. To show that j is injective, for any $\chi \in \widehat{G}$, we use the fact that for any $f \in \mathrm{L}^1(G)$ and any $a \in G$, we have

$$\zeta_\chi(\delta_a * f) = \zeta_\chi(\delta_a)\zeta_\chi(f) = \chi(a)\zeta_\chi(f),$$

because $\zeta_\chi(\delta_a) = \int \chi(b)\, d\delta_a(b) = \chi(a)$. If $\zeta_{\chi_1} = \zeta_{\chi_2}$, then the equation

$$\zeta_\chi(\delta_a * f) = \chi(a)\zeta_\chi(f), \quad \text{for all } a \in G \text{ and all } f \in \mathrm{L}^1(G)$$

applied to χ_1 and χ_2 shows that

$$\chi_1(a)\zeta_{\chi_1}(f) = \chi_2(a)\zeta_{\chi_2}(f), \quad \text{for all } a \in G \text{ and all } f \in \mathrm{L}^1(G).$$

Since by Proposition 10.5 there is some function $f \in \mathrm{L}^1(G)$ such that $\zeta_{\chi_1}(f) = \zeta_{\chi_2}(f) \neq 0$, we deduce that $\chi_1 = \chi_2$, so j is injective.

To prove surjectivity we use the following trick. For any $\zeta \in \mathsf{X}(\mathrm{L}^1(G))$ other than the zero function, pick some function $f \in \mathrm{L}^1(G)$ such that $\zeta(f) \neq 0$. Define $\chi\colon G \to \mathbb{C}$ by

$$\chi(a) = \zeta(\delta_a * f)/\zeta(f).$$

The goal is to show that χ is a character of G such that $\zeta = \zeta_\chi$.

It can be shown that the map $a \mapsto \delta_a * f$ is continuous, so χ is continuous. We also have

$$|\chi(a)| \leq \|\delta_a * f\|_1 / |\zeta(f)| = \|f\|_1 / |\zeta(f)|.$$

Therefore, χ is continuous and bounded. The next step is the most technical part of the proof. It can be shown that there is a filter base $\mathcal{B}$ of $e \in G$ consisting of compact subsets,

such that, for every $V \in \mathcal{B}$, there is a continuous function g_V which is positive, zero outside of V, and with $\int g_V \, d\lambda = 1$, and such that

$$\delta_a * f = \lim \delta_a * g_V * f.$$

Then, since $\zeta(\delta_a * g_V * f) = \zeta(\delta_a * g_V)\zeta(f)$, and by definition $\chi(a) = \zeta(\delta_a * f)/\zeta(f)$, we get

$$\chi(a) = \lim \zeta(\delta_a * g_V),$$

and for any $h \in \mathrm{L}^1(G)$,

$$\zeta(\delta_a * h) = \lim \zeta(\delta_a * g_V * h) = (\lim \zeta(\delta_a * g_V))\zeta(h) = \chi(a)\zeta(h).$$

Using Proposition 8.44(3) ($\delta_{ab} = \delta_a * \delta_b$) and the above equation, (with $h = \delta_b * f$), for all $a, b \in G$, we have

$$\chi(ab) = \zeta(\delta_a * \delta_b * f)/\zeta(f) = \zeta(\delta_a)\zeta(\delta_b * f)/\zeta(f) = \chi(a)\zeta(\delta_b * f)/\zeta(f) = \chi(a)\chi(b),$$

which proves that χ is a homomorphism. But since χ is also bounded, by Proposition 10.2, it is a (unitary) character of G. Since for any $f \in \mathrm{L}^1(G)$ we have $(\delta_a * f)(s) = f(a^{-1}s)$, (see just after Definition 8.24), we have

$$(g * f)(s) = \int g(a)f(a^{-1}s)\, d\lambda(a) = \int (\delta_a * f)(s)g(a)d\lambda(a). \tag{$\dagger$}$$

(Alternatively, see Bourbaki [7] (Chapter VIII, Section 1, Proposition 7).

The equation

$$\zeta(g * f) = \int \zeta(\delta_a * f)g(a)\, d\lambda(a)$$

is needed to finish the proof of surjectivity. Since by Theorem 9.19(1), every character $\zeta \in \mathsf{X}(\mathrm{L}^1(G))$ is continuous, ζ is a continuous linear form on $\mathrm{L}^1(G)$, so $\zeta \in \mathrm{L}^1(G)'$, the dual of $\mathrm{L}^1(G)$. Theorem 5.51 asserts that $\mathrm{L}^\infty(G)$ and $\mathrm{L}^1(G)'$ are isomorphic, and more precisely that there is some (unique) $\varphi \in \mathrm{L}^\infty(G)$ such that

$$\zeta(f) = \int f(s)\varphi(s)\, d\lambda(s) \quad \text{for all } f \in \mathrm{L}^1(G). \tag{$\dagger\dagger$}$$

Then using Fubini we have

$$\begin{aligned}
\zeta(g * f) &= \int (g * f)(s)\varphi(s)\, d\lambda(s) \\
&= \int\int (\delta_a * f)(s)g(a)\varphi(s)\, d\lambda(a)d\lambda(s) && \text{by } (\dagger) \\
&= \int \Big(\int (\delta_a * f)(s)\varphi(s)\, d\lambda(s)\Big) g(a)d\lambda(a) \\
&= \int \zeta(\delta_a * f)g(a)\, d\lambda(a). && \text{by } (\dagger\dagger)
\end{aligned}$$

Using the above fact, we get

$$\begin{aligned}\zeta(g)\zeta(f) &= \zeta(g * f)\\ &= \int \zeta(\delta_a * f)g(a)\,d\lambda(a)\\ &= \zeta(f)\int \chi(a)g(a)\,d\lambda(a)\\ &= \zeta_\chi(g)\zeta(f).\end{aligned}$$

If we pick f such that $\zeta(f) \neq 0$, we deduce that $\zeta = \zeta_\chi$, establishing the fact that j is surjective. Since j is injective and surjective, it is a bijection. It remains to prove that j is a homeomorphism. We skip this proof, referring the reader to Bourbaki [9] (Chapter 2, Section 1, No. 1). $\square$

As a corollary of Theorem 10.6, since by Proposition 9.20 the space $\mathsf{X}(\mathrm{L}^1(G))$ is locally compact, we have the following fact.

Corollary 10.7. *Let G be a locally compact abelian group. The group $\widehat{G}$ of characters of G with the compact-open topology is locally compact.*

Theorem 10.6, Proposition 10.5, and Proposition 9.41, also imply the following result. Recall from Definition 9.23 that the enveloping C^*-algebra of $\mathrm{L}^1(G)$ is denoted by $\mathrm{St}(G)$.

Proposition 10.8. *Let G be a locally compact abelian group. The characters of the algebra $\mathrm{L}^1(G)$ are hermitian. There is a homeomorphism between the set of characters $\mathsf{X}(\mathrm{St}(G))$ of the enveloping C^*-algebra $\mathrm{St}(G)$ of $\mathrm{L}^1(G)$ and the set of characters $\mathsf{X}(\mathrm{L}^1(G))$ of $\mathrm{L}^1(G)$.*

In summary, the spaces $\widehat{G}$, $\mathsf{X}(\mathrm{L}^1(G))$, and $\mathsf{X}(\mathrm{St}(G))$, are homeomorphic.

It is instructive to figure out the duals of various familiar locally compact abelian groups.

10.2 Characters Groups of some LCA Groups

Proposition 10.9. *The locally compact abelian groups, $\mathbb{Z}, \mathbb{T}$, $\mathbb{Z}/n\mathbb{Z}$, and $\mathbb{R}$, have the following characters and dual groups.*

(1) *For $\mathbb{Z}$, the homomorphisms $m \mapsto e^{im\theta} = (e^{i\theta})^m$, for any fixed $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, and any $m \in \mathbb{Z}$. Therefore, the dual group $\widehat{\mathbb{Z}}$ of $\mathbb{Z}$ is isomorphic to $\mathbb{T}$.*

(2) *For $\mathbb{T}$, the homomorphisms $e^{i\theta} \mapsto e^{im\theta} = (e^{im})^\theta$, for any fixed $m \in \mathbb{Z}$, and any $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Therefore, the dual group $\widehat{\mathbb{T}}$ of $\mathbb{T}$ is isomorphic to $\mathbb{Z}$.*

(3) *For $\mathbb{Z}/n\mathbb{Z}$, the homomorphisms $m \mapsto e^{2\pi imk/n} = (e^{2\pi ik/n})^m$, for any fixed $k \in \mathbb{Z}/n\mathbb{Z}$, and any $m \in \mathbb{Z}/n\mathbb{Z}$. Therefore, the dual group $\widehat{\mathbb{Z}/n\mathbb{Z}}$ of $\mathbb{Z}/n\mathbb{Z}$ is isomorphic to $\mathbb{Z}/n\mathbb{Z}$ (itself).*

(4) *For $\mathbb{R}$, the homomorphisms $x \mapsto e^{iyx} = (e^{iy})^x$, for any fixed $y \in \mathbb{R}$, and all $x \in \mathbb{R}$. Therefore, the dual group $\widehat{\mathbb{R}}$ of $\mathbb{R}$ is isomorphic to $\mathbb{R}$ (itself).*

Proof. (1) Since $\mathbb{Z}$ is a cyclic group generated by 1, every homomorphism $\varphi\colon \mathbb{Z} \to \mathbf{U}(1)$ satisfies the equation

$$\varphi(m) = (\varphi(1))^m, \qquad \text{for all } m \in \mathbb{Z}.$$

Thus φ is uniquely determined by picking $\varphi(1) = e^{i\theta}$ in $\mathbf{U}(1)$.

The characters of $\mathbb{T}$ are easily obtained from the characters of $\mathbb{R}$, so we consider (4) next.

(4) Folland has a particularly nice proof of (4). Any homomorphism $\varphi\colon \mathbb{R} \to \mathbf{U}(1)$ satisfies $\varphi(0) = 1$, and since φ is continuous, there is some $a > 0$ such that $\int_0^a \varphi(t)\, dt \neq 0$. Let $A = \int_0^a \varphi(t)\, dt$. Since $\varphi(x + t) = \varphi(x)\varphi(t)$, we have

$$A\varphi(x) = \left(\int_0^a \varphi(t)\ dt\right)\varphi(x) = \int_0^a \varphi(t)\varphi(x)\, dt = \int_0^a \varphi(t + x)\, dt = \int_x^{a+x} \varphi(u)\, du.$$

It follows that φ is differentiable, and we have

$$\varphi'(x) = A^{-1}(\varphi(a + x) - \varphi(x)) = A^{-1}(\varphi(a)\varphi(x) - \varphi(x)) = A^{-1}(\varphi(a) - 1)\varphi(x).$$

If we let $c = A^{-1}(\varphi(a) - 1)$, then (using the fact that $\varphi(0) = 1$) we deduce that

$$\varphi(x) = e^{cx}.$$

Since $|\varphi(x)| = 1$, c must be a pure imaginary number of the form $c = iy$, with $y \in \mathbb{R}$, so $\varphi(x) = e^{iyx}$.

(2) Recall that we have the surjective homomorphism $\sigma\colon \mathbb{R} \to \mathbb{T}$ given by $\sigma(\theta) = e^{i\theta}$, and that the kernel of σ is $2\pi\mathbb{Z}$. Therefore, there is a bijection between the continuous homomorphisms $\varphi\colon \mathbb{T} \to \mathbb{U}(1)$ and the continuous homomorphisms $\psi\colon \mathbb{R} \to \mathbb{U}(1)$ such that $\operatorname{Ker}\psi = 2\pi\mathbb{Z}$. By (4) the homomorphisms ψ are of the form $\psi(\theta) = (e^{iy})^\theta$ for some $y \in \mathbb{R}$, and for ψ to have kernel $2\pi\mathbb{Z}$, it must be the case that $\theta \equiv 0 \pmod{2\pi}$ implies that $y\theta \equiv 0 \pmod{2\pi}$, so $y = m \in \mathbb{Z}$, and $\varphi(e^{i\theta}) = (e^{im})^\theta$.

(3) We have the canonical surjective homomorphism $pr\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ whose kernel is $\operatorname{Ker} pr = \mathbb{Z}/n\mathbb{Z}$. It follows that there is a bijection between the homomorphisms $\varphi\colon \mathbb{Z}/n\mathbb{Z} \to \mathbf{U}(1)$ and the homomorphisms $\psi\colon \mathbb{Z} \to \mathbf{U}(1)$ such that $\operatorname{Ker}\psi = \mathbb{Z}/n\mathbb{Z}$. By (1) the homomorphisms ψ are of the form $\psi(m) = e^{im\theta} = (e^{i\theta})^m$, and for ψ to have kernel $\mathbb{Z}/n\mathbb{Z}$, it must be the case that $m \equiv 0 \pmod{n}$ implies that $\theta m \equiv 0 \pmod{2\pi}$, so for $m = dn$ with $d \in \mathbb{Z}$, we must have $\theta dn \equiv 0 \pmod{2\pi}$, which implies that $\theta = 2\pi k/n$ with $k \in \mathbb{Z}/n\mathbb{Z}$, and then $\psi(m) = (e^{2\pi ik/n})^m$. ∎

Remark: The proof of (4) shows that *any* continuous homomorphism $\varphi\colon \mathbb{R} \to \mathbb{C}^*$ is of the form $\varphi(x) = e^{cx}$ for some complex number $c \in \mathbb{C}$ (where $\mathbb{C}^*$ is the group of nonzero complex numbers under multiplication).

Given a locally compact abelian group G and its dual $\widehat{G}$, there is a canonical pairing $\langle -, - \rangle \colon G \times \widehat{G} \to \mathbb{T}$ given by evaluation,

$$\langle a, \chi \rangle = \chi(a), \qquad a \in G,\ \chi \in \widehat{G}.$$

In practice the dual group $\widehat{G}$ is only determined up to isomorphism. What this means is that if G_1 and G_2 are isomorphic to $\widehat{G}$, the two pairings are usually different. This issue comes up with the group $\mathbb{R}$. We figured out that one of the groups, $\widehat{\mathbb{R}}_1$, isomorphic to $\widehat{\mathbb{R}}$ consists of the characters $\chi^1(x) = (e^{iy})^x$, with $y \in \mathbb{R}$. The pairing is given by

$$\langle x, \chi_1 \rangle_1 = (e^{iy})^x.$$

However, another isomorphic copy $\widehat{\mathbb{R}}_2$ of $\widehat{\mathbb{R}}$ is often used, in which the characters are given by

$$\chi_2(x) = (e^{2\pi iy})^x,$$

with $x \in \mathbb{R}$, in which case the pairing is

$$\langle x, \chi_2 \rangle_2 = (e^{2\pi iy})^x.$$

A similar problem comes up with the groups $\widehat{\mathbb{Z}}$ and $\widehat{\mathbb{T}}$. The characters of $\mathbb{Z}$ can also be viewed as the homomorphisms $m \mapsto e^{2\pi im\theta} = (e^{2\pi i\theta})^m$, with $\theta \in \mathbb{R}/\mathbb{Z}$, and the characters of $\mathbb{T}$ can be viewed as the homomorphisms $e^{2\pi i\theta} \mapsto e^{2\pi im\theta} = (e^{2\pi im})^\theta$, with $\theta \in \mathbb{R}/\mathbb{Z}$.

This subtle issue is related to the choice of the normalization of the Haar measures on G and $\widehat{G}$ and will come up when we consider the inverse Fourier transform.

Products behave well with respect to characters.

Proposition 10.10. *If $G_1, \ldots, G_n$ are locally compact abelian groups, then*

$$(G_1 \times \cdots \times G_n)\widehat{\ } \cong \widehat{G_1} \times \cdots \times \widehat{G_n}.$$

Proof. Every tuple of characters $(\chi_1, \ldots, \chi_n)$ with $\chi_i \in \widehat{G_i}$ induces a character on $G_1 \times \cdots \times G_n$, by

$$(\chi_1, \ldots, \chi_n)(a_1, \ldots, a_n) = \chi_1(a_1) \cdots \chi_n(a_n)$$

for all $a_i \in G_i$, $i = 1, \ldots, n$. Conversely, every character $\chi \colon G_1 \times \cdots \times G_n \to \mathbb{T}$ can be written as

$$\chi(a_1, \ldots, a_n) = (\chi_1, \ldots, \chi_n)(a_1, \ldots, a_n) = \chi_1(a_1) \cdots \chi_n(a_n),$$

with χ_i given by

$$\chi_i(a_i) = \chi(1, \ldots, 1, a_i, 1, \ldots, 1).$$

This proves the isomorphism of the proposition. ∎

Propositions 10.9 and 10.10 imply the following facts.

Corollary 10.11. *We have the following isomorphisms:*

$$\widehat{\mathbb{R}^n} \cong \mathbb{R}^n, \quad \widehat{\mathbb{T}^n} \cong \mathbb{Z}^n, \quad \widehat{\mathbb{Z}^n} \cong \mathbb{T}^n.$$

In particular, the characters in $\widehat{\mathbb{R}^n}$ are the homomorphims from $\mathbb{R}^n$ to $\mathbb{T}$ given by

$$x \mapsto e^{iy\cdot x}, \qquad x, y \in \mathbb{R}^n,$$

where $y \cdot x$ is the Euclidean product in $\mathbb{R}^n$; that is, $y \cdot x = \sum_{k=1}^n y_k x_k$.

The characters in $\widehat{\mathbb{T}^n}$ are the homomorphims from $\mathbb{T}^n$ to $\mathbb{T}$ given by

$$(e^{i\theta_1}, \ldots, e^{i\theta_n}) \mapsto e^{im\cdot\theta}, \qquad m \in \mathbb{Z}^n,\ \theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n,$$

and the characters in $\widehat{\mathbb{Z}^n}$ are the homomorphims from $\mathbb{Z}^n$ to $\mathbb{T}$ given by

$$m \mapsto e^{im\cdot\theta}, \qquad \theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n,\ m \in \mathbb{Z}^n.$$

As a corollary of Proposition 10.10, since by the structure theorem for finitely generated abelian groups, every *finite* abelian group is isomorphic to a product of cyclic groups $\mathbb{Z}/p\mathbb{Z}$, by Proposition 10.9 and Proposition 10.10, we see that every finite abelian group is isomorphic to its dual. Here we give G the discrete topology so it is automatically compact.

Proposition 10.12. *If G is a finite abelian group, then G is isomorphic to its dual $\widehat{G}$.*

This fact can also be shown more directly, and there is no canonical isomorphism; see Apostol [2] (Chapter 6, Theorem 6.8).

If the abelian group is compact, then $\mathrm{L}^2(G)$ is a subspace of $\mathrm{L}^1(G)$, and a Hilbert space, with the hermitian inner product given by

$$\langle f, g\rangle = \int f(s)\overline{g(s)}\, d\lambda(s), \qquad \text{for all } f, g \in \mathrm{L}^2(G).$$

If we assume that $\int_G 1\, d\lambda = 1$, then it is remarkable that the set of characters is an orthonormal set in $\mathrm{L}^2(G)$.

Proposition 10.13. *Let G be a compact abelian group with a Haar measure normalized so that G has measure 1. Then for any character $\chi \in \widehat{G}$, we have $\langle \chi, \chi\rangle = 1$, and for any two distinct characters $\chi_1, \chi_2 \in \widehat{G}$, we have*

$$\langle \chi_1, \chi_2\rangle = 0;$$

that is, the characters form an orthonormal set in $\mathrm{L}^2(G)$.

Proof. We have
$$\langle \chi, \chi \rangle = \int \chi(s)\overline{\chi}(s)\, d\lambda(s) = \int 1\, d\lambda(s) = 1.$$
If $\chi_1 \neq \chi_2$, then there is some $a \in G$ such that $\chi_1(a) \neq \chi_2(a)$, which is equivalent to $(\chi_1\chi_2^{-1})(a) \neq 1$. Then by using properties of characters and left invariance of the Haar measure, we have

$$\begin{aligned}
\langle \chi_1, \chi_2 \rangle &= \int \chi_1(s)\overline{\chi_2}(s)\, d\lambda(s) \\
&= \int (\chi_1\chi_2^{-1})(s)\, d\lambda(s) \\
&= \int (\chi_1\chi_2^{-1})(aa^{-1}s)\, d\lambda(s) \\
&= (\chi_1\chi_2^{-1})(a) \int (\chi_1\chi_2^{-1})(a^{-1}s)\, d\lambda(s) \\
&= (\chi_1\chi_2^{-1})(a) \int (\chi_1\chi_2^{-1})(s)\, d\lambda(s) \\
&= (\chi_1\chi_2^{-1})(a)\langle \chi_1, \chi_2 \rangle.
\end{aligned}$$

Since $(\chi_1\chi_2^{-1})(a) \neq 1$, we conclude that $\langle \chi_1, \chi_2 \rangle = 0$. ☐

Proposition 10.13 implies the following fact.

Proposition 10.14. *For any compact abelian group G, for any character $\chi \in \widehat{G}$, if $\chi \neq 1$, that is, χ is not the trivial character with constant value 1, then $\int \chi(s)\, d\lambda(s) = 0$.*

Proof. Since $\chi \neq 1$, χ is orthogonal to the trivial character 1, so
$$0 = \langle \chi, 1 \rangle = \int f\overline{1}\, d\lambda = \int f d\lambda.$$
☐

If G is a compact abelian group, it is remarkable that the orthonormal set of characters $\widehat{G}$ is a Hilbert basis of $\mathrm{L}^2(G)$. This means that the set $\widehat{G}$ is dense in $\mathrm{L}^2(G)$, so for every function $f \in \mathrm{L}^2(G)$ there is a sequence of linear combinations of characters converging to f in the L^2-norm. The proof is nontrivial, and relies on Plancherel's theorem; see Section 10.8 (it it also a corollary of the Peter–Weyl theorem; see Dieudonné [21] (Chapter XXI, Section 3)).

In particular, for $G = \mathbb{T}$, it turns out that the characters are the functions $e^{i\theta} \mapsto e^{im\theta}$, so we obtain a proof of the fact that $(e^{im\theta})_{m\in\mathbb{Z}}$ is an orthonormal system and that every function in $\mathrm{L}^1(\mathbb{T})$ (a periodic function) is given by a Fourier series.

In the case of a finite locally compact group G, we have $\mathrm{L}^1(G) = \mathrm{L}^2(G)$, the functions in this space are just finite sequences $x = (x_a)_{a\in G}$ of complex numbers indexed by G, and we can figure out explicitly what is integration, the inner product, and convolution.

Example 10.1. Let G be a locally compact abelian group. If G is finite, since G must be Hausdorff, it must have the discrete topology (because every singleton set must be closed, and since G is finite, by closure under finite unions of closed sets, every subset is closed). Thus G is actually compact. The Haar measure λ is just the counting measure, and the integral $\int x\, d\lambda$ is the sum $\sum_{a\in G} x_a$. Here we have a choice; if $|G| = n$, then we can normalize the Haar measure so that G has measure 1, or assume that it has measure n. Let us adopt the first choice, $\lambda(G) = 1$, which implies that

$$\int x_a\, d\lambda(a) = \frac{1}{|G|}\sum_{a\in G} x_a.$$

The inner product $\langle x, y\rangle$ of $x, y \in \mathrm{L}^2(G)$ is

$$\langle x, y\rangle = \int x_a \overline{y}_a\, d\lambda(a) = \frac{1}{|G|}\sum_{a\in G} x_a \overline{y_a}.$$

The convolution $x * y$ of $x, y \in \mathrm{L}^2(G)$ is given by

$$(x * y)_a = \frac{1}{|G|}\sum_{b\in G} x_b y_{b^{-1}a} = \frac{1}{|G|}\sum_{\substack{b,c\in G\\ b+c=a}} x_b y_c.$$

In the special case where $G = \mathbb{Z}/n\mathbb{Z}$, if $x = (x_0, \ldots, x_{n-1})$ and $y = (y_0, \ldots, y_{n-1})$, for $k = 0, \ldots, n-1$, we have

$$(x * y)_k = \frac{1}{n}\sum_{\substack{i,j\in\mathbb{Z}/n\mathbb{Z}\\ i+j\equiv k \pmod n}} x_i y_j.$$

For example, if $G = \mathbb{Z}/3\mathbb{Z}$, the convolution of $x = (x_0, x_1, x_2)$ and $y = (y_0, y_1, y_2)$ is the sequence

$$x * y = \frac{1}{3}(x_0y_0 + x_1y_2 + x_2y_1,\ x_0y_1 + x_1y_0 + x_2y_2,\ x_0y_2 + x_1y_1 + x_2y_0).$$

Observe that

$$\begin{pmatrix} x_0 & x_2 & x_1\\ x_1 & x_0 & x_2\\ x_2 & x_1 & x_0 \end{pmatrix}\begin{pmatrix} y_0\\ y_1\\ y_2 \end{pmatrix} = \begin{pmatrix} x_0y_0 + x_1y_2 + x_2y_1\\ x_0y_1 + x_1y_0 + x_2y_2\\ x_0y_2 + x_1y_1 + x_2y_0 \end{pmatrix}.$$

The matrix

$$\begin{pmatrix} x_0 & x_2 & x_1\\ x_1 & x_0 & x_2\\ x_2 & x_1 & x_0 \end{pmatrix}$$

is called a *circulant matrix*. It is obtained from a given column vector by repeatedly making cyclic permutations on the coordinates. The formula giving the convolution of two vectors in terms of a circulant matrix holds for any n.

Since the vector space $\mathrm{L}^2(G)$ has dimension $|G|$ and since by Proposition 10.12 the dual group $\widehat{G}$ is isomorphic to G, by Proposition 10.13, the set of $\widehat{G}$ of characters of G is an orthonormal basis of $\mathrm{L}^2(G)$.

Another remarkable property of the dual group is that there is a natural homomorphism from G to its double dual $\widehat{\widehat{G}}$. In fact, a fundamental (and famous) theorem due to Pontrjagin asserts that this map is an isomorphism (and a homeomorphism), which will allow us to define an inverse of the Fourier transform (considering functions in $\mathrm{L}^2(G)$).

Proposition 10.15. *Let G be a locally compact abelian group. Given any $a \in G$, define the map $\eta_a \colon \widehat{G} \to \mathbb{C}$ by*

$$\eta_a(\chi) = \chi(a), \qquad \text{evaluation at } a.$$

The map $\eta \colon G \to \widehat{\widehat{G}}$ given by $\eta(a) = \eta_a$ is a continuous homomorphism from G to its double dual $\widehat{\widehat{G}}$.

Proof. First let us check that η_a is a character of the group $\widehat{G}$. For any $a \in G$, for any two characters $\chi_1, \chi_2 \in \widehat{G}$, we have

$$\eta_a(\chi_1\chi_2) = \chi_1(a)\chi_2(a) = \eta_a(\chi_1)\eta_a(\chi_2),$$

so η_a is a homomorphism. Since the characters have range $\mathbf{U}(1)$ and $\eta_a(\chi) = \chi(a)$, we see that $\eta_a \colon \widehat{G} \to \mathbf{U}(1)$. Since G is locally compact, the map $(a, \chi) \mapsto \chi(a)$ from $G \times \widehat{G}$ to $\mathbb{C}$ is continuous, and this implies that η_a is continuous; for details, see Bourbaki [9] (Chapter 2, Section 1, No. 1).

Let us now check that η is a homomorphism. For all $a, b \in G$ and all $\chi \in \widehat{G}$, we have

$$\eta_{ab}(\chi) = \chi(ab) = \chi(a)\chi(b) = \eta_a(\chi)\eta_b(\chi),$$

showing that η is a homomorphism. The fact that η is continuous is a consequence of the continuity of the map $(a, \chi) \mapsto \chi(a)$; for details, see Bourbaki [9] (Chapter 2, Section 1, No. 1). $\square$

Neither the injectivity nor the surjectivity of the map η is easy to prove.

10.3 The Fourier Transform and the Fourier Cotransform

Given a locally compact abelian group G equipped with a Haar measure λ, the Fourier transform $\mathcal{F}(f)$ of a function $f \colon G \to \mathbb{C}$ is a complex-valued function $\mathcal{F}(f) \colon \widehat{G} \to \mathbb{C}$ defined not on G, but on its dual group $\widehat{G}$ of characters, by the formula

$$\mathcal{F}(f)(\chi) = \int \overline{\chi(a)} f(a)\, d\lambda(a), \qquad \chi \in \widehat{G}. \tag{$*$}$$

The first issue it to determine when this integral converges. Since $|\chi|$ is a bounded continuous function $\chi\colon G \to \mathbb{T}$, by Proposition 5.36, the integral in $(*)$ is well-defined if $f \in \mathrm{L}^1(G)$. If G is not compact, in general, $\mathrm{L}^1(G)$ is not a subspace of $\mathrm{L}^2(G)$ and $\mathrm{L}^2(G)$ is not a subspace of $\mathrm{L}^1(G)$, so in general, the integral in $(*)$ does not converge if $f \in \mathrm{L}^2(G)$.

The second issue is to determine the class of functions on $\widehat{G}$ to which $\mathcal{F}(f)$ belongs.

We will see that if $f \in \mathrm{L}^1(G)$, then $\mathcal{F}(f) \in \mathcal{C}_0(\widehat{G};\mathbb{C})$, so in general $\mathcal{F}(f) \notin \mathrm{L}^1(\widehat{G})$. Remarkably, if $f \in \mathrm{L}^1(G) \cap \mathrm{L}^2(G)$, then $\mathcal{F}(f) \in \mathrm{L}^2(\widehat{G})$. In fact, because $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$ is dense in $\mathrm{L}^2(G)$, the Fourier transform $\mathcal{F}$ has a unique isometric extension from $\mathrm{L}^2(G)$ to $\mathrm{L}^2(\widehat{G})$. This is *Plancherel's theorem*.

If $\mathcal{F}(f) \in \mathrm{L}^1(\widehat{G})$ or if $\mathcal{F}(f) \in \mathrm{L}^2(\widehat{G})$, for some function $f\colon G \to \mathbb{C}$, then the question of finding an inverse of the Fourier transform arises. Is there a transform $\mathcal{G}$ defined on $\mathrm{L}^1(\widehat{G})$ or $\mathrm{L}^2(\widehat{G})$, or some subspace of them, such that

$$f = \mathcal{G}(\mathcal{F}(f))?$$

We call this equation a *Fourier inversion formula*. As stated, the problem does not make sense because $\mathcal{G}$, being defined on functions in $\mathrm{L}^1(\widehat{G})$ or $\mathrm{L}^2(\widehat{G})$, is a function defined on $\widehat{\widehat{G}}$, and not G. However, we showed in Proposition 10.15 that there is a homomorphism $\eta\colon G \to \widehat{\widehat{G}}$, and by Pontrjagin duality theorem, this map is an isomorphism, so the correct way to state Fourier inversion formula is so say that

$$f = (\mathcal{G} \circ \mathcal{F})(f) \circ \eta.$$

A bit less formally, since η is an isomorphism, by identifying G and $\widehat{\widehat{G}}$, we can drop η from the above formula. Then amazingly, the inverse $\mathcal{G}$ of $\mathcal{F}$ almost looks like $\mathcal{F}$, except that there is no conjugation on the character; that is, for every function $F \in \mathrm{L}^1(\widehat{G})$, for every character $\zeta \in \widehat{\widehat{G}}$, we have

$$\mathcal{G}(F)(\zeta) = \int \zeta(\chi) F(\chi)\, d\widehat{\lambda}(\chi), \qquad \zeta \in \widehat{\widehat{G}}. \tag{$**$}$$

In the above formula, $\widehat{\lambda}$ is a Haar measure on the dual group $\widehat{G}$ (suitably normalized), and χ is any element in $\widehat{G}$ (so, χ is a character of G).

Following Bourbaki, it seems fair to define simultaneously two notions of Fourier transforms.

Definition 10.3. Let G be a locally compact abelian group equipped with a Haar measure λ. For every function $f \in \mathrm{L}^1(G)$,

(1) The *Fourier transform* of f is the function $\mathcal{F}(f)\colon \widehat{G} \to \mathbb{C}$ given by

$$\mathcal{F}(f)(\chi) = \int \overline{\chi(a)} f(a)\, d\lambda(a), \qquad \chi \in \widehat{G}.$$

(2) The *Fourier cotransform* of f is the function $\overline{\mathcal{F}}(f)\colon \widehat{G} \to \mathbb{C}$ given by

$$\overline{\mathcal{F}}(f)(\chi) = \int \chi(a) f(a)\, d\lambda(a), \qquad \chi \in \widehat{G}.$$

Remark: We warn our readers that some authors define the Fourier transform as our notion of Fourier cotransform (and the notion of Fourier cotransform as our notion of Fourier transform). This is the convention adopted in Malliavin [68].

These transforms are not independent. In fact, each one can be obtained from the other. Recall that for any function $f\colon G \to \mathbb{C}$, the function $\check{f}$ is given by $\check{f}(s) = f(s^{-1})$ for all $s \in G$; see Definition 8.11.

Proposition 10.16. *The Fourier transform $\mathcal{F}$ and the Fourier cotransform $\overline{\mathcal{F}}$ are related as follows: for all $f \in \mathrm{L}^1(G)$ and all $\chi \in \widehat{G}$,*

$$\overline{\mathcal{F}}(f)(\chi) = \mathcal{F}(f)(\chi^{-1}) = \mathcal{F}(\check{f})(\chi) = \overline{\mathcal{F}(\overline{f})(\chi)}.$$

Proof. We have

$$\begin{aligned}\overline{\mathcal{F}}(f)(\chi) &= \int \chi(a) f(a)\, d\lambda(a)\\ &= \int \overline{\overline{\chi(a)}} f(a)\, d\lambda(a)\\ &= \int \overline{\chi^{-1}(a)} f(a)\, d\lambda(a) = \mathcal{F}(f)(\chi^{-1}),\end{aligned}$$

$$\int \overline{\overline{\chi(a) f(a)}}\, d\lambda(a) = \overline{\int \overline{\chi(a) f(a)}\, d\lambda(a)} = \overline{\mathcal{F}(\overline{f})(\chi)},$$

and

$$\begin{aligned}\int \overline{\chi^{-1}(a)} f(a)\, d\lambda(a) &= \int \overline{\chi(a^{-1})} f(a)\, d\lambda(a)\\ &= \int \overline{\chi(a)} f(a^{-1})\, d\lambda(a) = \mathcal{F}(\check{f})(\chi),\end{aligned}$$

where we used the fact that a commutative locally compact group is unimodular (Proposition 8.25) and Proposition 8.27 to change a to a^{-1} in the second equation above, so $\overline{\mathcal{F}}(f)(\chi) = \mathcal{F}(f)(\chi^{-1}) = \mathcal{F}(\check{f})(\chi) = \overline{\mathcal{F}(\overline{f})(\chi)}$. ∎

With these definitions, the Fourier inversion formula is

$$f = (\overline{\mathcal{F}} \circ \mathcal{F})(f) \circ \eta. \tag{finv}$$

In the above formula, $\overline{\mathcal{F}}$ is the Fourier cotransform on functions defined on the dual group $\widehat{G}$. If we replace $\overline{\mathcal{F}}$ by $\mathcal{F}$, then we don't quite the right formula. By Proposition 10.16, we have

$$(\mathcal{F}(\mathcal{F}(f))(\zeta) = (\overline{\mathcal{F}}(\mathcal{F}))(\zeta^{-1}), \quad \zeta \in \widehat{\widehat{G}},$$

so with $\zeta = \eta(g)$ (with $g \in G$), since η is a homomorphism, we obtain

$$(\mathcal{F}(\mathcal{F}(f))(\eta(g)) = (\overline{\mathcal{F}}(\mathcal{F}))(\eta(g^{-1})),$$

which by (finv) yields

$$(\mathcal{F}(\mathcal{F}(f))(\eta(g)) = (\overline{\mathcal{F}}(\mathcal{F}))(\eta(g^{-1})) = f(g^{-1}).$$

Therefore, instead of (finv), we have

$$\check{f} = (\mathcal{F} \circ \mathcal{F})(f) \circ \eta.$$

Of course, in the above formula the leftmost occurrence of $\mathcal{F}$ is the Fourier transform on functions defined on the dual group $\widehat{G}$. But if we "apply $\mathcal{F}$ four times," we get

$$f = (\mathcal{F} \circ \mathcal{F} \circ \mathcal{F} \circ \mathcal{F})(f) \circ \eta.$$

This looks a little silly to us and seems another justification for considering $\overline{\mathcal{F}}$ on an equal footing with $\mathcal{F}$.

Actually, in view of the isomorphism $j\colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$ given by Theorem 10.6, the Fourier cotransform $\overline{\mathcal{F}}$ can be viewed as the Gelfand transform from $\mathrm{L}^1(G)$ to $\mathsf{X}(\mathrm{L}^1(G))$. For any $f \in \mathrm{L}^1(G)$, the Gelfand transform $\mathcal{G}_f$ of f is given by

$$\mathcal{G}_f(\zeta) = \zeta(f), \qquad \zeta \in \mathsf{X}(\mathrm{L}^1(G)).$$

By Theorem 10.6, the map $j\colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$ given by

$$j(\chi)(f) = \zeta_\chi(f) = \int \chi(a) f(a)\, d\lambda(a), \qquad \chi \in \widehat{G},\ f \in \mathrm{L}^1(G),$$

is a homeomorphism of $\widehat{G}$ onto $\mathsf{X}(\mathrm{L}^1(G))$. But

$$\overline{\mathcal{F}}(f)(\chi) = \int \chi(a) f(a)\, d\lambda(a),$$

so

$$j(\chi)(f) = \zeta_\chi(f) = \overline{\mathcal{F}}(f)(\chi),$$

and since $\mathcal{G}_f(\zeta_\chi) = \zeta_\chi(f)$ and $\zeta_\chi = j(\chi)$, we see that

$$\mathcal{G}_f(j(\chi)) = \mathcal{G}_f(\zeta_\chi) = \overline{\mathcal{F}}(f)(\chi). \tag{\dagger}$$

In summary, we proved the following result.

Proposition 10.17. *Modulo the isomorphism* $j \colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$, *the Fourier cotransform* $\overline{\mathcal{F}}(f)$ *is the Gelfand transform* $\mathcal{G}_f$ *from* $\mathrm{L}^1(G)$ *to* $\mathsf{X}(\mathrm{L}^1(G))$.

This is good news because this shows that we can apply results known for the Gelfand transform to the Fourier cotransform, and thus to the Fourier transform.

The Fourier inversion formula turns out to hold in the following cases:

(1) If we identify G and $\widehat{\widehat{G}}$ using the Pontrjagin isomorphism theorem (and give suitably normalized Haar measures to G and $\widehat{G}$; see Section 10.9), then there is a unique extension of the Fourier tranform $\mathcal{F}$ to $\mathrm{L}^2(G)$ and a unique extension of the Fourier cotransform $\overline{\mathcal{F}}$ to $\mathrm{L}^2(\widehat{G})$, so that they are mutual inverses.

(2) If $A(G)$ is the subspace of $\mathrm{L}^1(G)$ spanned by all functions of the form $f * g$ with $f, g \in \mathrm{L}^1(G) \cap \mathrm{L}^2(G)$, then $A(G)$ is an ideal of $\mathrm{L}^1(G)$ contained in $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$; if $f \in A(G)$, then $\mathcal{F}(f) \in \mathrm{L}^1(\widehat{G})$ and the Fourier inversion formula holds.

(3) If $B(G) = \{f \in \mathrm{L}^1(G) \mid \mathcal{F}(f) \in \mathrm{L}^1(\widehat{G})\}$, then the restriction of $\mathcal{F}$ to $B(G)$ is a bijection onto $B(\widehat{G})$, and its inverse is the restriction of $\overline{\mathcal{F}}$ to $B(\widehat{G})$.

Thus it appears that the L^2 theory has the best behavior with respect to the Fourier transform. This had been observed for $G = \mathbb{R}$ long ago.

We now redefine the Fourier transform and the Fourier cotransform so that they apply to complex measures $\mu \in \mathcal{M}^1(G)$.

Definition 10.4. Let G be a locally compact abelian group equipped with a Haar measure λ. For every complex measure $\mu \in \mathcal{M}^1(G)$, the *Fourier transform* $\mathcal{F}(\mu)$ and the *Fourier cotransform* $\overline{\mathcal{F}}(\mu)$ of μ are the functions $\mathcal{F}(\mu) \colon \widehat{G} \to \mathbb{C}$ and $\overline{\mathcal{F}}(\mu) \colon \widehat{G} \to \mathbb{C}$ defined on the group $\widehat{G}$ by

$$\begin{aligned}\mathcal{F}(\mu)(\chi) &= \int \overline{\chi(a)}\, d\mu(a)\\ \overline{\mathcal{F}}(\mu)(\chi) &= \int \chi(a)\, d\mu(a),\end{aligned}$$

for all $\chi \in \widehat{G}$.

For every function $f \in \mathrm{L}^1(G)$, the *Fourier transform* $\mathcal{F}(f)$ and the *Fourier cotransform* $\overline{\mathcal{F}}(f)$ of f are the functions $\mathcal{F}(f) \colon \widehat{G} \to \mathbb{C}$ and $\overline{\mathcal{F}}(f) \colon \widehat{G} \to \mathbb{C}$ defined on the group $\widehat{G}$ by

$$\begin{aligned}\mathcal{F}(f)(\chi) &= \int \overline{\chi(a)} f(a)\, d\lambda(a)\\ \overline{\mathcal{F}}(f)(\chi) &= \int \chi(a) f(a)\, d\lambda(a),\end{aligned}$$

for all $\chi \in \widehat{G}$.

Remark: The Fourier cotransform is also called the *inverse Fourier transform* by some authors, including Hewitt and Ross.

As in the case of functions,

$$\overline{\mathcal{F}}(\mu)(\chi) = \mathcal{F}(\mu)(\chi^{-1}) = \mathcal{F}\big(\check{\mu}\big)(\chi) = \overline{\mathcal{F}(\overline{\mu})(\chi)}.$$

Here is our first result. In particular, it gives the fundamental property of the Fourier transform (and cotransform), which is to convert a convolution into a (pointwise) product of functions. Recall that for a function $f\colon G \to \mathbb{C}$, we have $f^*(a) = \overline{f(a^{-1})}$.

Proposition 10.18.

(1) The Fourier transform $\mathcal{F}$ and the Fourier cotransform $\overline{\mathcal{F}}$ are involutive homomorphisms from the unital involutive Banach algebra $\mathcal{M}^1(G)$ to the unital involutive Banach algebra $\mathcal{C}_b(\widehat{G};\mathbb{C})$ of continuous bounded functions on $\widehat{G}$. In particular for any two complex measures $\mu, \nu \in \mathcal{M}^1(G)$, we have

$$\mathcal{F}(\mu * \nu) = \mathcal{F}(\mu)\mathcal{F}(\nu), \qquad \overline{\mathcal{F}}(\mu * \nu) = \overline{\mathcal{F}}(\mu)\overline{\mathcal{F}}(\nu),$$

and

$$\mathcal{F}(\overline{\check{\mu}}) = (\mathcal{F}(\mu))^*, \qquad \overline{\mathcal{F}}(\overline{\check{\mu}}) = (\overline{\mathcal{F}}(\mu))^*.$$

(2) The Fourier transform $\mathcal{F}$ and the Fourier cotransform $\overline{\mathcal{F}}$ are injective involutive homomorphisms from the involutive Banach algebra $\mathrm{L}^1(G)$ to the involutive Banach algebra $\mathcal{C}_0(\widehat{G};\mathbb{C})$ of continuous functions on $\widehat{G}$ that tend to zero at infinity. In particular for any two functions $f, g \in \mathrm{L}^1(G)$, we have

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g), \qquad \overline{\mathcal{F}}(f * g) = \overline{\mathcal{F}}(f)\overline{\mathcal{F}}(g),$$

and

$$\mathcal{F}(f^*) = (\mathcal{F}(f))^*, \qquad \overline{\mathcal{F}}(f^*) = (\overline{\mathcal{F}}(f))^*.$$

Proof. (1) Observe that for any character $\chi \in \widehat{G}$, we have $\overline{\mathcal{F}}(\mu)(\chi) = \zeta_\chi(\mu)$, as defined in Proposition 10.5, and by this proposition, ζ_χ is a character of the algebra $\mathcal{M}^1(G)$, so $\overline{\mathcal{F}}$ is a homomorphism. Since $\mathcal{F}(\mu)(\chi) = \overline{\mathcal{F}}(\mu)(\chi^{-1})$, the Fourier transform $\mathcal{F}$ is also a homomorphism. Since $|\chi(a)| = 1$ for all $a \in G$, we have

$$|\mathcal{F}(\mu)(\chi)| = \left| \int \overline{\chi(a)}\, d\mu(a) \right| \leq \|\mu\|,$$

and the same argument applies to $\overline{\mathcal{F}}(\mu)$. Therefore $\mathcal{F}(\mu)$ and $\overline{\mathcal{F}}(\mu)$ are bounded functions on $\widehat{G}$.

Finally, by the definition of the compact-open topology on $\widehat{G}$, if χ_1 tends to χ_2 in $\widehat{G}$, then the function χ_1 defined on G tends to χ_2 uniformly on every compact set, while being

bounded by 1. It follows that $\mathcal{F}(\mu)(\chi_1)$ tends to $\mathcal{F}(\mu)(\chi_2)$. Thus $\mathcal{F}(\mu)$ is continuous. A similar reasoning shows that $\overline{\mathcal{F}}(\mu)$ is continuous.

(2) Since $\mathrm{L}^1(G)$ is a subalgebra of $\mathcal{M}^1(G)$, we see immediately that $\mathcal{F}$ and $\overline{\mathcal{F}}$ are homomorphisms. Since $\widehat{G}$ and $\mathsf{X}(\mathrm{L}^1(G))$ are homeomorphic, $\overline{\mathcal{F}}(f)$ can be identified with the Gelfand transform $\mathcal{G}_f$, by the remark following Theorem 9.23, the Gelfand transform $\mathcal{G}$ maps $\mathrm{L}^1(G)$ into $\mathcal{C}_0(\mathsf{X}(\mathrm{L}^1(G)) \cong \mathcal{C}_0(\widehat{G})$, so $\overline{\mathcal{F}}$ maps $\mathrm{L}^1(G)$ to $\mathcal{C}_0(\widehat{G})$. Since $\mathcal{F}(\mu)(\chi) = \overline{\mathcal{F}}(\mu)(\chi^{-1})$, the Fourier transform $\mathcal{F}$ also maps $\mathrm{L}^1(G)$ to $\mathcal{C}_0(\widehat{G})$.

By Proposition 9.27, the Gelfand transform $\mathcal{G}$ is injective because it can be shown that $\mathrm{L}^1(G)$ has radical (0); see Proposition 9.43. As a consequence $\overline{\mathcal{F}}$ is injective, and since $\mathcal{F}(\mu)(\chi) = \overline{\mathcal{F}}(\mu)(\chi^{-1})$, the Fourier transform is also injective. □

Remarks:

(1) If $G = \mathbb{R}^n$, the fact that the Fourier transform $\mathcal{F}(f)$ is a continuous function that tends to zero at infinity is known as the *Riemann–Lebesgue lemma*; see Folland [34] (Chapter 8, Theorem 8.22).

(2) Since $\mathrm{L}^1(G)$ is a commutative Banach algebra, one may wonder whether a quicker proof of the injectivity of $\mathcal{G}$ could be obtained using the Gelfand-Naimark theorem (Theorem 9.37). Unfortunately, $\mathrm{L}^1(G)$ is *not* a unital C^*-algebra so this theorem does not apply. However, $\mathrm{L}^1(G)$ is dense in its enveloping C^*-algebra $\mathrm{St}(G)$ (see Definition 9.23), so $\mathcal{G}$ extends to an isomorphism between $\mathrm{St}(G)$ and $\mathcal{C}_0(\widehat{G})$.

Example 10.2.

(1) Let $G = \mathbb{R}$ and equip G with the Lebesgue measure dx. Recall from Proposition 10.9 that $\widehat{\mathbb{R}}$ is isomorphic to $\mathbb{R}$. If we choose the characters to be the maps $x \mapsto e^{iyx}$ for some $y \in \mathbb{R}$, then it turns out that for the Fourier inversion formula to come out right, $\widehat{\mathbb{R}}$ needs to be equipped with $dx/2\pi$. The Fourier transform on $\mathbb{R}$ is

$$\widehat{f}(x) = \mathcal{F}(f)(x) = \int f(y) e^{-ixy}\, dy,$$

and on $\widehat{\mathbb{R}}$ it is

$$\mathcal{F}(f)(x) = \int f(y) e^{-ixy}\, \frac{dy}{2\pi}.$$

The Fourier cotransform is obtained by changing the sign $-$ in the exponent to a $+$ sign. Then the inversion formula is

$$f(x) = \int \widehat{f}(y) e^{ixy}\, \frac{dy}{2\pi}.$$

This is the choice made in Malliavin [68].

Another choice that works is to equip both $\mathbb{R}$ and $\widehat{\mathbb{R}}$ with the normalized Lebesgue measure $dx/\sqrt{2\pi}$, as in Rudin [79, 80], and in Chapter 6.

With the choice of characters $x \mapsto e^{2\pi i y x}$, the Lebesgue measure is self-dual; see Folland [34]. The Fourier transform on $\mathbb{R}$ and $\widehat{\mathbb{R}}$ is

$$\widehat{f}(x) = \mathcal{F}(f)(x) = \int f(y) e^{-2\pi i x y}\, dy.$$

The Fourier cotransform is obtained by changing the sign $-$ in the exponent to a $+$ sign, and the inversion formula is

$$f(x) = \int \widehat{f}(y) e^{2\pi i x y}\, dy.$$

The trick to pick the right normalization factor is that the Fourier transform $\widehat{g}$ of the function $g(x) = e^{-\pi x^2}$ should be g itself; $\widehat{g} = g$; see Folland [34] (Chapter 8, Proposition 8.24).

(2) Let $G = \mathbb{T}$, equipped with the Haar measure $d\nu_1/2\pi$ inherited from $\mathbb{R}$, by viewing $\mathbb{T}$ as $\mathbb{R}/2\pi\mathbb{Z}$, so that $\mathbb{T}$ has measure 1; see Example 8.4. The characters of $\mathbb{T}$ are the maps $e^{i\theta} \mapsto e^{im\theta}$, with $m \in \mathbb{Z}$. We equip $\mathbb{Z}$ with the counting measure, and the characters are the maps $m \mapsto e^{im\theta}$, with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. The Fourier transform $\widehat{f} = \mathcal{F}(f)$ of a function $f \in \mathrm{L}^1(\mathbb{T})$ is the $\mathbb{Z}$-indexed sequence whose mth element $\widehat{f}_m$, the *mth Fourier coefficient* of f, is given by

$$\widehat{f}_m = \mathcal{F}(f)(m) = \int_{-\pi}^{\pi} f(\theta) e^{-im\theta} \frac{d\theta}{2\pi}.$$

The Fourier transform $\mathcal{F}(c)$ of a sequence $c = (c_m)_{m\in\mathbb{Z}} \in \mathrm{L}^1(\mathbb{Z}) = l^1(\mathbb{Z})$ is the function

$$f(e^{i\theta}) = \mathcal{F}(c)(e^{i\theta}) = \sum_{m\in\mathbb{Z}} c_m e^{-im\theta}.$$

The Fourier cotransform is obtained by changing the sign $-$ in the exponent to a $+$ sign. This is traditionally called the *Fourier series* associated with $c = (c_m)_{m\in\mathbb{Z}}$. The inversion formula is

$$f(e^{i\theta}) = \sum_{m\in\mathbb{Z}} \widehat{f}_m e^{im\theta}.$$

(3) Consider the group $G = (\mathbb{R}_+^*, *)$, the group of positive reals under multiplication. This group comes up in the *Mellin transform*. Since we have the isomorphism $\varphi\colon \mathbb{R} \to \mathbb{R}_+^*$ given by $\varphi(x) = e^x$, where the group operation on $\mathbb{R}$ is addition, it is easy to figure out the characters of $\mathbb{R}_+^*$, which are homomorphisms $\chi\colon \mathbb{R}_+^* \to \mathbf{U}(1)$. Indeed, there is a bijection between the characters $\zeta \in \widehat{\mathbb{R}}$, which are homomorphisms $\zeta\colon \mathbb{R} \to \mathbf{U}(1)$, and the characters $\chi \in \widehat{\mathbb{R}_+^*}$, with $\chi\colon \mathbb{R}_+^* \to \mathbf{U}(1)$, given by

$$\zeta = \chi \circ \varphi,$$

as illustrated by the following diagram.

$$\begin{tikzcd} \mathbb{R} \arrow[rr, "\varphi"] \arrow[dr, "\zeta"'] & & \mathbb{R}_+^* \arrow[dl, "\chi"] \\ & \mathbf{U}(1) & \end{tikzcd}$$

By Proposition 10.9, every character $\zeta \in \widehat{\mathbb{R}}$ is of the form $\zeta(z) = e^{ixz}$ for some $x \in \mathbb{R}$, and since $\chi(y) = \zeta(\varphi^{-1}(y)) = \zeta(\log y)$ for any $y > 0$, we get

$$\chi(y) = \zeta(\log y) = e^{ix \log y} = (e^{\log y})^{ix} = y^{ix}.$$

Therefore, the characters of $\mathbb{R}_+^*$ are of the form

$$y \mapsto y^{ix}, \quad y \in \mathbb{R}_+^*,$$

for some fixed $x \in \mathbb{R}$, which shows that

$$\widehat{\mathbb{R}_+^*} \cong \mathbb{R}.$$

We can also easily determine the characters of $\widehat{\mathbb{R}_+^*}$.

We know from Proposition 10.9 that every $\zeta \in \widehat{\mathbb{R}}$ is of the form $\zeta(z) = e^{ixz}$ for some $x \in \mathbb{R}$. If we let $y = e^x$, then $y > 0$, and

$$\zeta(z) = e^{ixz} = y^{iz}.$$

Therefore, every character of $\widehat{\mathbb{R}_+^*}$ is of the form

$$z \mapsto y^{iz}, \quad z \in \mathbb{R},$$

for some fixed $y \in \mathbb{R}_+^*$. Thus

$$\widehat{\widehat{\mathbb{R}_+^*}} \cong \mathbb{R}_+^*,$$

which confirms the general theory. If we pick the left-invariant Haar measure dy/y on $\mathbb{R}_+^*$ (see Example 8.3), then the Fourier transform on $\mathrm{L}^1(\mathbb{R}_+^*)$ is given by

$$\mathcal{F}(f)(x) = \int_0^\infty y^{-ix} f(y)\, \frac{dy}{y},$$

for any $f \in \mathrm{L}^1(\mathbb{R}_+^*)$ and all $x \in \mathbb{R}$. This is one of the formulations of the *Mellin transform*, often denoted $\mathcal{M}(f)$.

If we give $\widehat{\mathbb{R}_+^*}$ the Haar measure $dx/2\pi$, then the Fourier cotransform on $\mathrm{L}^1(\widehat{\mathbb{R}_+^*}) = \mathrm{L}^1(\mathbb{R})$ is given by

$$\overline{\mathcal{F}}(g)(y) = \frac{1}{2\pi} \int_{\mathbb{R}} y^{ix} g(x)\, dx,$$

for any $g \in \mathrm{L}^1(\mathbb{R})$ and all $y \in \mathbb{R}_+^*$. This is one of the formulations of the *inverse Mellin transform*, often denoted $\mathcal{M}^{-1}(g)$. The inversion formula is

$$f(y) = \frac{1}{2\pi} \int_{\mathbb{R}} y^{ix} \mathcal{M}(f)(x)\, dx, \quad y \in \mathbb{R}_+^*.$$

Observe that on $\mathrm{L}^1(\mathbb{R}_+^*)$, convolution is given by

$$(f * g)(y) = \int_0^\infty f(z) g\left(\frac{y}{z}\right) \frac{dz}{z},$$

and by the general theory,

$$\mathcal{M}(f * g) = \mathcal{M}(f)\mathcal{M}(g).$$

Some useful properties of the Fourier transform are listed below.

Proposition 10.19. *Let G be locally compact abelian group. The Fourier transforms and the Fourier cotransforms satisfy the following equations: For all $f \in \mathrm{L}^1(G)$, and all $\chi \in \widehat{G}$,*

(1)

$$\overline{\mathcal{F}}(f)(\chi) = \mathcal{F}(f)(\chi^{-1}) = \mathcal{F}(\check{f})(\chi) = \overline{\mathcal{F}(\overline{f})(\chi)}.$$

These equations also hold with f replaced by a complex measure $\mu \in \mathcal{M}^1(G)$. We also have

$$\mathcal{F}(\delta_a)(\chi) = \overline{\chi(a)}, \qquad \overline{\mathcal{F}}(\delta_a)(\chi) = \chi(a).$$

(2)

$$\|\mathcal{F}(f)\|_\infty = \left\|\overline{\mathcal{F}}(f)\right\|_\infty \leq \|f\|_1 .$$

These equations also hold with f replaced by a complex measure $\mu \in \mathcal{M}^1(G)$.

(3)

$$\begin{aligned} \mathcal{F}(\lambda_a(f))(\chi) &= \overline{\chi(a)}\mathcal{F}(f)(\chi) \\ \overline{\mathcal{F}}(\lambda_a(f))(\chi) &= \chi(a)\overline{\mathcal{F}}(f)(\chi). \end{aligned}$$

These equations also hold with f replaced by a complex measure $\mu \in \mathcal{M}^1(G)$.

(4) For all $f \in \mathrm{L}^1(G)$, and all $\chi, \xi \in \widehat{G}$,

$$\begin{aligned} \mathcal{F}(\xi f)(\chi) &= \mathcal{F}(f)(\xi^{-1}\chi) = \lambda_\xi(\mathcal{F}(f))(\chi) \\ \overline{\mathcal{F}}(\xi f)(\chi) &= \overline{\mathcal{F}}(f)(\xi^{-1}\chi) = \lambda_\xi(\overline{\mathcal{F}}(f))(\chi). \end{aligned}$$

These equations also hold with f replaced by a complex measure $\mu \in \mathcal{M}^1(G)$.

Proof. (1) We have already proven Equations (1) in Proposition 10.16. Part (2) is proven in the proof of Part (1) of Proposition 10.18.

(3) We have

$$\begin{aligned}
\mathcal{F}(\lambda_a(f))(\chi) &= \int \overline{\chi(b)} f(a^{-1}b)\, d\lambda(b) \\
&= \int \overline{\chi(ab)} f(b)\, d\lambda(b) \\
&= \int \overline{\chi(a)\chi(b)} f(b)\, d\lambda(b) \\
&= \overline{\chi(a)} \int \overline{\chi(b)} f(b)\, d\lambda(b) = \overline{\chi(a)}\mathcal{F}(f)(\chi).
\end{aligned}$$

The second equation is proven in a similar way.

(4) We have

$$\begin{aligned}
\mathcal{F}(\xi f)(\chi) &= \int \overline{\chi(a)} \xi(a) f(a)\, d\lambda(a) \\
&= \int \overline{(\xi^{-1}\chi)(a)} f(a)\, d\lambda(a) \\
&= \mathcal{F}(f)(\xi^{-1}\chi) = \lambda_\xi(\mathcal{F}(f))(\chi).
\end{aligned}$$

We leave the proof of the other equations as exercises. ☐

Example 10.3. If $G = \mathbb{R}$, since the characters are of the form $x \mapsto e^{iyx}$ (with $y \in \mathbb{R}$), then, with a slight abuse of notation, Equations (3) yields the well-known formula

$$\mathcal{F}(\lambda_a(f))(x) = \mathcal{F}(f(x-a)) = e^{-iax}\mathcal{F}(f)(x),$$

and Equations (4) yields

$$\mathcal{F}(e^{iax} f(x)) = \mathcal{F}(f)(x-a);$$

see Rudin [79] (Chapter 9) or Folland [34] (Chapter 8); recall that $\lambda_a(f)(x) = f(x-a)$.

If $G = \mathbb{T}$, since the characters are of the form $e^{i\theta} \mapsto e^{im\theta}$ (with $m \in \mathbb{Z}$), then, with a slight abuse of notation, Equations (3) yields the formula

$$\mathcal{F}(\lambda_{e^{i\varphi}}(f))(m) = e^{-im\varphi}\mathcal{F}(f)(m),$$

and Equations (4) yields

$$\mathcal{F}(e^{in\theta} f(e^{i\theta}))(m) = \mathcal{F}(f)(m-n).$$

Recall that $f\colon \mathbb{T} \to \mathbb{C}$, and that $\mathcal{F}(f)$ is a $\mathbb{Z}$-indexed sequence of complex numbers, namely the Fourier coefficients of f.

If $G = \mathbb{Z}$, since the characters are of the form $m \mapsto e^{im\theta}$ with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, Equations (3) yields the formula

$$\mathcal{F}(\lambda_n(f))(\theta) = e^{-in\theta}\mathcal{F}(f)(\theta),$$

and Equations (4) yields

$$\mathcal{F}(e^{im\varphi}f(m))(\theta) = \mathcal{F}(f)(\theta - \varphi).$$

Recall that $f \colon \mathbb{Z} \to \mathbb{C}$ is a $\mathbb{Z}$-indexed sequence, and that $\mathcal{F}(f)(\theta) = \sum_{m\in\mathbb{Z}} f(m)e^{-im\theta}$ is a Fourier series.

10.4 The Fourier Transform on a Finite Abelian Group

Let G be a finite locally compact abelian group, and as in Example 10.1, assume that the Haar measure is normalized so that $\lambda(G) = 1$. It is possible and very instructive to work out explicitly the Fourier transform on $\mathrm{L}^2(G)$ and the Fourier cotransform on $\mathrm{L}^2(\widehat{G})$. We will also prove directly that Fourier inversion holds.

Recall that the inner product $\langle x, y\rangle$ of $x, y \in \mathrm{L}^2(G)$ is given by

$$\langle x, y\rangle = \frac{1}{|G|}\sum_{a\in G} x_a\overline{y_a}.$$

Then the Fourier transform and the Fourier cotransform of $x = (x_a)_{a\in G} \in \mathrm{L}^2(G)$ are given by

$$\mathcal{F}(x)(\chi) = \frac{1}{|G|}\sum_{a\in G} x_a\overline{\chi(a)} \qquad (\dagger_1)$$

$$\overline{\mathcal{F}}(x)(\chi) = \frac{1}{|G|}\sum_{a\in G} x_a\chi(a), \qquad (\dagger_2)$$

where $\chi \colon G \to \mathbb{T}$ is a character of G.

Observe that

$$\mathcal{F}(x)(\chi) = \frac{1}{|G|}\sum_{a\in G} x_a\overline{\chi(a)} = \langle x, \chi\rangle.$$

Recall that $\widehat{G} \cong G$ by Proposition 10.12, so we can write $\widehat{G} = \{\chi_1, \ldots, \chi_n\}$, with $n = |G|$. We may call $\widehat{x}(\chi_i) = \mathcal{F}(x)(\chi_i) = \langle x, \chi_i\rangle$ the ith *Fourier coefficient* of x.

Proposition 10.20. *(Fourier inversion formula) Let G be a finite abelian group of order n, and let $\widehat{G} = \{\chi_1, \ldots, \chi_n\}$. For every $x \in \mathrm{L}^2(G)$, we have*

$$x = \sum_{j=1}^{n} \mathcal{F}(x)(\chi_j)\chi_j = \sum_{j=1}^{n} \widehat{x}(\chi_j)\chi_j,$$

with

$$\mathcal{F}(x)(\chi_j) = \langle x, \chi_j\rangle = \frac{1}{|G|}\sum_{a\in G} x_a\overline{\chi_j(a)}.$$

Proof. By Proposition 10.13 the characters $\{\chi_1, \ldots, \chi_n\}$ are orthonormal, and since $\mathrm{L}^2(G)$ is a vector space of dimension n, they form a basis. Consequently we can write

$$x = \sum_{j=1}^{n} c_j \chi_j$$

for some $c_j \in \mathbb{C}$. Taking the inner product with χ_j, we obtain

$$\langle x, \chi_j \rangle = c_j \langle \chi_j, \chi_j \rangle = c_j,$$

which concludes the proof. $\square$

Let us equip $\widehat{G}$ with the counting measure $\widehat{\lambda}$ normalized so that $\widehat{\lambda}(\widehat{G}) = |\widehat{G}| = |G|$. Then the integral of a function $H \in \mathrm{L}^2(\widehat{G})$ is given by

$$\int H(\chi)\, d\widehat{\lambda}(\chi) = \sum_{\chi \in \widehat{G}} H_\chi,$$

the inner product of $F, H \in \mathrm{L}^2(\widehat{G})$ is given by

$$\langle F, H \rangle = \sum_{\chi \in \widehat{G}} F_\chi \overline{H_\chi},$$

and the Fourier transform and the Fourier cotransform of $H = (H_\chi)_{\chi \in \widehat{G}} \in \mathrm{L}^2(\widehat{G})$ are given by

$$\mathcal{F}(H)(\zeta) = \sum_{\chi \in \widehat{G}} H_\chi \overline{\zeta(\chi)} \tag{$\dagger_3$}$$

$$\overline{\mathcal{F}}(H)(\zeta) = \sum_{\chi \in \widehat{G}} H_\chi \zeta(\chi), \tag{$\dagger_4$}$$

where $\zeta \colon \widehat{G} \to \mathbb{T}$ is a character of $\widehat{G}$. Observe that the factor $1/|G|$ is missing.

Recall that $\eta_a \in \widehat{\widehat{G}} \subseteq \mathrm{L}^2(\widehat{G})$ is a character on $\widehat{G}$ such that $\eta_a(\chi) = \chi(a)$.

Proposition 10.21. *(Fourier inversion) For any $x \in \mathrm{L}^2(G)$ and any $a \in G$, we have*

$$(\overline{\mathcal{F}} \circ \mathcal{F})(x)(\eta_a) = x_a,$$

that is,

$$x = (\overline{\mathcal{F}} \circ \mathcal{F})(x) \circ \eta.$$

Proof. Let us compute $(\overline{\mathcal{F}} \circ \mathcal{F})(x)(\eta_a) = \overline{\mathcal{F}}(\mathcal{F}(x))(\eta_a)$, for any $x = (x_b)_{b\in G}$ and any $a \in G$. Since $\eta_a(\chi) = \chi(a)$, we have

$$\begin{aligned}
(\overline{\mathcal{F}} \circ \mathcal{F})(x)(\eta_a) &= \overline{\mathcal{F}}(\mathcal{F}(x))(\eta_a) \\
&= \sum_{\chi\in\widehat{G}} \mathcal{F}(x)_\chi \, \eta_a(\chi) \\
&= \frac{1}{|G|} \sum_{\chi\in\widehat{G}} \sum_{b\in G} x_b \overline{\chi(b)} \chi(a) \\
&= \frac{1}{|G|} \sum_{b\in G} x_b \sum_{\chi\in\widehat{G}} \chi(a)\overline{\chi(b)} \\
&= \frac{1}{|G|} \sum_{b\in G} x_b \sum_{\chi\in\widehat{G}} \eta_a(\chi)\overline{\eta_b}(\chi) \\
&= \frac{1}{|G|} \sum_{b\in G} x_b \langle \eta_a, \eta_b\rangle.
\end{aligned}$$

But $\eta_a, \eta_b \in \widehat{\widehat{G}} \subseteq \mathrm{L}^2(\widehat{G})$, so by Proposition 10.13 (applied to $\widehat{\widehat{G}}$), all terms $\langle \eta_a, \eta_b\rangle$ are zero if $a \neq b$, and $\langle \eta_a, \eta_a\rangle = |G|$, because the Haar measure on $\widehat{G}$ is normalized so that $\widehat{\lambda}(\widehat{G}) = |\widehat{G}| = |G|$, so the factor $1/|G|$ is missing. Therefore,

$$(\overline{\mathcal{F}} \circ \mathcal{F})(x)(\eta_a) = \frac{1}{|G|} \sum_{b\in G} x_b \langle \eta_a, \eta_b\rangle = \frac{1}{|G|} x_a \langle \eta_a, \eta_a\rangle = \frac{1}{|G|} x_a |G| = x_a,$$

which proves that

$$x = (\overline{\mathcal{F}} \circ \mathcal{F})(x) \circ \eta,$$

namely, the Fourier inversion formula. $\square$

Observe that in order for the inversion formula to be correct, the normalization factor of the Haar measure λ on G and the normalization factor of the Haar measure $\widehat{\lambda}$ on $\widehat{G}$ have to be chosen carefully. In our case, we chose $\lambda(G) = 1$ and $\widehat{\lambda}(\widehat{G}) = |G|$. Instead we could have chosen $\lambda(G) = |G|$ and $\widehat{\lambda}(\widehat{G}) = 1$. The reader should check that the self-dual choice $\lambda(G) = \widehat{\lambda}(\widehat{G}) = \sqrt{|G|}$ also works.

The reader should also check that if we use $\mathcal{F}$ instead of $\overline{\mathcal{F}}$ on $\mathrm{L}^2(\widehat{G})$, then we get

$$\check{x} = (\mathcal{F} \circ \mathcal{F})(x) \circ \eta,$$

where $\check{x}(a) = x_{a^{-1}}$.

Proposition 10.22. *(Plancherel theorem) For all $x, y \in \mathrm{L}^2(G)$, we have*

$$\langle x, y\rangle = \langle \mathcal{F}(x), \mathcal{F}(y)\rangle.$$

As a consequence, $\mathrm{L}^2(G)$ and $\mathrm{L}^2(\widehat{G})$ are isometric.

Proof. We have

$$\begin{aligned}
\langle \mathcal{F}(x), \mathcal{F}(y)\rangle &= \sum_{\chi\in\widehat{G}} \mathcal{F}(x)_\chi \overline{\mathcal{F}(y)_\chi} \\
&= \frac{1}{|G|^2} \sum_{\chi\in\widehat{G}} \sum_{a\in G} \sum_{b\in G} x_a \overline{\chi(a)}\,\overline{y_b}\chi(b) \\
&= \frac{1}{|G|^2} \sum_{a\in G} \sum_{b\in G} x_a \overline{y_b} \sum_{\chi\in\widehat{G}} \chi(b)\overline{\chi(a)} \\
&= \frac{1}{|G|^2} \sum_{a\in G} \sum_{b\in G} x_a \overline{y_b} \sum_{\chi\in\widehat{G}} \eta_b(\chi)\overline{\eta_a(\chi)} \\
&= \frac{1}{|G|^2} \sum_{a\in G} \sum_{b\in G} x_a \overline{y_b} \langle \eta_b, \eta_a\rangle \\
&= \frac{1}{|G|^2} \sum_{a\in G} x_a \overline{y_a} |G| \\
&= \langle x, y\rangle.
\end{aligned}$$

Again, we used the fact that in $\mathrm{L}^2(\widehat{G})$, the measure $\widehat{\lambda}$ is normalized so that $\widehat{\lambda}(\widehat{G}) = |G|$, so $\langle \eta_a, \eta_a\rangle = |G|$ (and $\langle \eta_a, \eta_b\rangle = 0$ whenever $a \neq b$). $\square$

Proposition 10.22 is a special case of Plancherel theorem.

Proposition 10.23. *(Convolution rule) For all $x, y \in \mathrm{L}^2(G)$, we have*

$$\mathcal{F}(x * y) = \mathcal{F}(x)\mathcal{F}(y).$$

Proof. Recall from Example 10.1 that the convolution $x * y$ of $x, y \in \mathrm{L}^2(G)$ is given by

$$(x * y)_a = \frac{1}{|G|} \sum_{b\in G} x_b y_{b^{-1}a} = \frac{1}{|G|} \sum_{\substack{b,c\in G \\ b+c=a}} x_b y_c.$$

Thus we have

$$\begin{aligned}
\mathcal{F}(x * y)(\chi) &= \frac{1}{|G|} \sum_{c\in G} (x * y)_c \overline{\chi(c)} \\
&= \frac{1}{|G|} \sum_{c\in G} \frac{1}{|G|} \sum_{d\in G} x_d y_{d^{-1}c} \overline{\chi(c)}.
\end{aligned}$$

If we replace c by dc, since G is a group the sum does not change, and we get

$$\begin{aligned}\frac{1}{|G|}\sum_{c\in G}\frac{1}{|G|}\sum_{d\in G}x_d y_{d^{-1}c}\overline{\chi(c)} &= \frac{1}{|G|}\sum_{dc\in G}\frac{1}{|G|}\sum_{d\in G}x_d y_{d^{-1}dc}\overline{\chi(dc)}\\ &= \frac{1}{|G|}\sum_{d\in G}x_d\frac{1}{|G|}\sum_{c\in G}y_c\overline{\chi(d)}\,\overline{\chi(c)}\\ &= \frac{1}{|G|}\sum_{d\in G}x_d\overline{\chi(d)}\frac{1}{|G|}\sum_{c\in G}y_c\overline{\chi(c)}\\ &= \mathcal{F}(x)(\chi)\mathcal{F}(y)(\chi).\end{aligned}$$

Therefore, we proved that

$$\mathcal{F}(x * y) = \mathcal{F}(x)\mathcal{F}(y).$$

□

10.5 Dirichlet Characters

Let $G = (\mathbb{Z}/m\mathbb{Z})^*$ equipped with the counting measure. The group $(\mathbb{Z}/m\mathbb{Z})^*$ is the multiplicative group of units in $\mathbb{Z}/m\mathbb{Z}$, that is, those elements $a \in \mathbb{Z}/m\mathbb{Z}$ such that there is some $b \in \mathbb{Z}/m\mathbb{Z}$ and $ab = ba = 1 \pmod m$. It is well known that $a \in \mathbb{Z}/m\mathbb{Z}$ is a unit if and only if $\gcd(a, m) = 1$, and the group $(\mathbb{Z}/m\mathbb{Z})^*$ has $\varphi(m)$ elements, where φ is the Euler phi-function ($\varphi(m)$ is the number of integers a, with $1 \leq a \leq m$, such that $\gcd(a, m) = 1$). The group $(\mathbb{Z}/m\mathbb{Z})^*$ is not always cyclic, (for example, if $m = 2^k, k \geq 3$), but Gauss determined when this happens. However, when $(\mathbb{Z}/m\mathbb{Z})^*$ is cyclic, finding a generator for it is computationally hard.

Definition 10.5. The characters $\chi\colon (\mathbb{Z}/m\mathbb{Z})^* \to \mathbf{U}(1)$ are called *Dirichlet characters*. Every such character can be extended to a function $\chi\colon \mathbb{Z} \to \mathbb{C}$ given by

$$\chi(n) = \begin{cases} \chi(n \bmod m) & \text{if } \gcd(m, n) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

It is immediately verified that such functions are multiplicative, which means that

$$\chi(rs) = \chi(r)\chi(s) \qquad \text{for all } r, s \in \mathbb{Z}.$$

They are also periodic with period m ($\chi(n + m) = \chi(n)$ for all $n \in \mathbb{Z}$). These functions are called *Dirichlet characters modulo m*. The trivial Dirichlet character is the Dirichlet character such that $\chi_0(m) = 1$ iff m and n are relatively prime, and 0 otherwise.

Definition 10.6. For every $\ell \in \mathbb{N} - \{0\}$, define the function $\delta_\ell\colon (\mathbb{Z}/m\mathbb{Z})^* \to \mathbb{C}$ given by

$$\delta_\ell(n) = \begin{cases} 1 & \text{if } n \equiv \ell \pmod m \\ 0 & \text{otherwise.} \end{cases}$$

Example 10.4. Let $G = (\mathbb{Z}/10\mathbb{Z})^* = \{1, 3, 7, 9\}$. The multiplication table for $(\mathbb{Z}/10\mathbb{Z})^*$ is shown below.

$\cdot$	1	3	7	9
1	1	3	7	9
3	3	9	1	7
7	7	1	9	3
9	9	7	3	1

Let $\ell = 13$. Since $13 \equiv 3 \pmod{10}$, we find that

$$\delta_{13}(1) = 0, \qquad \delta_{13}(3) = 1, \qquad \delta_{13}(7) = 0, \qquad \delta_{13}(9) = 0.$$

By the Fourier inversion formula (Proposition 10.20) we can write

$$\delta_\ell(n) = \sum_{\chi \in \widehat{G}} \mathcal{F}(\delta_\ell)(\chi)\chi(n),$$

with

$$\mathcal{F}(\delta_\ell)(\chi) = \frac{1}{\varphi(m)} \sum_{k \in G} \delta_\ell(k)\overline{\chi(k)} = \frac{1}{\varphi(m)}\overline{\chi(\ell)}.$$

Therefore,

$$\delta_\ell(n) = \frac{1}{\varphi(m)} \sum_{\chi \in \widehat{G}} \chi(n)\overline{\chi(\ell)}.$$

Like the characters, the functions δ_ℓ can be extended to $\mathbb{Z}$, by setting $\delta_\ell(n) = 0$ if m and n are not relatively prime. The above shows that

$$\delta_\ell(n) = \frac{1}{\varphi(m)} \sum_{\chi} \chi(n)\overline{\chi(\ell)}, \quad n \in \mathbb{Z},$$

where the sum is over the Dirichlet characters modulo m.

The above result is one of the steps in the proof of Dirichlet's theorem on arithmetic progressions of integers $mk + \ell$ with $\gcd(\ell, m) = 1$ and $k \in \mathbb{N}$, which says that such a sequence contains infinitely many primes.

Dirichlet's theorem is a consequence of the fact that the sum

$$\sum_{\substack{p \equiv \ell \pmod{m} \\ p \text{ prime}}} \frac{1}{p}$$

is infinite. This follows from the fact the the limit of the sum

$$\sum_{\substack{p \equiv \ell \pmod{m} \\ p \text{ prime}}} \frac{1}{p^s}$$

tends to infinity when $s > 1$ tends to 1. To simplify notation, write

$$\sum_{p\equiv\ell} \frac{1}{p^s},$$

where it is understood that p is prime and congruent to ℓ modulo m.

We can write

$$\sum_{p\equiv\ell} \frac{1}{p^s} = \sum_{p \text{ prime}} \frac{\delta_\ell(p)}{p^s} = \frac{1}{\varphi(m)} \sum_{\chi} \overline{\chi(\ell)} \sum_{p \text{ prime}} \frac{\chi(p)}{p^s}.$$

We can divide the above sum into two parts depending on whether or not χ is trivial, and we get

$$\begin{aligned}\sum_{p\equiv\ell} \frac{1}{p^s} &= \frac{1}{\varphi(m)} \sum_{p} \frac{\chi_0(p)}{p^s} + \frac{1}{\varphi(m)} \sum_{\chi\neq\chi_0} \overline{\chi(\ell)} \sum_{p} \frac{\chi(p)}{p^s}\\ &= \frac{1}{\varphi(m)} \sum_{p\nmid m} \frac{1}{p^s} + \frac{1}{\varphi(m)} \sum_{\chi\neq\chi_0} \overline{\chi(\ell)} \sum_{p} \frac{\chi(p)}{p^s},\end{aligned}$$

where the sums on the right-hand side are over all primes p. Since there are only finitely many primes dividing m, the sum

$$\sum_{p\nmid m} \frac{1}{p^s}$$

tends to infinity when $s > 1$ tends to 1, because the series

$$\sum_{p \text{ prime}} \frac{1}{p}$$

diverges. This is a classical result of number theory going back to Euler (1737); see Stein and Shakarchi [94] (Chapter 8), or Apostol [2] (Chapter 1, Section 1.6).

If we could prove that the sum

$$\sum_{p \text{ prime}} \frac{\chi(p)}{p^s}$$

remains bounded when $s > 1$ tends to 1, we would be done, because then

$$\sum_{p\equiv\ell} \frac{1}{p^s}$$

tends to infinity as $s > 1$ tends to 1.

The above suggests studying the behavior of the functions $L(s, \chi)$ given by

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}, \qquad s > 1,$$

where χ is a Dirichlet character, called *L-functions*. The series $L(s,\chi)$ is absolutely convergent for $s > 1$. Actually, if $\chi \neq \chi_0$, the series $L(s,\chi)$ converges for $s > 0$ and is continuously differentiable for $s > 0$; see Stein and Shakarchi [94] (Chapter 8). There is a also a remarkable product formula: For $s > 1$, we have

$$L(s,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - \chi(p)p^{-s}}.$$

The above is a generalization of Euler's formula for expressing the zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

as the infinite product

$$\prod_{p \text{ prime}} \frac{1}{1 - p^{-s}}.$$

In a tour de force Dirichlet proved that $L(1,\chi)$ is finite and that $L(1,\chi) \neq 0$ if χ is not the trivial character. By taking the logarithm of both sides of the product formula and using some properties of the log function one obtains

$$\log L(s,\chi) = \sum_{p \text{ prime}} \frac{\chi(p)}{p^s} + O(1),$$

so if $L(1,\chi)$ is nonzero, then the sum

$$\sum_{p \text{ prime}} \frac{\chi(p)}{p^s}$$

remains bounded when $s > 1$ tends to 1, and the proof of Dirichlet's theorem is completed.

The above sketch lacks rigor on several fronts, and a rigorous proof involves a lot of rather difficult technical details, the hardest step being the proof that $L(1,\chi) \neq 0$ if χ is not the trivial character. We refer the interested reader to Stein and Shakarchi [94] (Chapter 8), Apostol [2] (Chapters 6 and 7), or Serre [89] (Chapter VI).

10.6 Fourier Transform and Cotransform in Terms of Matrices

In this section we formulate the Fourier transform (and cotransform) on a finite abelian group in terms of matrices. Write $n = |G|$, and denote by $[G \to \mathbb{C}]$ the vector space of all functions $x \colon G \to \mathbb{C}$. Since G is finite, we have $\mathrm{L}^1(G) = \mathrm{L}^2(G) = [G \to \mathbb{C}]$. Of course, the space $[G \to \mathbb{C}]$ is isomorphic to $\mathbb{C}^n$, but it is better to stick with $[G \to \mathbb{C}]$. Our goal is to

extend the Fourier transform $\mathcal{F}$ on $\mathrm{L}^2(G) = [G \to \mathbb{C}]$ (defined on $\widehat{G}$) to a sesquilinear form on $[G \to \mathbb{C}]^*$, and to extend the Fourier transform $\overline{\mathcal{F}}$ on $\mathrm{L}^2(\widehat{G}) = [\widehat{G} \to \mathbb{C}]$ (defined on $\widehat{\widehat{G}}$) to a bilinear form on $[G \to \mathbb{C}]^{**}$. We will also prove the Fourier inversion theorem for these extensions.

For any $a \in G$, we denote by $e_a \in [G \to \mathbb{C}]$ the map given by

$$e_a(b) = \begin{cases} 1 & \text{if } b = a \\ 0 & \text{if } b \neq a. \end{cases}$$

If we order G as $(a_1, \ldots, a_n)$, then $(e_{a_1}, \ldots, e_{a_n})$ is a basis of $[G \to \mathbb{C}]$. Viewed as an element of $\mathbb{C}^n$, the vector e_{a_i} corresponds to the canonical ith basis vector e_i. For any function $x \in [G \to \mathbb{C}]$, we can write

$$x = \sum_{i=1}^{n} x(a_i) e_{a_i}.$$

For simplicity of notation, we write x_i (or x_{a_i}) instead of $x(a_i)$. To show that Fourier inversion holds, we need to view a function $x \in [G \to \mathbb{C}]$ as a linear form in $[G \to \mathbb{C}]^*$.

Definition 10.7. Let G be a finite abelian group and write $G = \{a_1, \ldots, a_n\}$. Every vector $x \in [G \to \mathbb{C}]$ determines uniquely the linear form $\widetilde{x} \in [G \to \mathbb{C}]^*$ defined on the basis $(e_{a_1}, \ldots, e_{a_n})$ of $[G \to \mathbb{C}]$ by

$$\widetilde{x}(e_{a_j}) = x(a_j).$$

In terms of the dual basis $(e^*_{a_1}, \ldots, e^*_{a_n})$ of the basis $(e_{a_1}, \ldots, e_{a_n})$, we have

$$\widetilde{x} = \sum_{j=1}^{n} x(a_j) e^*_{a_j}.$$

Since G is finite, we know that $\widehat{G}$ is isomorphic to G, and we order $\widehat{G}$ as $(\chi_1, \ldots, \chi_n)$. Every character $\chi_i \in \widehat{G}$ is a function $\chi_i \colon G \to \mathbb{C}$, so $\chi_i \in [G \to \mathbb{C}]$, and the orthogonality conditions of Proposition 10.13 imply that $(\chi_1, \ldots, \chi_n)$ are linearly independent.

As above, every character χ_i determines uniquely the linear form $\widetilde{\chi_i} \in [G \to \mathbb{C}]^*$ given by

$$\widetilde{\chi_i} = \sum_{j=1}^{n} \chi_i(a_j) e^*_{a_j}.$$

If there is a linear dependence

$$\sum_{j=1}^{n} \alpha_j \widetilde{\chi_j} = 0,$$

by applying the above linear form to the n vectors e_{a_k}, we obtain the n equations

$$\sum_{j=1}^{n} \alpha_j \widetilde{\chi_j}(e_{a_k}) = \sum_{j=1}^{n} \alpha_j \chi_j(a_k) = 0, \qquad k = 1, \ldots, n,$$

which are equivalent to

$$\sum_{j=1}^{n} \alpha_j(\chi_j(a_1), \ldots, \chi_j(a_n)) = 0,$$

that is

$$\sum_{j=1}^{n} \alpha_j \chi_j = 0,$$

and since $(\chi_1, \ldots, \chi_n)$ are linearly independent, we must have $\alpha_1 = \cdots = \alpha_n = 0$, and so the linear forms $(\widetilde{\chi_1}, \ldots, \widetilde{\chi_n})$ are also linearly independent. Thus they form a basis of $[G \to \mathbb{C}]^*$.

Our first goal is to extend the Fourier transform $\mathcal{F}$ on $\mathrm{L}^2(G) = [G \to \mathbb{C}]$ to a sesquilinear form on $[G \to \mathbb{C}]^*$. Given any $x \in [G \to \mathbb{C}]$, its Fourier transform is the map $\mathcal{F}(x)\colon \widehat{G} \to \mathbb{C}$ given by

$$\mathcal{F}(x)(\chi_i) = \frac{1}{n}\sum_{j=1}^{n} x(a_j)\overline{\chi_i(a_j)} = \frac{1}{n}\sum_{j=1}^{n} \widetilde{x}(e_{a_j})\overline{\widetilde{\chi_i}(e_{a_j})}. \tag{$*$}$$

Definition 10.8. Let G be a finite abelian group and write $G = \{a_1, \ldots, a_n\}$. For all $f, \gamma \in [G \to \mathbb{C}]^*$ we define the *Fourier transform* $\mathcal{F}$ as the sequilinear form on $[G \to \mathbb{C}]^*$ given by

$$\mathcal{F}(f)(\gamma) = \frac{1}{n}\sum_{j=1}^{n} f(e_{a_j})\overline{\gamma(e_{a_j})}.$$

If $f = \widetilde{x}$ and $\gamma = \widetilde{\chi_i}$, the right-hand side of the above definition is equal to the right-hand side of $(*)$, so this definition extends the definition of the Fourier transform on $[G \to \mathbb{C}]$ and $\widehat{G}$.

If we write $f = \sum_{i=1}^{n} x_i e^*_{a_i}$, then we have

$$\mathcal{F}(f)(\gamma) = \sum_{i=1}^{n} x_i \mathcal{F}(e^*_{a_i})(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i e^*_{a_i}(e_{a_j})\overline{\gamma(e_{a_j})} = \frac{1}{n}\sum_{j=1}^{n} x_j \overline{\gamma(e_{a_j})}.$$

Then if $\gamma = \sum_{i=1} w_i \widetilde{\chi_i}$, we get

$$\begin{aligned}
\mathcal{F}(f)\left(\sum_{i=1}^{n} w_i \widetilde{\chi_i}\right) &= \sum_{i=1}^{n} \overline{w_i}\mathcal{F}(f)(\widetilde{\chi_i}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \overline{w_i}\sum_{j=1}^{n} x_j \overline{\widetilde{\chi_i}(e_{a_j})} \\
&= \frac{1}{n}\sum_{i=1}^{n} \overline{w_i}\sum_{j=1}^{n} x_j \overline{\chi_i(a_j)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \overline{\chi_i(a_j)} x_j\right)\overline{w_i}.
\end{aligned}$$

Definition 10.9. Let G be a finite abelian group of order n, and write $G = \{a_1, \ldots, a_n\}$. Define the $n \times n$ matrix $F = (F_{ij})$, called the *Fourier matrix* of G with respect to the characters $\{\chi_1, \ldots, \chi_n\}$ of G, by

$$F_{ij} = \overline{\chi_i(a_j)}.$$

Example 10.5. As a concrete example of the preceding calculations set $n = 3$. Then

$$f = x_1 e^*_{a_1} + x_2 e^*_{a_2} + x_3 e^*_{a_3}, \qquad \gamma = w_1 \widetilde{\chi_1} + w_2 \widetilde{\chi_2} + w_3 \widetilde{\chi_3},$$

and

$$\widetilde{\chi_j} = \sum_{i=1}^{3} \chi_j(a_i) e^*_{a_i}, \qquad 1 \leq j \leq 3.$$

Since $f(e_{a_j}) = x_j$ whenever $1 \leq j \leq 3$, Definition 10.8 implies that

$$\mathcal{F}(f)(\gamma) = \frac{1}{3} \sum_{i=1}^{3} x_i \mathcal{F}(e^*_{a_i})(\gamma) = \frac{1}{3} \left[x_1 \overline{\gamma(e_{a_1})} + x_2 \overline{\gamma(e_{a_2})} + x_3 \overline{\gamma(e_{a_3})} \right].$$

Next we need to evaluate $\overline{\gamma(e_{a_j})}$ for $1 \leq j \leq 3$. We will demonstrate the evaluation of $\overline{\gamma(e_{a_1})}$ and leave the other two cases to the reader. By using the definition of γ and the fact each $\widetilde{\chi_i}$ can be expanded in terms of the dual basis $\left(e^*_{a_1}, e^*_{a_2}, e^*_{a_3}\right)$, we find that

$$\begin{aligned}
\gamma(e_{a_1}) &= w_1 \widetilde{\chi_1}(e_{a_1}) + w_2 \widetilde{\chi_2}(e_{a_1}) + w_3 \widetilde{\chi_3}(e_{a_1}) \\
&= w_1 \left[\chi_1(a_1) e^*_{a_1}(e_{a_1}) + \chi_1(a_2) e^*_{a_2}(e_{a_1}) + \chi_1(a_3) e^*_{3_1}(e_{a_1}) \right] \\
&\quad + w_2 \left[\chi_2(a_1) e^*_{a_1}(e_{a_1}) + \chi_2(a_2) e^*_{a_2}(e_{a_1}) + \chi_2(a_3) e^*_{3_1}(e_{a_1}) \right] \\
&\quad + w_3 \left[\chi_3(a_1) e^*_{a_1}(e_{a_1}) + \chi_3(a_2) e^*_{a_2}(e_{a_1}) + \chi_3(a_3) e^*_{3_1}(e_{a_1}) \right] \\
&= w_1 \chi_1(a_1) + w_2 \chi_2(a_1) + w_3 \chi_3(a_1).
\end{aligned}$$

Hence

$$\overline{\gamma(e_{a_1})} = \overline{w_1} \overline{\chi_1(a_1)} + \overline{w_2} \overline{\chi_2(a_1)} + \overline{w_3} \overline{\chi_3(a_1)}.$$

Similar calculations show that

$$\begin{aligned}
\overline{\gamma(e_{a_2})} &= \overline{w_1} \overline{\chi_1(a_2)} + \overline{w_2} \overline{\chi_2(a_2)} + \overline{w_3} \overline{\chi_3(a_2)}, \\
\overline{\gamma(e_{a_3})} &= \overline{w_1} \overline{\chi_1(a_3)} + \overline{w_2} \overline{\chi_2(a_3)} + \overline{w_3} \overline{\chi_3(a_3)}.
\end{aligned}$$

Our calculations, when written in matrix form, become

$$\mathcal{F}(f)(\gamma) = \frac{1}{3} \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} \overline{\gamma(e_{a_1})} \\ \overline{\gamma(e_{a_2})} \\ \overline{\gamma(e_{a_3})} \end{pmatrix},$$

where

$$\begin{pmatrix} \overline{\gamma(e_{a_1})} \\ \overline{\gamma(e_{a_2})} \\ \overline{\gamma(e_{a_3})} \end{pmatrix} = \begin{pmatrix} \overline{\chi_1(a_1)} & \overline{\chi_2(a_1)} & \overline{\chi_3(a_1)} \\ \overline{\chi_1(a_2)} & \overline{\chi_2(a_2)} & \overline{\chi_2(a_3)} \\ \overline{\chi_1(a_3)} & \overline{\chi_2(a_3)} & \overline{\chi_3(a_3)} \end{pmatrix} \begin{pmatrix} \overline{w_1} \\ \overline{w_2} \\ \overline{w_3} \end{pmatrix} = F^{\top} \begin{pmatrix} \overline{w_1} \\ \overline{w_2} \\ \overline{w_3} \end{pmatrix},$$

with F being the corresponding Fourier matrix of Definition 10.9, namely

$$F = \begin{pmatrix} \overline{\chi_1(a_1)} & \overline{\chi_1(a_2)} & \overline{\chi_1(a_3)} \\ \overline{\chi_2(a_1)} & \overline{\chi_2(a_2)} & \overline{\chi_2(a_3)} \\ \overline{\chi_3(a_1)} & \overline{\chi_3(a_2)} & \overline{\chi_3(a_3)} \end{pmatrix}.$$

Hence

$$\mathcal{F}(f)(\gamma) = \frac{1}{3} \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} F^\top \begin{pmatrix} \overline{w_1} \\ \overline{w_2} \\ \overline{w_3} \end{pmatrix}.$$

In summary we obtained the following result.

Proposition 10.24. *If $f \in [G \to \mathbb{C}]^*$ is expressed over the basis $(e^*_{a_1}, \ldots, e^*_{a_n})$ as $f = \sum_{j=1}^n x_j e^*_{a_j}$, and if $\gamma \in [G \to \mathbb{C}]^*$ is expressed over the basis $(\widetilde{\chi_1}, \ldots, \widetilde{\chi_n})$ as $\gamma = \sum_{i=1}^n w_i \widetilde{\chi_i}$, then the value $\mathcal{F}(f)(\gamma)$ of the sequilinear form $\mathcal{F}$ is given by*

$$\mathcal{F}(f)(\gamma) = \frac{1}{n} w^* F x = \frac{1}{n} x^\top F^\top \overline{w}.$$

As a semilinear map from $[G \to \mathbb{C}]^*$ to $\mathbb{C}$, the matrix of $\mathcal{F}(f)$ over the basis $(\widetilde{\chi_1}, \ldots, \widetilde{\chi_n})$ is the row vector $\frac{1}{n} x^\top F^\top$. As a semilinear form on $[G \to \mathbb{C}]^*$, the semilinear form $\mathcal{F}(f)$ is represented by the column vector $\frac{1}{n} F x$ over the dual basis $(\widetilde{\chi_1}^*, \ldots, \widetilde{\chi_n}^*)$.

Our next goal is to extend the Fourier cotransform $\overline{\mathcal{F}}$ on $\mathrm{L}^2(\widehat{G}) = [\widehat{G} \to \mathbb{C}]$ to a bilinear form on $[G \to \mathbb{C}]^{**}$. Given any function $\xi \in [\widehat{G} \to \mathbb{C}]$, the Fourier cotransform $\overline{\mathcal{F}}(\xi)$ of ξ is the map $\overline{\mathcal{F}}(\xi) \colon \widehat{\widehat{G}} \to \mathbb{C}$ given by

$$\overline{\mathcal{F}}(\xi)(\zeta) = \sum_{j=1}^n \xi(\chi_j) \zeta(\chi_j), \qquad \zeta \in \widehat{\widehat{G}},$$

with $\widehat{G} = \{\chi_1, \ldots, \chi_n\}$.

Since every character in $\widehat{G}$ is a function in $[G \to \mathbb{C}]$, every function $\xi \in [\widehat{G} \to \mathbb{C}]$ can be viewed as a function in $[[G \to \mathbb{C}] \to \mathbb{C}]$. Similarly, every character $\zeta \in \widehat{\widehat{G}}$ is a function in $[[G \to \mathbb{C}] \to \mathbb{C}]$, and we know that $\widehat{\widehat{G}}$ is isomorphic to G.

Recall that every function $x \in [G \to \mathbb{C}]$ defines uniquely a linear form $\widetilde{x} \in [G \to \mathbb{C}]^*$. Similarly, every function $\xi \in [[G \to \mathbb{C}] \to \mathbb{C}]$ can be extended uniquely to the linear form $\widetilde{\xi} \in [G \to \mathbb{C}]^{**}$ defined on the basis $(\widetilde{\chi_1}, \ldots, \widetilde{\chi_n})$ of $[G \to \mathbb{C}]^*$ by

$$\widetilde{\xi}(\widetilde{\chi_i}) = \xi(\chi_i).$$

In terms of the dual basis $(\widetilde{\chi_1}^*, \ldots, \widetilde{\chi_n}^*)$ of the basis $(\widetilde{\chi_1}, \ldots, \widetilde{\chi_n})$, we have

$$\widetilde{\xi} = \sum_{i=1}^{n} \xi(\chi_i)\widetilde{\chi_i}^*.$$

The vector space $[\widehat{G} \to \mathbb{C}]$ is isomorphic to $[G \to \mathbb{C}]^{**}$. Using the basis $(\widetilde{\chi_1}^*, \ldots, \widetilde{\chi_n}^*)$ in $[G \to \mathbb{C}]^{**}$, any $\omega \in [G \to \mathbb{C}]^{**}$ can be expressed as $\omega = \sum_{i=1}^{n} y_i\widetilde{\chi_i}^*$.

Since $\overline{\mathcal{F}}(\xi)(\zeta)$ is defined as

$$\overline{\mathcal{F}}(\xi)(\zeta) = \sum_{j=1}^{n} \xi(\chi_j)\zeta(\chi_j),$$

we have

$$\overline{\mathcal{F}}(\xi)(\zeta) = \sum_{j=1}^{n} \xi(\chi_j)\zeta(\chi_j) = \sum_{j=1}^{n} \widetilde{\xi}(\widetilde{\chi_j})\widetilde{\zeta}(\widetilde{\chi_j}).$$

Definition 10.10. Let G be a finite abelian group, write $G = \{a_1, \ldots, a_n\}$, and let $\{\chi_1, \ldots, \chi_n\}$ be the characters of G. For all $\omega, \gamma \in [G \to \mathbb{C}]^{**}$, we define the *Fourier cotransform* $\overline{\mathcal{F}}$ as the bilinear form on $[G \to \mathbb{C}]^{**}$ given by

$$\overline{\mathcal{F}}(\omega)(\gamma) = \sum_{j=1}^{n} \omega(\widetilde{\chi_j})\gamma(\widetilde{\chi_j}).$$

If $\omega = \widetilde{\xi}$ and $\gamma = \widetilde{\zeta}$, then the right-hand side of the above equation is equal to $\overline{\mathcal{F}}(\xi)(\zeta)$, so this definition extends the Fourier cotransform on $[\widehat{G} \to \mathbb{C}]$ and $\widehat{\widehat{G}}$.

If we write

$$\omega = \sum_{i=1}^{n} y_i\widetilde{\chi_i}^*,$$

then we have

$$\overline{\mathcal{F}}(\omega)(\gamma) = \sum_{i=1}^{n} y_i\overline{\mathcal{F}}(\widetilde{\chi_i}^*)(\gamma) = \sum_{i=1}^{n}\sum_{j=1}^{n} y_i\widetilde{\chi_i}^*(\widetilde{\chi_j})\gamma(\widetilde{\chi_j}) = \sum_{j=1}^{n} y_j\gamma(\widetilde{\chi_j}).$$

We also have a natural isomorphism η from $[G \to \mathbb{C}]$ to $[G \to \mathbb{C}]^{**}$, defined such that the linear form $\eta_u \in [G \to \mathbb{C}]^{**}$ is given by

$$\eta_u(\chi) = \chi(u) \qquad \chi \in [G \to \mathbb{C}]^*,\ u \in [G \to \mathbb{C}].$$

Observe that

$$e_{a_i}^{**} = \eta_{e_{a_i}},$$

because on the basis $(e^*_{a_1}, \ldots, e^*_{a_n})$ of $[G \to \mathbb{C}]^*$, we have

$$e^{**}_{a_i}(e^*_{a_j}) = \delta_{ij} = e^*_{a_j}(e_{a_i}) = \eta_{e_{a_i}}(e^*_{a_j}).$$

Then $(\eta_{e_{a_1}}, \ldots, \eta_{e_{a_n}}) = (e^{**}_{a_1}, \ldots, e^{**}_{a_n})$ is a basis of $[G \to \mathbb{C}]^{**}$, and if $\omega = \sum_{i=1}^n y_i \widetilde{\chi_i}^*$, then $\overline{\mathcal{F}}(\omega)$ is given by

$$\begin{aligned}
\overline{\mathcal{F}}(\omega)\left(\sum_{i=1}^n z_i \eta_{e_{a_i}}\right) &= \sum_{i=1}^n z_i \overline{\mathcal{F}}(\omega)(\eta_{e_{a_i}}) \\
&= \sum_{i=1}^n z_i \sum_{j=1}^n y_j \eta_{e_{a_i}}(\widetilde{\chi_j}) \\
&= \sum_{i=1}^n z_i \sum_{j=1}^n y_j \widetilde{\chi_j}(e_{a_i}) \\
&= \sum_{i=1}^n \left(\sum_{j=1}^n y_j \chi_j(a_i)\right) z_i \\
&= y^\top \overline{F} z = z^\top F^* y.
\end{aligned}$$

Example 10.6. Once again we provide a concrete demonstration of the preceding calculations. Set $n = 3$ and write Definition 10.10 as

$$\overline{\mathcal{F}}(\omega)(\gamma) = \omega(\widetilde{\chi_1})\gamma(\widetilde{\chi_1}) + \omega(\widetilde{\chi_2})\gamma(\widetilde{\chi_2}) + \omega(\widetilde{\chi_3})\gamma(\widetilde{\chi_3}).$$

Since $\omega = y_1\widetilde{\chi_1}^* + y_2\widetilde{\chi_2}^* + y_3\widetilde{\chi_3}^*$, a familiar calculation shows that preceding line becomes

$$\overline{\mathcal{F}}(\omega)(\gamma) = y_1\gamma(\widetilde{\chi_1}) + y_2\gamma(\widetilde{\chi_2}) + y_3\gamma(\widetilde{\chi_3}).$$

Now it is a matter of calculating $\gamma(\widetilde{\chi_i})$ for $1 \le i \le 3$. We will demonstrate in detail the calculation of $\gamma(\widetilde{\chi_1})$ and leave the other two cases to the reader. Since $\gamma = z_1 e^{**}_{a_1} + z_2 e^{**}_{a_2} + z_3 e^{**}_{a_3}$, we see that

$$\gamma(\widetilde{\chi_1}) = z_1 e^{**}_{a_1}(\widetilde{\chi_1}) + z_2 e^{**}_{a_2}(\widetilde{\chi_1}) + z_3 e^{**}_{a_3}(\widetilde{\chi_1}).$$

However, recall that

$$\widetilde{\chi_i} = \chi_i(a_1)e^*_{a_1} + \chi_i(a_2)e^*_{a_2} + \chi_i(a_3)e^*_{a_3}, \qquad 1 \le i \le 3.$$

Thus

$$\begin{aligned}
\gamma(\widetilde{\chi_1}) &= z_1\left[\chi_1(a_1)e^{**}_{a_1}(e^*_{a_1}) + \chi_1(a_2)e^{**}_{a_1}(e^*_{a_2}) + \chi_1(a_3)e^{**}_{a_1}(e^*_{a_3})\right] \\
&\quad + z_2\left[\chi_1(a_1)e^{**}_{a_2}(e^*_{a_1}) + \chi_1(a_2)e^{**}_{a_2}(e^*_{a_2}) + \chi_1(a_3)e^{**}_{a_2}(e^*_{a_3})\right] \\
&\quad + z_3\left[\chi_1(a_1)e^{**}_{a_3}(e^*_{a_1}) + \chi_1(a_2)e^{**}_{a_3}(e^*_{a_2}) + \chi_1(a_3)e^{**}_{a_3}(e^*_{a_3})\right] \\
&= z_1\chi_1(a_1) + z_2\chi_1(a_2) + z_3\chi_1(a_3).
\end{aligned}$$

Similar calculations show that

$$\gamma(\widetilde{\chi_2}) = z_1\chi_2(a_1) + z_2\chi_2(a_2) + z_3\chi_2(a_3),$$
$$\gamma(\widetilde{\chi_3}) = z_1\chi_3(a_1) + z_2\chi_3(a_2) + z_3\chi_3(a_3).$$

It remains to write these calculations in matrix form. Observe that

$$\overline{\mathcal{F}}(\omega)(\gamma) = \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix} \begin{pmatrix} \gamma(\widetilde{\chi_1}) \\ \gamma(\widetilde{\chi_2}) \\ \gamma(\widetilde{\chi_3}) \end{pmatrix},$$

with

$$\begin{pmatrix} \gamma(\widetilde{\chi_1}) \\ \gamma(\widetilde{\chi_2}) \\ \gamma(\widetilde{\chi_3}) \end{pmatrix} = \begin{pmatrix} \chi_1(a_1) & \chi_1(a_2) & \chi_1(a_3) \\ \chi_2(a_1) & \chi_2(a_2) & \chi_2(a_3) \\ \chi_3(a_1) & \chi_3(a_2) & \chi_3(a_3) \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \overline{F} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix},$$

where F is the Fourier matrix of Example 10.5. Thus we ultimately obtain

$$\overline{\mathcal{F}}(\omega)(\gamma) = \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix} \overline{F} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}.$$

In summary, we proved the following result.

Proposition 10.25. *If $\omega \in [G \to \mathbb{C}]^{**}$ is expressed over the basis $(\widetilde{\chi_1}^*, \ldots, \widetilde{\chi_n}^*)$ as $\omega = \sum_{i=1}^n y_i \widetilde{\chi_i}^*$, and if $\zeta \in [G \to \mathbb{C}]^{**}$ is expressed over the basis $(e_{a_1}^{**}, \ldots, e_{a_n}^{**})$ as $\zeta = \sum_{i=1}^n z_i e_{a_i}^{**}$, then the value $\overline{\mathcal{F}}(\omega)(\zeta)$ of the bilinear form $\overline{\mathcal{F}}$ is given by*

$$\overline{\mathcal{F}}(\omega)(\zeta) = y^\top \overline{F} z = z^\top F^* y.$$

As a linear map from $[G \to \mathbb{C}]^{**} \cong [G \to \mathbb{C}]$ to $\mathbb{C}$, the matrix of $\overline{\mathcal{F}}(\omega)$ over the basis $(e_{a_1}^{**}, \ldots, e_{a_n}^{**})$ is the row vector $y^\top \overline{F}$. As an element of $[G \to \mathbb{C}]^{***} \cong [G \to \mathbb{C}]^*$, the linear form $\overline{\mathcal{F}}(\omega)$ is represented by the column vector $F^* y$ over the dual basis $(e_{a_1}^{***}, \ldots, e_{a_n}^{***})$.

The orthogonality conditions of Proposition 10.13 imply that

$$\frac{1}{n} F F^* = \frac{1}{n} F^* F = I,$$

so F^* is the inverse of $\frac{1}{n} F$.

Now if $f \in [G \to \mathbb{C}]^*$ with $f = \sum_{j=1}^n x_j e_{a_j}^*$, then $\mathcal{F}(f) \in [G \to \mathbb{C}]^{**}$, and since $\mathcal{F}(f)$ is expressed over the basis $(\widetilde{\chi_1}^*, \ldots, \widetilde{\chi_n}^*)$ and $\overline{\mathcal{F}}$ is also defined over this basis, with $\omega = \mathcal{F}(f)$, we know that ω is represented by the column vector $y = Fx$ over the basis $(\widetilde{\chi_1}^*, \ldots, \widetilde{\chi_n}^*)$, and if $\zeta = \sum_{i=1}^n z_i e_{a_i}^{**} = \sum_{i=1}^n z_i \eta_{e_{a_i}}$, then $\overline{\mathcal{F}}(\omega)(\zeta)$ is given by

$$\overline{\mathcal{F}}(\omega)(\zeta) = z^\top F^* y,$$

so we get

$$(\overline{\mathcal{F}}(\mathcal{F}(f)))(\zeta) = z^\top \frac{1}{n} F^* F x = z^\top x = f\left(\sum_{i=1}^n z_i e_{a_i}\right).$$

But

$$\zeta = \sum_{i=1}^n z_i \eta_{e_{a_i}} = \eta_{\sum_{i=1}^n z_i e_{a_i}},$$

so the above equation shows that

$$(\overline{\mathcal{F}} \circ \mathcal{F})(f) \circ \eta_{\sum_{i=1}^n z_i e_{a_i}} = f\left(\sum_{i=1}^n z_i e_{a_i}\right),$$

that is,

$$f = (\overline{\mathcal{F}} \circ \mathcal{F})(f) \circ \eta.$$

In summary, we proved the following theorem.

Theorem 10.26. *Let G be a finite abelian group of order n. The Fourier transform $\mathcal{F}$ defined on $[G \to \mathbb{C}]^*$ in Definition 10.8 and the Fourier cotransform $\overline{\mathcal{F}}$ defined on $[G \to \mathbb{C}]^{**}$ in Definition 10.10 satisfy the Fourier inversion equation*

$$f = (\overline{\mathcal{F}} \circ \mathcal{F})(f) \circ \eta,$$

for all $f \in [G \to \mathbb{C}]^*$.

A nice feature of Definition 10.8 and Definition 10.10 is that they are intrinsic, that is, independent the choice of bases. A slight disadvantage is that the cotransform is defined on the double dual $[G \to \mathbb{C}]^{**}$ of $[G \to \mathbb{C}]$. But $[G \to \mathbb{C}]$ and $[G \to \mathbb{C}]^{**}$ are isomorphic (in fact, canonically), so there is an alternative method to define directly a Fourier cotransform on $[G \to \mathbb{C}]$. The slight complication is that we would like Fourier inversion to hold, and for this it appears that we need to use a noncanonical isomorphism, namely the isomorphism $\theta\colon [G \to \mathbb{C}] \to [G \to \mathbb{C}]^{**}$ defined as follows: if $c = \sum_{i=1}^n y_i e_{a_i}$, then

$$\theta_c = \sum_{i=1}^n y_i \widetilde{\chi}_i^{\,*}.$$

Definition 10.11. For any vector $c = \sum_{i=1}^n y_i e_{a_i} \in [G \to \mathbb{C}]$, the *Fourier cotransform* $\overline{\mathcal{F}}_2(c)$ of c is the linear form on $[G \to \mathbb{C}]$ defined such that for all $a \in [G \to \mathbb{C}]$, we have

$$\overline{\mathcal{F}}_2(c)(a) = \sum_{j=1}^n y_j \widetilde{\chi}_j(a).$$

Observe that as long as we express the vector $c \in [G \to \mathbb{C}]$ over the basis $(e_{a_1}, \ldots, e_{a_n})$, the map $\overline{\mathcal{F}}_2$ is bilinear. We may think of the components of c as the Fourier coefficients of some form $f \in [G \to \mathbb{C}]^*$, and

$$\overline{\mathcal{F}}_2(c) = \sum_{j=1}^{n} y_j \widetilde{\chi}_j$$

is the "Fourier series" associated with c (a linear form on $[G \to \mathbb{C}]$).

We can then repeat our familiar computation to prove that if $a = \sum_{j=1}^{m} z_j e_{a_j}$, then

$$\overline{\mathcal{F}}_2(c)(a) = y^\top \overline{F} z = z^\top F^* y.$$

As a linear map from $[G \to \mathbb{C}]$ to $\mathbb{C}$, the matrix of $\overline{\mathcal{F}}_2(c)$ over the basis $(e_{a_1}, \ldots, e_{a_n})$ is the row vector $y^\top \overline{F}$, and as an element of $[G \to \mathbb{C}]^*$, the linear form $\overline{\mathcal{F}}_2(c)$ is represented by the column vector $F^* y$.

In order to be able to compose $\mathcal{F}$ and $\overline{\mathcal{F}}_2$, we need to convert $\mathcal{F}(f) \in [G \to \mathbb{C}]^{**}$, the result of applying $\mathcal{F}$ to $f \in [G \to \mathbb{C}]^*$, to a vector in $[G \to \mathbb{C}]$, and we can do this by applying the isomorphism $\theta^{-1} \colon [G \to \mathbb{C}]^{**} \to [G \to \mathbb{C}]$. Then Fourier inversion becomes the identity

$$f = (\overline{\mathcal{F}}_2 \circ \theta^{-1} \circ \mathcal{F})(f),$$

for all $f \in [G \to \mathbb{C}]^*$, which is another way of stating Proposition 10.20.

Since the vector spaces $[G \to \mathbb{C}]$, $[G \to \mathbb{C}]^*$, $[G \to \mathbb{C}]^{**}$, and $[G \to \mathbb{C}]^{***}$, are all isomorphic (and isomorphic to $\mathbb{C}^n$, where $n = |G|$), if we are just interested in transformations on sequences of complex numbers of length n indexed by the elements of the group G, namely elements of $[G \to \mathbb{C}]$, we can formulate versions of the Fourier transform and of the Fourier cotransform in terms of the Fourier matrix defined in Definition 10.9.

Definition 10.12. Let G be a finite abelian group, $G = \{a_1, \ldots, a_n\}$, and let $\{\chi_1, \ldots, \chi_n\}$ be the characters of G. If $F = (\overline{\chi_i(a_j)})$ is the Fourier matrix of G (as in Definition 10.9), then for every sequence $x \in [G \to \mathbb{C}]$, the sequence

$$\widehat{x} = \mathcal{F}(x) = \frac{1}{n} F x$$

is called the *Fourier transform* of x, and given any sequence $\xi \in [G \to \mathbb{C}]$, the sequence

$$\overline{\mathcal{F}}(\xi) = F^* \xi$$

is called the *inverse Fourier transform* or *Fourier cotransform* of ξ.

Recall that

$$\frac{1}{n} F F^* = \frac{1}{n} F^* F = I,$$

so the two transforms are mutual inverses.

Recall that we proved earlier that

$$\mathcal{F}(x * y) = \mathcal{F}(x)\mathcal{F}(y),$$

where $x * y$ is the convolution of x and y, given by

$$(x * y)_a = \frac{1}{|G|}\sum_{b\in G} x_b y_{b^{-1}a} = \frac{1}{|G|}\sum_{\substack{b,c\in G\\ b+c=a}} x_b y_c.$$

In matrix terms, $\mathcal{F}(x)\mathcal{F}(y) = \widehat{x}\,\widehat{y}$ is the vector whose ath entry $(\widehat{x}\,\widehat{y})_a$ is the product of the ath entry $\widehat{x}_a$ of the vector $\widehat{x}$ by the ath entry $\widehat{y}_a$ of the vector $\widehat{y}$. In matrix terms, it can be expressed as

$$\mathrm{diag}(\widehat{x})\widehat{y},$$

where $\mathrm{diag}(\widehat{x})$ is the diagonal matrix whose diagonal entries are the entries in the vector $\widehat{x}$.

Other aspects of harmonic analysis on finite abelian groups can be found in Terras [97]. In the next section we consider the special case where $G = \mathbb{Z}/n\mathbb{Z}$.

10.7 The Discrete Fourier Transform (on $\mathbb{Z}/n\mathbb{Z}$)

If $G = \mathbb{Z}/n\mathbb{Z}$, then we know from Proposition 10.9(3) that the characters of $\mathbb{Z}/n\mathbb{Z}$ are the n homomorphisms χ_k given by

$$m \mapsto e^{2\pi imk/n}, \quad k = 0, 1, \ldots, n-1, \quad m \in \mathbb{Z}/n\mathbb{Z}.$$

Observe that the characters are indexed by $0, 1, \ldots, n-1$ rather than $1, 2, \ldots, n$, but this is actually more convenient in what follows. The complex numbers

$$\{1, e^{2\pi i/n}, e^{2\pi i2/n}, \ldots, e^{2\pi im/n}, \ldots, e^{2\pi i(n-1)/n}\}$$

(with $0 \leq m \leq n-1$) are the *nth roots of unity* (because obviously, $(e^{2\pi ik/n})^n = e^{2\pi ik} = 1$). They form a subgroup of $\mathbb{C}$ denoted $\mu_n(\mathbb{C})$ isomorphic to $\mathbb{Z}/n\mathbb{Z}$ under the isomorphism $k \mapsto e^{2\pi ik/n}$, for $k \in \mathbb{Z}/n\mathbb{Z}$.

Definition 10.13. The Fourier matrix $F_n = \left(\overline{\chi_k(m)}\right)_{\substack{0\leq k\leq n-1\\ 0\leq m\leq n-1}}$ is given by

$$F_n = \left(e^{-2\pi ikm/n}\right)_{\substack{0\leq k\leq n-1\\ 0\leq m\leq n-1}}.$$

The first row ($k = 0$) consists of 1's, the second row ($k = 1$) consists of the consecutive inverse powers ζ^{-m} of $\zeta = e^{2\pi i/n}$ (a primitive nth root of unity) for $m = 0, \ldots, n-1$,

$$(1 \quad e^{-2\pi i/n} \quad e^{-2\pi i2/n} \quad \ldots \quad e^{-2\pi im/n} \quad \ldots \quad e^{-2\pi i(n-1)/n}),$$

and the $(k+1)$th row ($0 \leq k \leq n-1$) consists of the kth powers of the entries in the second row,

$$\left(1 \quad (e^{-2\pi i/n})^k \quad (e^{-2\pi i2/n})^k \quad \ldots \quad (e^{-2\pi im/n})^k \quad \ldots \quad (e^{-2\pi i(n-1)/n})^k\right).$$

Observe that F_n is symmetric. It is also a Vandermonde matrix for the roots of unity

$$\{1, e^{-2\pi i/n}, e^{-2\pi i 2/n}, \ldots, e^{-2\pi i m/n}, \ldots, e^{-2\pi i(n-1)/n}\},$$

namely, with $\omega = \zeta^{-1} = e^{-2\pi i/n}$ (also a primitive nth root of unity), we have

$$F_n = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & \omega & \cdots & \omega^{n-2} & \omega^{n-1} \\ 1 & \omega^2 & \cdots & \omega^{2(n-2)} & \omega^{2(n-1)} \\ 1 & \omega^3 & \cdots & \omega^{3(n-2)} & \omega^{3(n-1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \omega^{n-2} & \cdots & \omega^{(n-2)^2} & \omega^{(n-2)(n-1)} \\ 1 & \omega^{n-1} & \cdots & \omega^{(n-1)(n-2)} & \omega^{(n-1)^2} \end{pmatrix}.$$

For example, if $n = 5$, we have

$$\begin{aligned} F_5 &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & e^{-2\pi i/5} & e^{-2\pi i 2/5} & e^{-2\pi i 3/5} & e^{-2\pi i 4/5} \\ 1 & e^{-2\pi i 2/5} & e^{-2\pi i 4/5} & e^{-2\pi i 6/5} & e^{-2\pi i 8/5} \\ 1 & e^{-2\pi i 3/5} & e^{-2\pi i 6/5} & e^{-2\pi i 9/5} & e^{-2\pi i 12/5} \\ 1 & e^{-2\pi i 4/5} & e^{-2\pi i 8/5} & e^{-2\pi i 12/5} & e^{-2\pi i 16/5} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & e^{-2\pi i/5} & e^{-2\pi i 2/5} & e^{-2\pi i 3/5} & e^{-2\pi i 4/5} \\ 1 & e^{-2\pi i 2/5} & e^{-2\pi i 4/5} & e^{-2\pi i/5} & e^{-2\pi i 3/5} \\ 1 & e^{-2\pi i 3/5} & e^{-2\pi i/5} & e^{-2\pi i 4/5} & e^{-2\pi i 2/5} \\ 1 & e^{-2\pi i 4/5} & e^{-2\pi i 3/5} & e^{-2\pi i 2/5} & e^{-2\pi i/5} \end{pmatrix}. \end{aligned}$$

Definition 10.14. Given a sequence $x = (x_0, \ldots, x_{n-1}) \in \mathbb{C}^n$, its Fourier transform, also called *discrete Fourier transform*, is

$$\widehat{x} = \frac{1}{n} F_n x.$$

We can think of $c = \widehat{x}$ as the sequence of Fourier coefficients of x.

Definition 10.15. Similarly, given a sequence $c = (c_0, \ldots, c_{n-1}) \in \mathbb{C}^n$, its *discrete inverse Fourier transform* (or *discrete Fourier cotransform*) is

$$\overline{\mathcal{F}}(c) = F_n^* c = \overline{F_n} c.$$

Definition 10.16. Every sequence $c = (c_0, \ldots, c_{n-1}) \in \mathbb{C}^n$ of "Fourier coefficients" determines a periodic function $f_c \colon \mathbb{R} \to \mathbb{C}$ (of period 2π) known as *discrete Fourier series, or phase polynomial*, defined such that

$$f_c(\theta) = c_0 + c_1 e^{i\theta} + \cdots + c_{n-1} e^{i(n-1)\theta} = \sum_{k=0}^{n-1} c_k e^{ik\theta}.$$

Then given any sequence $f = (f_0, \ldots, f_{n-1})$ of data points, it is desirable to find the "Fourier coefficients" $c = (c_0, \ldots, c_{n-1})$ of the discrete Fourier series f_c such that

$$f_c(2\pi k/n) = f_k,$$

for every k, $0 \leq k \leq n-1$.

The problem amounts to solving the linear system of n equations

$$c_0 + c_1 e^{2\pi i k/n} + \cdots + c_{n-1} e^{2\pi i k(n-1)/n} = f_k, \quad k = 0, \ldots, n-1,$$

which is just the system

$$\overline{F_n} c = f.$$

Since

$$\frac{1}{n} F_n \overline{F_n} = \frac{1}{n} \overline{F_n} F_n = I,$$

we see that c is given by

$$c = \frac{1}{n} F_n f = \widehat{f},$$

the discrete Fourier transform of f, so

$$c_k = \frac{1}{n} \sum_{j=0}^{n-1} f_j e^{-2\pi i jk/n}, \qquad k = 0, \ldots, n-1.$$

Example 10.7. Let us see how to obtain the "Fourier coefficients" $c = (c_0, c_1, c_2)$ when $n = 3$ and $f = (f_0, f_1, f_2)$. The condition $f_c(2\pi k/n) = f_k$ for $0 \leq k \leq 2$ translates into three equations

$$f_c(0) = f_0, \qquad f_c(2\pi/3) = f_1, \qquad f_c(4\pi/3) = f_2.$$

After expanding via the definition $f_c(\theta) = \sum_{k=0}^{n-1} c_k e^{ik\theta}$, the above three equations become

$$\begin{aligned} c_0 + c_1 + c_2 &= f_0 \\ c_0 + c_1 e^{2\pi i/3} + c_2 e^{4\pi i/3} &= f_1 \\ c_0 + c_1 e^{4\pi i/3} + c_2 e^{2\pi i/3} &= f_2. \end{aligned}$$

These three equations can be written as

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & e^{2\pi i/3} & e^{4\pi i/3} \\ 1 & e^{4\pi i/3} & e^{2\pi i/3} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}.$$

Since

$$F_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & e^{-2\pi i/3} & e^{-4\pi i/3} \\ 1 & e^{-4\pi i/3} & e^{-2\pi i/3} \end{pmatrix}$$

the above matrix system is equivalent to

$$\overline{F}_3 \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix},$$

which implies that

$$\begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \frac{1}{3} F_3 \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}.$$

Note the analogy with the case of $\mathbb{T}$ and $\mathbb{Z}$, where the Fourier cotransform $\overline{\mathcal{F}}(c)$ of the sequence $(c_m)_{m\in\mathbb{Z}}$ is given by

$$f(\theta) = \sum_{k=-\infty}^{\infty} c_k e^{ik\theta},$$

and the Fourier coefficients of the function $f\colon \mathbb{T} \to \mathbb{C}$ are given by the formulae

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx.$$

In $\mathbb{Z}/n\mathbb{Z}$, the convolution of two sequences $f = (f_0, \ldots, f_{n-1})$ and $g = (g_0, \ldots, g_{n-1})$ is given by

$$(f * g)_k = \frac{1}{n} \sum_{\substack{i,j\in\mathbb{Z}/n\mathbb{Z} \\ i+j\equiv k \pmod n}} f_i g_j, \qquad k = 0, \ldots, n-1.$$

It is remarkable that the convolution $f * g$ can be expressed in matrix form as

$$f * g = \frac{1}{n} H(f) g$$

for some matrix $H(f)$. The matrix $H(f)$ is a *circulant matrix*.

Definition 10.17. The *circular shift matrix S_n (of order n)* is defined as the matrix

$$S_n = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

consisting of cyclic permutations of its first column. For any sequence $f = (f_0, \ldots, f_{n-1}) \in \mathbb{C}^n$, we define the *circulant matrix $H(f)$* as

$$H(f) = \sum_{j=0}^{n-1} f_j S_n^j,$$

where $S_n^0 = I_n$, as usual.

For example, the circulant matrix associated with the sequence $f = (a, b, c, d)$ is

$$\begin{pmatrix} a & d & c & b \\ b & a & d & c \\ c & b & a & d \\ d & c & b & a \end{pmatrix}$$

It is not hard to prove that the convolution $f * g$ of two sequences $f = (f_0, \ldots, f_{n-1})$ and $g = (g_0, \ldots, g_{n-1})$ is given by

$$f * g = \frac{1}{n} H(f)\, g,$$

viewing f and g as column vectors.

Then the miracle (which is not too hard to prove!), is that we have

$$H(f)\overline{F_n} = \overline{F_n}\,\mathrm{diag}(n\widehat{f}), \tag{$\dagger$}$$

where $\mathrm{diag}(n\widehat{f})$ is the diagonal matrix whose diagonal entries are the elements of the vector $n\widehat{f}$, which means that the columns of the Fourier matrix $\overline{F_n}$ are the eigenvectors of the circulant matrix $H(f)$, and that the eigenvalue associated with the kth eigenvector is $(n\widehat{f})_k$, that is, n times the kth component of the Fourier transform $\widehat{f}$ of f (counting from 0).

To prove ($\dagger$), we first prove that the eigenvectors u_k of the circular shift matrix S_n (indexing from 0 to $n-1$) are the columns of $\overline{F_n}$, where the column of index k whose entries are

$$\begin{pmatrix} 1 & (e^{2\pi i/n})^k & (e^{2\pi i2/n})^k & \cdots & (e^{2\pi im/n})^k & \cdots & (e^{2\pi i(n-1)/n})^k \end{pmatrix}$$

is associated with the eigenvalue $e^{-2\pi ik/n}$.

Indeed, applying S_n to u_k, the last entry $e^{2\pi i(n-1)k/n} = e^{-2\pi ik/n}$ in u_k becomes the first entry in $S_n u_k$, so the corresponding eigenvalue is $e^{-2\pi ik/n}$. For example, if $n = 4$, since $(1, e^{2\pi i/4}, e^{2\pi i2/4}, e^{2\pi i3/4}) = (1, i, -1, -i)$ and $(1, e^{-2\pi i/4}, e^{-2\pi i2/4}, e^{-2\pi i3/4}) = (1, -i, -1, i)$, we have

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & i \end{pmatrix}.$$

Since $H(f) = \sum_{j=0}^{n-1} f_j S_n^j$, the eigenvectors remain the same, and it is easy to see that the kth eigenvalue of $H(f)$ (indexing from 0) is $\sum_{j=0}^{n-1} f_j e^{-2\pi ijk/n} = n\widehat{f}_k$.

If we recall that $F_n\overline{F_n} = \overline{F_n}F_n = n\, I_n$, multiplying the equation $H(f)\overline{F_n} = \overline{F_n}\,\mathrm{diag}(n\widehat{f})$ both on the left and on the right by F_n, we get

$$F_n H(f)(n\, I_n) = (n\, I_n)\,\mathrm{diag}(n\widehat{f})\, F_n,$$

that is,

$$F_n H(f) = \mathrm{diag}(n\widehat{f})\, F_n.$$

If we apply both sides to any sequence $g \in \mathbb{C}^n$, we get

$$F_n H(f)g = \mathrm{diag}(n\widehat{f})\, F_n g.$$

Since $\widehat{f} = \frac{1}{n}F_n f$, $\widehat{g} = \frac{1}{n}F_n g$, $f * g = \frac{1}{n}H(f)g$, and $\widehat{f * g} = \frac{1}{n}F_n(f * g)$, multiplying both sides by $1/n^2$, the above equation yields

$$\frac{1}{n}F_n \frac{1}{n}H(f)g = \mathrm{diag}(\widehat{f})\, \frac{1}{n}F_n g,$$

which means that

$$\widehat{f * g} = \mathrm{diag}(\widehat{f})\, \widehat{g} = \widehat{f}\, \widehat{g},$$

where $\widehat{f}\,\widehat{g}$ is the column vector obtained by pointwise multiplication $((\widehat{f}\,\widehat{g})(j) = \widehat{f}(j)\widehat{g}(f))$.

Therefore, we have given another proof of the convolution rule.

10.8 Plancherel's Theorem and Fourier Inversion

Let G be a locally compact abelian group equipped with a Haar measure λ. In general, given a function $f \in \mathrm{L}^1(G)$, its Fourier transform $\mathcal{F}(f)$ does not belong to $\mathrm{L}^1(\widehat{G})$.

Plancherel's theorem (Theorem 10.27) asserts that there is a Haar measure $\widehat{\lambda}$ on the dual group $\widehat{G}$ such that the map $f \mapsto \mathcal{F}(f)$ sends $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$ into $\mathrm{L}^2(\widehat{G})$ and has a unique extension which is an isometry from $\mathrm{L}^2(G)$ to $\mathrm{L}^2(\widehat{G})$.

We will follow Bourbaki's proof [9] (Chapter 2, Section 1, No. 3). The crucial step is to define a subspace $A(G)$ of $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$ which is dense in both $\mathrm{L}^1(G)$ and $\mathrm{L}^2(G)$, and to show that there is a Haar measure ν on $\widehat{G}$ such that

$$\int_{\widehat{G}} |\mathcal{F}(f)|^2\, d\nu = \int_G |f|^2\, d\lambda$$

for all $f \in A(G)$. There are many technical details so we will focus on the main ideas.

Definition 10.18. Let $A(G)$ be the subspace of $\mathrm{L}^1(G)$ spanned by the set of functions of the form $f * g$, with $f, g \in \mathrm{L}^1(G) \cap \mathrm{L}^2(G)$.

It is immediately verified that $A(G)$ is an ideal of $\mathrm{L}^1(G)$ contained in $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$.

The first step is to show that there is a filter base $\mathcal{B}$ defined on $A(G) \cap \mathcal{K}_{\mathbb{C}}(G)$, where the functions in this filter base approximate the Dirac measure, so that

(1) $\delta_e = \lim_{\mathcal{B}} \varphi d\lambda$ for all φ in any subset in $\mathcal{B}$.

(2) $\lim_{\mathcal{B}} \mathcal{F}(\varphi) = 1$, and $\|\mathcal{F}(\varphi)\|_\infty \leq 1$, for all φ in any subset in $\mathcal{B}$ (recall from Proposition 10.18 that $\mathcal{F}(\varphi)$ is bounded).

(3) $\lim_{\mathcal{B}} \varphi * f = f$, for all $f \in \mathrm{L}^p(G)$, $p = 1, 2$.

Condition (3) implies that $A(G)$ is dense in both $\mathrm{L}^1(G)$ and $\mathrm{L}^2(G)$, and it can also be shown that $\mathcal{F}(A(G))$ is dense in $\mathcal{C}_0(\widehat{G})$.

The second step is to show that there is a Haar measure ν on $\widehat{G}$ such that

$$\int_{\widehat{G}} |\mathcal{F}(f)|^2 \, d\nu = \int_G |f|^2 \, d\lambda \tag{$*$}$$

for all $f \in A(G)$. This goes as follows.

It can be shown that for every $f \in A(G)$, there is a unique positive measure μ_f on $\widehat{G}$ such that

$$(g * f)(e) = \int_{\widehat{G}} \mathcal{F}(g) \, d\mu_f, \qquad \text{for all } g \in A(G).$$

Then for every $f \in A(G)$, define Ω_f as the open subset of $\widehat{G}$ given by

$$\Omega_f = \{\chi \in \widehat{G} \mid \mathcal{F}(f)(\chi) \neq 0\}.$$

It can be shown that the subsets Ω_f form an open cover of $\widehat{G}$ when f ranges over $A(G)$. For every Ω_f, let ν_f be the positive measure on Ω_f associated with the Radon functional given by $\Phi_{\nu_f} = 1/(\mathcal{F}(f)) \, \Phi_{\mu_f}$, where Φ_{μ_f} is the Radon functional associated with the measure μ_f. It can be shown that the local measures ν_f patch to a global Haar measure ν on $\widehat{G}$, and that $(*)$ is satisfied.

Definition 10.19. Let G be a locally compact abelian group. For every Haar measure λ on G, the Haar measure $\widehat{\lambda} = \nu$ given by the previous construction is called the *measure associated with* λ or the *dual measure*.

Theorem 10.27. *(Plancherel) Let G be a locally compact abelian group equipped with a Haar measure λ. There is a Haar measure $\widehat{\lambda}$ on the dual $\widehat{G}$ such that for any $f \in \mathrm{L}^1(G) \cap \mathrm{L}^2(G)$, we have $\mathcal{F}(f) \in \mathrm{L}^2(\widehat{G})$. Furthermore, the map $f \mapsto \mathcal{F}(f)$ from $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$ to $\mathrm{L}^2(\widehat{G})$ has a unique extension which is an isometry from $\mathrm{L}^2(G)$ to $\mathrm{L}^2(\widehat{G})$.*

Proof sketch. By Property (3) of the filter base, $A(G)$ is dense in $\mathrm{L}^2(G)$. By $(*)$, the Fourier transform is an isometry from $A(G) \subseteq \mathrm{L}^2(G)$ to a subspace of $\mathrm{L}^2(\widehat{G})$, which is complete Therefore, by Proposition A.61, it has a unique extension Φ to $\mathrm{L}^2(G)$ (an isometry is uniformly continuous). To finish the proof, is suffices to show that $\Phi(L^2(G)) = \mathrm{L}^2(\widehat{G})$ and that $\Phi(f) = \mathcal{F}(f)$ for all $f \in \mathrm{L}^1(G) \cap \mathrm{L}^2(G)$.

Since $\mathcal{F}$ is an isometry between $A(G) \subseteq \mathrm{L}^2(G)$ and the subspace $\mathcal{F}(A(G))$ of $\mathrm{L}^2(\widehat{G})$, a sequence $(\mathcal{F}(f_n))_n$ of functions in $\mathcal{F}(A(G))$ is a Cauchy sequence iff $(f_n)_n$ is a Cauchy

sequence in $A(G)$, and since $\mathrm{L}^2(G)$ and $\mathrm{L}^2(\widehat{G})$ are complete and Φ is continuous, the sequence $(\mathcal{F}(f_n))_n$ converges to a function $g \in \mathrm{L}^2(\widehat{G})$ iff the sequence $(f_n)_n$ converges to a function $f \in \mathrm{L}^2(G)$ such that $\Phi(f) = g$. Therefore, to prove that $\Phi(L^2(G)) = \mathrm{L}^2(\widehat{G})$, it suffices to show that $\mathcal{F}(A(G))$ is dense in $\mathrm{L}^2(\widehat{G})$.

Assume that some $h \in \mathrm{L}^2(\widehat{G})$ is orthogonal to $\mathcal{F}(A(G))$. For all $f, g \in A(G)$, we have $\mathcal{F}(f)\mathcal{F}(g) = \mathcal{F}(f * g) \in \mathcal{F}(A(G))$. Since $\check{f}(a) = f(a^{-1})$ and $f^*(a) = \overline{f(a^{-1})} = \overline{f}(a^{-1})$, we have $f^* = \check{\overline{f}}$, so by Proposition 10.19(1), the equation

$$\mathcal{F}(\check{f})(\chi) = \overline{\mathcal{F}(\overline{f})(\chi)}$$

implies that $\mathcal{F}(f^*)(\chi) = \overline{\mathcal{F}(f)(\chi)}$. Then

$$\begin{aligned}
\langle h\mathcal{F}(f), \mathcal{F}(g)\rangle &= \int h(\chi)\mathcal{F}(f)(\chi)\overline{\mathcal{F}(g)(\chi)}\, d\widehat{\lambda}(\chi) \\
&= \int h(\chi)\overline{\overline{\mathcal{F}(f)(\chi)}\,\mathcal{F}(g)(\chi)}\, d\widehat{\lambda}(\chi) \\
&= \int h(\chi)\overline{\mathcal{F}(f^*)(\chi)\,\mathcal{F}(g)(\chi)}\, d\widehat{\lambda}(\chi) \\
&= \int h(\chi)\overline{\mathcal{F}(f^* * g)(\chi)}\, d\widehat{\lambda}(\chi) \\
&= \langle h, \mathcal{F}(f^* * g)\rangle = 0,
\end{aligned}$$

where Proposition 10.18 was used in the next to the last equation, and because $f^* * g \in A(G)$ and h is orthogonal to $\mathcal{F}(A(G))$. The above shows that $h\mathcal{F}(f)$ is orthogonal to $\mathcal{F}(A(G))$. Then it can be shown that this implies that $h = 0$. By a well known fact of Hilbert space theory, $\mathcal{F}(A(G))$ is dense in $\mathrm{L}^2(\widehat{G})$. The last step is to show that $\Phi(f) = \mathcal{F}(f)$ for all $f \in \mathrm{L}^1(G) \cap \mathrm{L}^2(G)$. The details are a bit involved, so we refer the reader to Bourbaki [9] (Chapter 2, Section 1, No. 3, Theorem 1). Another proof of Plancherel's theorem is given in Folland [33] (Chapter 4, Section 2, Theorem 4.25). It follows similar lines but uses a class of functions $\mathcal{B}^1$ different from $A(G)$. $\square$

The unique extension of $\mathcal{F}$ is also denoted $\mathcal{F}$. By using the same techniques as above, it is easy to see that $\overline{\mathcal{F}}$ also has a unique extension to $\mathrm{L}^2(G)$, and that it is an isometry.

On should realize that Theorem 10.27 does not say that the Fourier transform $\mathcal{F}$ (or the Fourier cotransform $\overline{\mathcal{F}}$) is defined on $\mathrm{L}^2(G)$, because in general the integral will not converge for f outside of $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$. What is happening is more subtle. It is always possible by using a limit process to define the Fourier transform of any $f \in \mathrm{L}^2(G)$, and this extension of $\mathcal{F}$ to $\mathrm{L}^2(G)$ is an isometry. This is still quite remarkable because there is no such result for $\mathrm{L}^1(G)$. Since the extension of $\mathcal{F}$ to $\mathrm{L}^2(G)$ is an isometry, it has an inverse, but it is far from obvious that this inverse has any relation to the Fourier cotransform on $\mathrm{L}^2(\widehat{G})$. In fact it does, but this requires proving Gelfand's duality theorem, that G and its double dual $\widehat{\widehat{G}}$ are isomorphic.

As a corollary of Theorem 10.27 and Proposition 10.13, we can prove the fact announced just after Proposition 10.13.

Proposition 10.28. *If G is a compact abelian group endowed with a Haar measure λ normalized so that G has measure $\lambda(G) = 1$, then $\widehat{G}$ is a Hilbert basis for $\mathrm{L}^2(G)$ (it is orthonormal and dense in $\mathrm{L}^2(G)$).*

Proof. By a well known fact of Hilbert space theory, it suffices to show that there is no nonzero function $f \in \mathrm{L}^2(G)$ orthogonal to every character $\chi \in \widehat{G}$. Assume $f \in \mathrm{L}^2(G)$ is orthogonal to every character $\chi \in \widehat{G}$. This means that $\int f(a)\overline{\chi(a)}\, d\lambda(a) = 0$, and since

$$\int f(a)\overline{\chi(a)}\, d\lambda(a) = \mathcal{F}(f)(\chi),$$

we get

$$\mathcal{F}(f)(\chi) = 0, \qquad \text{for all } \chi \in \widehat{G}.$$

Since Plancherel's theorem asserts that $\mathcal{F}$ is an isometry, it is injective, so $f = 0$, establishing our result. $\square$

We now turn to the issue of Fourier inversion. If $f, g \in \mathrm{L}^2(G)$, the convolution $f * g$ is given by

$$(f * g)(s) = \int f(t)g(t^{-1}s)\, d\lambda(t)$$

so for $s = e$,

$$(f * g)(e) = \int f(t)g(t^{-1})\, d\lambda(s) = \int f(t)\check{g}(t)\, d\lambda(s) = \int f(t)\overline{\overline{\check{g}}(t)}\, d\lambda(s) = \int f(t)\overline{g^*(t)}\, d\lambda(s).$$

We also have

$$\langle f, g\rangle = \int f(t)\overline{g(t)}\, d\lambda(t),$$

so we deduce

$$(f * g^*)(e) = \langle f, g\rangle. \tag{$\dagger_1$}$$

But Plancherel's theorem implies that

$$\langle f, g\rangle = \langle \mathcal{F}(f), \mathcal{F}(g)\rangle = \int \mathcal{F}(f)(\chi)\overline{\mathcal{F}(g)(\chi)}\, d\widehat{\lambda}(\chi),$$

and we conclude that

$$(f * g^*)(e) = \int \mathcal{F}(f)(\chi)\overline{\mathcal{F}(g)(\chi)}\, d\widehat{\lambda}(\chi). \tag{$\dagger_2$}$$

Proposition 10.29. *(Fourier inversion for $A(G)$) If $f \in A(G)$, then $\mathcal{F}(f) \in \mathrm{L}^1(\widehat{G})$, and*

$$f(a) = \int_{\widehat{G}} \chi(a)\mathcal{F}(f)(\chi)\, d\widehat{\lambda}(\chi) \qquad \textit{for all } a \in G.$$

Equivalenty,

$$f = (\overline{\mathcal{F}} \circ \mathcal{F})(f) \circ \eta, \qquad \textit{for all } f \in A(G),$$

where $\eta\colon G \to \widehat{\widehat{G}}$ is the canonical map.

Sketch of proof. Using functions g that approximate the Dirac measure δ_e, Formula ($\dagger_2$) yields

$$f(e) = \int \mathcal{F}(f)(\chi)\, d\widehat{\lambda}(\chi),$$

which is the result of Proposition 10.29 for $a = e$ since $\chi(e) = 1$. For any arbitrary $a \in G$, replace f by $\lambda_{a^{-1}} f$ and use Proposition 10.19(3). See Bourbaki [9] (Chapter 2, Section 1, No. 4). $\square$

Since the ideal $A(G)$ of Definition 10.18 is contained in $\mathrm{L}^1(G) \cap \mathrm{L}^2(G)$, unlike the situation in Plancherel's theorem, there is no need to extend $\mathcal{F}$ (and $\overline{\mathcal{F}}$ on $\mathrm{L}^1(\widehat{G})$).

It is also shown in Bourbaki that the inversion formula holds for all $f \in \mathrm{L}^2(G)$ such that $\mathcal{F}(f) \in \mathrm{L}^1(\widehat{G})$.

In order to proceed any further, we need Pontrjagin's duality theorem asserting that η is an isomorphism.

10.9 Pontrjagin Duality and Fourier Inversion

The Pontrjagin duality theorem is one of the most important and most beautiful theorems of the theory of locally compact abelian groups. Recall that we have a canonical map $\eta\colon G \to \widehat{\widehat{G}}$ given by

$$\eta_a(\chi) = \chi(a), \qquad a \in G,\ \chi \in \widehat{G},$$

which is a homomorphism.

Theorem 10.30. *(Pontrjagin duality theorem) Let G be a locally compact abelian group endowed with a Haar measure λ, let $\widehat{G}$ be its dual group endowed with the associated Haar measure $\widehat{\lambda}$ (see Definition 10.19), and let $\widehat{\widehat{G}}$ be its double dual endowed with the associated measure $\widehat{\widehat{\lambda}}$. The map $\eta\colon G \to \widehat{\widehat{G}}$ is an isomorphism and a homeomorphism between the topological groups G and $\widehat{\widehat{G}}$ that maps the measure λ to the measure $\widehat{\widehat{\lambda}}$, which means that $\lambda = \eta^{-1}(\widehat{\widehat{\lambda}})$, as in Definition 8.14. If we identify G and $\widehat{\widehat{G}}$ using the isomorphism η, then the extension $\mathcal{F}\colon \mathrm{L}^2(G) \to \mathrm{L}^2(\widehat{G})$ of the Fourier transform to $\mathrm{L}^2(G)$ and the extension*

$\overline{\mathcal{F}}\colon \mathrm{L}^2(\widehat{G}) \to \mathrm{L}^2(G)$ *of the Fourier cotransform to* $\mathrm{L}^2(\widehat{G})$ *are mutual inverses. In particular, Fourier inversion holds; that is,*

$$f = (\overline{\mathcal{F}} \circ \mathcal{F})(f) \circ \eta, \qquad \textit{for all } f \in \mathrm{L}^2(G).$$

Proof idea. The proof of Theorem 10.30 is too technical to be presented in full detail here. A proof can found in Bourbaki [9] (Chapter 2, Section 1, No. 5, Theorem 2) and in Folland [33] (Chapter 4, Section 4.3, Theorem 4.31).

The first part of the proof of Pontrjagin duality establishes the fact that η is injective and a homeomorphism onto its image, which is closed in $\widehat{\widehat{G}}$. To prove that η is injective and that η^{-1} is continuous, it suffices to prove that for every neighborhood U of e in G there is some neighborhood W of $\widehat{\widehat{e}}$ in $\widehat{\widehat{G}}$, such that $\eta^{-1}(W) \subseteq U$.

We can find a compact symmetric neighborhood V of e in G such that $V^2 \subseteq U$, and some positive function $f \in \mathcal{K}_{\mathbb{C}}(G)$ whose support is contained in V. If we let $g = f^* * f$, then we see that $g \in A(G)$, $\mathrm{supp}(g) \subseteq U$, and $g(e) > 0$, which follows from Equation ($\dagger_1$) of the previous section. Since $\widehat{\widehat{G}}$ has the compact open topology, which is equivalent to the topology of pointwise convergence on $\mathrm{L}^1(\widehat{G})$ (by Theorem 10.6 applied to $\widehat{G}$), there is a neighborhood W of the identity $\widehat{\widehat{e}}$ in $\widehat{\widehat{G}}$ such that

$$|\overline{\mathcal{F}}(\mathcal{F}(g))(\zeta) - \overline{\mathcal{F}}(\mathcal{F}(g))(\widehat{\widehat{e}})| < \frac{1}{2} g(e), \qquad \zeta \in W,$$

so if $a \in \eta^{-1}(W)$, since $g \in A(G)$, by Proposition 10.29 we have $g = (\overline{\mathcal{F}} \circ \mathcal{F})(g) \circ \eta$, and we obtain

$$|g(a) - g(e)| < \frac{1}{2} g(e).$$

Therefore, $g(a) \neq 0$, and since $\mathrm{supp}(g) \subseteq U$, we have $a \in U$, which shows that $\eta^{-1}(W) \subseteq U$, as desired.

The second part is to prove that η is surjective. Bourbaki's proof uses the following fact which shows the existence of certain kinds of bump functions on $\mathrm{L}^1(G)$.

Proposition 10.31. *Given any closed subset P of $\widehat{G}$ and any $\chi \in \widehat{G}$, if $\chi \notin P$, then there exists some $f \in \mathrm{L}^1(G)$ such that $\mathcal{F}(f)(\chi) = 1$ and $\mathcal{F}(f)$ vanishes on P.*

Assume that there is some $\zeta \in \widehat{\widehat{G}}$ such that $\zeta \notin \eta(G)$. By Proposition 10.31 applied to $\widehat{G}$, since we proved in the first part of the proof that $\eta(G)$ is closed, there is some nonzero function $f \in \mathrm{L}^1(\widehat{G})$ such that $\mathcal{F}(f)$ vanishes on $\eta(G)$, that is (since $\eta_a(\chi) = \chi(a)$),

$$\mathcal{F}(f)(\eta_a) = \int f(\chi)\overline{\eta_a(\chi)}\, d\widehat{\lambda}(\chi) = \int f(\chi)\overline{\chi(a)}\, d\widehat{\lambda}(\chi) = 0, \qquad \text{for all } a \in G.$$

Using Fubini's theorem, for any $g \in \mathrm{L}^1(G)$, we have

$$\begin{aligned}\int f(\chi)\mathcal{F}(g)(\chi)\, d\widehat{\lambda}(\chi) &= \int f(\chi)\left(\int g(a)\overline{\chi(a)}\, d\lambda(a)\right) d\widehat{\lambda}(\chi)\\ &= \int \left(\int f(\chi)\overline{\chi(a)}\, d\widehat{\lambda}(\chi)\right) g(a)\, d\lambda(a) = 0.\end{aligned}$$

By the second remark just after Proposition 10.18, $\mathcal{F}(\mathrm{L}^1(G))$ is dense in $\mathcal{C}_0(\widehat{G})$, so

$$\int f(\chi)\mathcal{F}(g)(\chi)\, d\widehat{\lambda}(\chi) = 0$$

for all $g \in \mathrm{L}^1(G)$ implies that $f \equiv 0$, a contradiction. Therefore, η is surjective, thus an isomorphism. Then using Plancherel's Theorem (Theorem 10.27) and Proposition 10.29, we can shown that $\overline{\mathcal{F}}$ (defined on $\mathrm{L}^2(\widehat{G})$) is an isometry between $\mathrm{L}^2(\widehat{G})$ and $\mathrm{L}^2(\widehat{\widehat{G}})$, and that $\lambda = \eta^{-1}(\widehat{\widehat{\lambda}})$. ∎

From now on we identify G and $\widehat{\widehat{G}}$ unless specified otherwise.

We will now show that there is another class of functions for which $\mathcal{F}$ and $\overline{\mathcal{F}}$ are mutual inverses. For this we need the following result.

Proposition 10.32. *For every $f \in \mathrm{L}^1(G)$ and every $H \in \mathrm{L}^1(\widehat{G})$, we have*

$$\int_G f(a)\mathcal{F}(H)(a)\, d\lambda(a) = \int_{\widehat{G}} \mathcal{F}(f)(\chi)H(\chi)\, d\widehat{\lambda}(\chi).$$

The proof of Proposition 10.32 is an application of Fubini's theorem.

Definition 10.20. Let G be a locally compact abelian group. We define $B(G)$ as the set of functions

$$B(G) = \{f \in \mathrm{L}^1(G) \mid \mathcal{F}(f) \in \mathrm{L}^1(\widehat{G})\}.$$

Recall that we are identifying G and $\widehat{\widehat{G}}$ so that we view $\mathcal{F}$ and $\overline{\mathcal{F}}$ as mutual inverses. Then $B(\widehat{G})$ is defined by

$$B(\widehat{G}) = \{f \in \mathrm{L}^1(\widehat{G}) \mid \overline{\mathcal{F}}(f) \in \mathrm{L}^1(G)\}.$$

Theorem 10.33. *The restriction of $\mathcal{F}$ to $B(G)$ is a bijection from $B(G)$ to $B(\widehat{G})$, whose inverse is the restriction of $\overline{\mathcal{F}}$ to $B(\widehat{G})$.*

Proof. If $f \in B(G)$, then $\mathcal{F}(f) \in \mathrm{L}^1(\widehat{G}) \cap \mathcal{C}_0(\widehat{G}) \subseteq \mathrm{L}^1(\widehat{G}) \cap \mathrm{L}^2(\widehat{G})$. Let $h = (\overline{\mathcal{F}} \circ \mathcal{F})(f) \in \mathrm{L}^2(G)$. For every $g \in \mathcal{K}_{\mathbb{C}}(\widehat{G})$, by Theorem 10.30, $\overline{\mathcal{F}}$ is an isometry, $\mathcal{F}$ and $\overline{\mathcal{F}}$ are mutual inverses, so that $\mathcal{F}(h) = \mathcal{F}(f)$, and by Proposition 10.32, we have

$$\begin{aligned}
\int_G h(a)\mathcal{F}(g)(a)\,d\lambda &= \langle \mathcal{F}(g), \overline{h}\rangle \\
&= \langle g, \overline{\mathcal{F}}(\overline{h})\rangle && \mathcal{F} \text{ is an isometry} \\
&= \langle g, \overline{\mathcal{F}(h)}\rangle && \text{by Proposition 10.16} \\
&= \int_{\widehat{G}} \mathcal{F}(h)(\chi)g(\chi)\,d\widehat{\lambda}(\chi) \\
&= \int_{\widehat{G}} \mathcal{F}(f)(\chi)g(\chi)\,d\widehat{\lambda}(\chi) && \mathcal{F}(h) = \mathcal{F}(f) \\
&= \int_G f(a)\mathcal{F}(g)(a)\,d\lambda(a) && \text{by Proposition 10.32.}
\end{aligned}$$

It follows that

$$\int_G h(a)\mathcal{F}(g)(a)\,d\lambda = \int_G f(a)\mathcal{F}(g)(a)\,d\lambda(a) \qquad \text{for all } g \in \mathcal{K}_{\mathbb{C}}(\widehat{G}),$$

and we deduce that $h = f \in \mathrm{L}^1(G)$. Since $h = (\overline{\mathcal{F}} \circ \mathcal{F})(f)$, we conclude that $\mathcal{F}(f) \in B(\widehat{G})$. We also proved that $\overline{\mathcal{F}} \circ \mathcal{F}$ is the identity on $B(G)$. By exchanging the roles of G and $\widehat{G}$, we can show that $\mathcal{F} \circ \overline{\mathcal{F}}$ is the identity on $B(\widehat{G})$. The theorem follows immediately. $\square$

Note that unlike the situation in Theorem 10.30, in Theorem 10.33 there is no need to extend $\mathcal{F}$ (and $\overline{\mathcal{F}}$ on $B(\widehat{G})$).

Remark: The space $B(G)$ is an algebra for the pointwise product and the convolution product. It is also easy to see that $\mathcal{F}(fg) = \mathcal{F}(f) * \mathcal{F}(g)$, in addition to $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$.

The property $\mathcal{F}(fg) = \mathcal{F}(f) * \mathcal{F}(g)$ is also satisfied by $\mathrm{L}^2(G)$.

Proposition 10.34. *For any two functions $f, g \in \mathrm{L}^2(G)$, we have*

$$\mathcal{F}(fg) = \mathcal{F}(f) * \mathcal{F}(g).$$

Proposition 10.34 is proven in Bourbaki [9] (Chapter 2, Section 1, No. 6, Theorem 3), and Folland [33] (Chapter 4, Section 4.3, Proposition 4.36).

The following proposition explains a phenomenon that we have already observed for $G = \mathbb{T}$ and $G = \mathbb{Z}$.

Proposition 10.35. *For any locally compact abelian group G, the group G is discrete if and only if $\widehat{G}$ is compact (and by duality, G is compact if and only if $\widehat{G}$ is discrete). Furthermore, if G is compact and endowed with the Haar measure λ normalized so that $\lambda(G) = 1$, then the associated measure $\widehat{\lambda}$ on $\widehat{G}$ is the counting measure. If G is discrete and endowed with the counting measure, then the associated measure $\widehat{\lambda}$ on $\widehat{G}$ is normalized so that $\widehat{\lambda}(\widehat{G}) = 1$.*

Proof. Assume that G is discrete. If so $\mathrm{L}^1(G)$ is unital with identity δ_e. By Theorem 9.19, the algebra $\mathsf{X}(\mathrm{L}^1(G))$ is compact, and since $\widehat{G}$ is homeomorphic to $\mathsf{X}(\mathrm{L}^1(G))$, we deduce that $\widehat{G}$ is compact.

Assume now that $\widehat{G}$ is compact. Since the characters are uniformly continuous, there is an open subset V containing the identity in $\widehat{G}$ such that for all $\chi \in V$, for all $a \in G$, we have $|\chi(a) - 1| \leq 1$. Since $a \in G$ is arbitrary, we can replace it by a^n for any $n \in \mathbb{Z}$, and since $\chi(a^n) = \chi(a)^n$, we get

$$|\chi(a)^n - 1| \leq 1, \qquad \text{for all } a \in G \text{ and all } n \in \mathbb{Z}.$$

Since $\chi(a)$ is a complex number of unit length, say $\chi(a) = \cos\theta + i\sin\theta$, with $0 \leq \theta < 2\pi$, we have $\chi(a)^n = \cos n\theta + i \sin n\theta$, and

$$|\cos n\theta - 1 + i\sin n\theta|^2 = (\cos n\theta - 1)^2 + \sin^2 n\theta = \cos^2 n\theta - 2\cos n\theta + 1 + \sin^2 n\theta = 2(1 - \cos n\theta).$$

Unless $\theta = 0$, we can find some n so that $\cos n\theta < 0$, and we get a contradiction to the inequality $|\chi(a)^n - 1| \leq 1$. Therefore, $\chi(a) = 1$ for all $a \in G$, which implies that $V = \{e\}$ (since the other characters are not the constant character 1). Thus we proved that $\{e\}$ is an open subset of $\widehat{G}$, so every singleton subset $\{\chi\}$ is open, which means that $\widehat{G}$ is discrete. The second part of the proposition is proven in Folland [33] (Chapter 4, Section 2, Proposition 4.24). $\square$

We haven't discussed functions of positive type yet. They play an important role in the theory of unitary representations of a locally compact group. A function $\varphi \in \mathrm{L}^\infty(G)$ is of *positive type* if

$$\int (f^* * f)\varphi \, d\lambda \geq 0, \qquad \text{for all } f \in \mathrm{L}^1(G).$$

Let $\mathcal{P}_+(G)$ be the space of functions of positive type. There is a connection with the dual group $\widehat{G}$. Indeed, for any measure $\mu \in \mathcal{M}(\widehat{G})$, define φ_μ by

$$\varphi_\mu(a) = \int_{\widehat{G}} \chi(a) d\mu(\chi).$$

Then a theorem of Bochner states that for any function of positive type $\varphi \in \mathcal{P}_+(G)$, there is a *unique positive measure* $\mu \in \mathcal{M}(\widehat{G})$ such that $\varphi = \varphi_\mu$. We will return to positive functions in Chapter 12 (Section 12.5) and Chapter 17, and refer the reader to Folland for a discussion of this topic; see [33] Chapters 3 and Chapter 4, Theorem 4.18.

Chapter 11

Representations of Algebras and Hilbert Algebras

In order to generalize harmonic analysis to compacts groups, we need to introduce group representations. However, it turns out that in order to prove the main theorem of the subject, the Peter–Weyl theorem, one needs the notion of representation of algebras, because there is a bijection between the set of unitary representations of a locally compact group G and the set of nondegenerate representations of the involutive algebra $\mathrm{L}^1(G)$. When G is compact, $\mathrm{L}^2(G)$ is actually a Hilbert algebra, and there is a beautiful structure theorem for Hilbert algebras which says that such an algebra splits as a Hilbert sum of minimal left ideals, and this result can be used to prove the Peter–Weyl theorem.

The purpose of this chapter is to define the notion of Hilbert algebra and to develop the machinery needed to prove three fundamental theorems (Theorem 11.31, Theorem 11.32, and Theorem 11.34) about complete Hilbert algebras. We also state two important theorems about commutative Hilbert algebras; the Plancherel–Godement theorem and the Bochner-Godement theorem. These theorems will be needed later when we discuss Gelfand pairs. We mostly follow Dieudonné [24] (Chapter XV, Sections 15.5-15.9), occasionally borrowing from Folland [33] (Chapters 1 and 3). This is a rather technical chapter and we do not give all proofs, relying on the above references for details.

We begin with the definition of the notion of representation of an involutive algebra in a Hilbert space. We define the notion of Hilbert sum of a finite or an infinite family of Hilbert spaces. Then we introduce the crucial concepts of invariant subspace, of a topologically irreducible representation, of a nondegenerate representation, of a totalizing (or cyclic) vector, and of a topologically cyclic representation.

In Section 11.3 we define positive linear foms and positive Hilbert forms on an involutive algebra A. Positive Hilbert forms are positive hermitian forms which may fail to be positive definite but satisfy a kind of adjunction property, namely a positive hermitian form g satisfying the condition

$$g(xy, z) = g(y, x^* z) \qquad \text{for all } x, y, z \in A.$$

They can be used to define topologically cyclic representations. A good method for producing positive Hilbert forms is to use positive linear forms which satisfy the property $f(s^*s) \geq 0$, for all $s \in A$, a kind of positive semidefinite property. Then the map g given by $g(x, y) = f(y^*x)$ is a positive Hilbert form.

In Section 11.4 we introduce bitraces and Hilbert algebras. A bitrace is a positive Hilbert form $g\colon A \times A \to \mathbb{C}$ such that

$$g(t^*, s^*) = g(s, t), \qquad \text{for all } s, t \in A.$$

The most important concept of this section is the notion of Hilbert algebra. An involutive algebra A is a *Hilbert algebra* if its underlying vector space is a hermitian space whose hermitian inner product $\langle -, - \rangle$ is a bitrace satisfying two extra conditions (U) and (N) (see Definition 11.14). Specifically, the conditions for being a bitrace hold

$$\langle y^*, x^* \rangle = \langle x, y \rangle \tag{1}$$

$$\langle xy, z \rangle = \langle y, x^*z \rangle, \tag{2}$$

and the following two conditions hold: for every $x \in A$, there is some $M_x \geq 0$ such that

$$\langle xy, xy \rangle \leq M_x \langle y, y \rangle, \quad \text{for all } y \in A, \tag{U}$$

and

$$\text{the subspace spanned by the set } \{xy \mid x, y \in A\} \text{ is dense in } A. \tag{N}$$

In general the map $(x, y) \mapsto xy$ is not continuous, so a Hilbert algebra is not a normed algebra in the sense of Definition 9.4. However, if the Hilbert algebra A is complete, then it can be shown that the map $(x, y) \mapsto xy$ is continuous, and thus that A is a normable algebra which is a Banach space (see Proposition 11.15).

An important example of a complete Hilbert algebra is the algebra $\mathcal{L}_2(H)$ of Hilbert–Schmidt operators on a separable Hilbert space H. Another very important example of a complete Hilbert algebra is $\mathrm{L}^2(G)$, where G is a compact, metrizable group.

Section 11.5 which is about complete separable Hilbert algebras contains the most important results of this chapter regarding the structure of such algebras. To understand the structure of complete separable Hilbert algebras we need to study minimal left ideals and the irreducible self-adjoint idempotents which generate them. Recall that an element e of an algebra A is *idempotent* if $e^2 = e$, and *self-adjoint* if $e = e^*$.

Roughly speaking, the master decomposition theorem (Theorem 11.31) states that given a complete separable Hilbert algebra A, there is an irredundant list $(\mathfrak{l}_k)_{k \in J}$ of the minimal left ideals of A, and A is the Hilbert sum of two-sided ideals $\mathfrak{a}_k$,

$$A = \bigoplus_{k \in J} \mathfrak{a}_k,$$

where each $\mathfrak{a}_k$ is the Hilbert sum obtained by picking a certain number of copies of the minimal left ideal $\mathfrak{l}_k$ of A,

$$\mathfrak{a}_k = \bigoplus_{j \in I_k} \mathfrak{l}'_j,$$

with $\mathfrak{l}'_j$ isomorphic to $\mathfrak{l}_k$.

Each two-sided ideal $\mathfrak{a}_k$ contains no closed two-sided ideal other than (0) and $\mathfrak{a}_k$. They are said to be *topologically simple*.

Theorem 11.32 gives the structure of a topologically simple Hilbert algebra. Theorem 11.32 implies that in the Hilbert sum

$$A = \bigoplus_{k \in J} \mathfrak{a}_k$$

given by Theorem 11.31, the Hilbert algebra $\mathfrak{a}_k$, which is a building block of the decomposition, is either isomorphic to the algebra $\mathcal{L}_2(\mathfrak{l}_k)$ of Hilbert–Schmidt operators on $\mathfrak{l}_k$, or to the finite-dimensional algebra $\mathrm{End}_{\mathbb{C}}(\mathfrak{l}_k)$ of all endomorphisms of the vector space $\mathfrak{l}_k$. If G is a metrizable compact group and $A = \mathrm{L}^2(G)$, then every $\mathfrak{a}_k$ in the Hilbert sum for A is isomorphic to the finite-dimensional algebra $\mathrm{End}_{\mathbb{C}}(\mathfrak{l}_k)$.

The master decomposition for a nondegenerate continuous representation $V\colon A \to \mathcal{L}(H)$ (Theorem 11.34) states that the Hilbert space H is a Hilbert sum $H = \bigoplus_{k \in J} H_k$ of subspaces invariant under V, and that the restriction V_k of V to $\mathfrak{a}_k$ can be considered as a representation of $\mathfrak{a}_k$ in H_k. Furthermore, each representation V_k is the Hilbert sum of irreducible representations, each equivalent to a representation $U_{\mathfrak{l}_k}$ canonically associated with a minimal ideal $\mathfrak{l}_k$ of $\mathfrak{a}_k$.

Section 11.8 discusses the Plancherel–Godement theorem and the Bochner–Godement theorem without proofs. These theorems apply to a *commutative* Hilbert algebra (not necessarily complete) arising from the quotient of a commutative Hilbert algebra by a left ideal induced by a bitrace satisfying two additional conditions. Therefore we go back to positive Hilbert forms to describe the construction of a certain representation.

The idea is that if g is a positive Hilbert form on an involutive (not necessarily commutative) algebra A, it almost defines an inner product, but in general it fails to be positive definite because there may be nonzero elements $s \in A$ such that $g(s, s) = 0$. However, if we take the quotient of A by the set $\mathfrak{n} = \{s \in A \mid g(s, s) = 0\}$, which is a left ideal because g is a positive Hilbert form, then we can define an inner product on the quotient vector space $A/\mathfrak{n}$. If g is a bitrace, then $A/\mathfrak{n}$ is an involutive algebra (see Proposition 11.35).

If a positive Hilbert form g satisfies the analog of Condition (U) of Definition 11.14, namely, for every $s \in A$, there is some $M_s \geq 0$ such that

$$g(st, st) \leq M_s g(t, t), \quad \text{for all } t \in A, \tag{U}$$

and if the hermitian space $A/\mathfrak{n}_g$ is separable, where $\mathfrak{n}_g = \{s \in A \mid g(s,s) = 0\}$, then g defines a unitary representation $U_g \colon A \to \mathcal{L}(H_g)$, where H_g is the Hilbert space which is the completion of $A/\mathfrak{n}_g$ (see Proposition 11.37).

In general, the representation $U_g \colon A \to \mathcal{L}(H_g)$ given by Proposition 11.37 may be degenerate. It is nondegenerate if and only if the following condition holds:

$$\text{the subspace spanned by the set } \{\pi_g(st) \mid s,t \in A\} \text{ is dense in } A/\mathfrak{n}_g. \tag{N}$$

If A is a commutative Hilbert algebra and if Property (U) holds, then the representation $U_g \colon A \to \mathcal{L}(H_g)$ is nondegenerate and the image of A under U_g is a commutative subalgebra of the C^*-algebra $\mathcal{L}(H_g)$. Let $\mathcal{A}_g$ be the closure of $U_g(A)$ in $\mathcal{L}(H_g)$, so that $\mathcal{A}_g$ is a commutative C^*-algebra.

Roughly speaking, the Plancherel–Godement theorem (Theorem 11.41) states that if A is a commutative involutive algebra, if g is a bitrace on A satisfying Conditions (U) and (N), and if the hermitian space $A/\mathfrak{n}_g$ and the C*-algebra $\mathcal{A}_g \subseteq \mathcal{L}(H_g)$ are separable, then g is obtained from a positive measure by a process of integration from hermitian characters.

Sections 11.9, 11.10 and 11.11 present results generally called spectral theorems. The important notion of projection-valued measure is introduced.

Section 11.9 contains a technically crucial characterization of a representation of the algebra $\mathcal{C}_{\mathbb{C}}(K)$ of continuous functions on a compact metrizable space K, a result proven using the theorems from Section 11.8, the Plancherel–Godement theorem and the Bochner–Godement theorem. This theorem shows that every topologically cyclic representation $U \colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(H)$ of the commutative unital C^*-algebra $\mathcal{C}_{\mathbb{C}}(K)$ (for K compact) in a separable Hilbert space H is equivalent to a representation $M_\mu \colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(\mathrm{L}^2_\mu(K;\mathbb{C}))$ such that for every $u \in \mathcal{C}_{\mathbb{C}}(K)$, $M_\mu(u) \colon \mathrm{L}^2_\mu(K;\mathbb{C}) \to \mathrm{L}^2_\mu(K;\mathbb{C})$ is the continuous linear map multiplication by u (μ is some positive Radon measure on K); see Theorem 11.43.

A particularly interesting case for the space K arises if we consider a commutative unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(H)$. In this case, by the Gelfand–Naimark theorem (Theorem 9.37), the Gelfand transform $\mathcal{G} \colon \mathcal{A} \to \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$ is an isometric isomorphism between $\mathcal{A}$ and $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$. Furthermore, $K = \mathsf{X}(\mathcal{A})$ is compact (see Theorem 9.19). Thus the inverse Gelfand transform $\mathcal{G}^{-1} \colon \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A})) \to \mathcal{A}$ is a representation of $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$ and Theorem 11.43 can be used to prove Theorem 11.45 (Spectral Theorem I), which can be viewed as a generalization of the spectral theorem for normal linear maps on a finite-dimensional hermitian space. This result applies to any commutative unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(H)$. An interesting special case is the subalgebra $\mathcal{A}_T$ of $\mathcal{L}(H)$ generated by T, T^* and I, where T is a normal continuous linear map T on a Hilbert space H. We have Theorem 11.46, a first version of a spectral theorem for normal continuous linear maps. The end of this section presents a condition for a scalar in the spectrum $\sigma(T)$ of T to be an eigenvalue.

The next step taken in Section 11.10 is to realize that a representation $U \colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(H)$ as above determines certain complex Radon measures $\mu_{u,v}$ on K, and that conversely these

measures determine U. Indeed for any two vectors $u, v \in H$ there is a unique complex Radon measure $\mu_{u,v}$ on K such that

$$\langle U(f)(u), v\rangle = \int_K f\, d\mu_{u,v}, \quad f \in \mathcal{C}_{\mathbb{C}}(K).$$

The measure $\mu_{u,v}$ is often called a *spectral measure*.

Then it is possible to extend the representation U of $\mathcal{C}_{\mathbb{C}}(K)$ to the larger commutative unital C^*-algebra $B(K)$ of bounded Borel measurable functions on K. We obtain the representation $\widetilde{U}\colon B(K) \to \mathcal{L}(H)$ which is completely determined by the equation

$$\langle \widetilde{U}(f)(u), v\rangle = \int_K f\, d\mu_{u,v} \quad \text{for all } u, v \in H \text{ and for all } f \in B(K). \tag{$*_3$}$$

The above equation defines a "weak integral" with respect to the family of measures $\mu_{u,v}$ denoted

$$\widetilde{U}(f) = \int f\, d\mu.$$

For simplicity of notation we denote $\widetilde{U}$ as U. Since for any commutative unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(H)$ the inverse Gelfand transform $\mathcal{G}^{-1}\colon \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A})) \to \mathcal{A}$ is a representation of $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$, we obtain a representation $U\colon B(\mathsf{X}(\mathcal{A})) \to \mathcal{L}(H)$ of $\mathcal{G}^{-1}\colon \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A})) \to \mathcal{A}$. Consequently we obtain Theorem 11.51 which states that there is a family of complex Radon measures $(\mu_{u,v})_{(u,v)\in H\times H}$ on $\mathsf{X}(\mathcal{A})$ and we have

$$T = \int \mathcal{G}_T\, d\mu, \quad U(f) = \int f\, d\mu$$

for all $T \in \mathcal{A}$ and all $f \in B(\mathsf{X}(\mathcal{A}))$. This is another spectral theorem for a commutative unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(H)$.

What we gain in doing all this is the fact that we can apply the extended representation U to the characteristic functions χ_E of Borel sets E (on K) (the functions χ_E are not continuous), and such operators $P(E) = U(\chi_E)$ turn out to be orthogonal projections in $\mathcal{L}(H)$. These families of projections have properties that make them *projection-valued measures* (also called *spectral measures*), and such measures can be used to define representations of $B(K)$ that generalize the notion of integral.

Projection-valued measures are defined and used to prove more spectral theorems in Section 11.11. The connection between a family P of projection-valued measures and families of complex Radon measures as above is that if for all $u, v \in E$ we define $P_{u,v}$ by

$$P_{u,v}(E) = \langle P(E)(u), v\rangle,$$

then the $P_{u,v}$ are complex Radon measures (with $P_{u,u}$ a positive measure) with the same properties as the $\mu_{u,v}$. This allows to define a notion of weak integral with respect to a

projection-valued measure. We obtain the important Theorem 11.54 which states that for any function $f \in B(K)$, the integral

$$U(f) = \int f\, dP$$

is defined by the equation

$$\langle U(f)(u), v\rangle = \int f\, dP_{u,v} \quad \text{for all } u, v \in H \text{ and all } f \in B(K).$$

Furthermore the map $U\colon B(K) \to \mathcal{L}(H)$ is a representation.

Theorem 11.54 yields more spectral theorems in terms of projection-valued mesures, in particular another spectral theorem (Spectral Theorem II) for any commutative unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(H)$ (see Theorem 11.55).

Remarkably Theorem 11.55 (Spectral Theorem III) can be generalized to unital commutative Banach algebras. Theorem 11.57 states that for any commutative unital involutive Banach algebra $\mathcal{A}$, for any representation $U\colon \mathcal{A} \to \mathcal{L}(H)$ of $\mathcal{A}$ in a Hilbert space H, there is a regular projection-valued measure P on $\mathsf{X}(\mathcal{A})$ such that

$$U(a) = \int \mathcal{G}_a\, dP, \quad a \in \mathcal{A},$$

where $\mathcal{G}_a$ is the Gelfand tranform. In fact the projection-valued measure P is unique.

There is one more generalization (Spectral Theorem IV) where the involutive Banach algebra $\mathcal{A}$ is not necessarily unital, but the representation $U\colon \mathcal{A} \to \mathcal{L}(H)$ is nondegenerate; see Theorem 11.58. This theorem is crucial to the proof of Theorem 12.17 characterizing the unitary representations of an *abelian* locally compact group. Intuitively, the characters of G are glued by a suitable projection-valued measure. In turn Theorem 12.17 is a key result used in Mackey's theory for constructing induced representations; see Chapter 16, Proposition 16.1.

As corollary of Theorem 11.58 we also obtain a spectral theorem for nondegenerate representations $U\colon \mathcal{C}_0(X;\mathbb{C}) \to \mathcal{L}(H)$ of $\mathcal{C}_0(X;\mathbb{C})$; see Theorem 11.59. This theorem is used in Section 16.2 to give an alternate definition of a system of imprimitivity; see Definition 16.4.

11.1 Representations of Algebras with Involution

Let A be an algebra with an involution (not necessarily a normed algebra, nor a commutative or a unital algebra). Since representations of algebras involve Hilbert spaces, the reader may want to review Chapter D, especially Sections D.1 and D.2. For the reader's convenience we quickly review some basic notions, including Hilbert bases.

If H is a complex vector space, recall that a map $\langle -, - \rangle \colon H \times H \to \mathbb{C}$ is a *hermitian form* if it satisfies the following properties for all $x, y, x_1, x_2, y_1, y_2 \in H$ and all $\lambda \in \mathbb{C}$: it is *sesquilinear*, which means that

$$\begin{aligned}\langle x_1 + x_2, y\rangle &= \langle x_1, y\rangle + \langle x_2, y\rangle \\ \langle x, y_1 + y_2\rangle &= \langle x, y_1\rangle + \langle x, y_2\rangle \\ \langle \lambda x, y\rangle &= \lambda \langle x, y\rangle \\ \langle x, \lambda y\rangle &= \overline{\lambda}\langle x, y\rangle,\end{aligned}$$

and satisfies the *hermitian property*,

$$\langle y, x\rangle = \overline{\langle x, y\rangle}.$$

The hermitian property implies that $\langle x, x\rangle \in \mathbb{R}$ for all $x \in H$.

A hermitian form $\langle -, - \rangle \colon H \times H \to \mathbb{C}$ is *positive* if

$$\langle x, x\rangle \geq 0 \qquad \text{for all } x \in H.$$

A positive hermitian form satisfies the *Cauchy-Schwarz inequality*:

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle, \qquad \text{for all } x, y \in H.$$

A positive hermitian form is *positive definite* if for all $x \in H$,

$$\langle x, x\rangle = 0 \quad \text{implies that } x = 0,$$

or equivalently,

$$\langle x, x\rangle > 0 \qquad \text{for all } x \neq 0.$$

A positive definite hermitian form on H is often called a *hermitian inner product* on H, and H is called a *hermitian space* (sometimes a *pre-Hilbert space*).

If H is a hermitian space with a hermitian inner product $\langle -, - \rangle$, then the map $x \mapsto \|x\| = \sqrt{\langle x, x\rangle}$ is a norm on H. We say that H is a *Hilbert space* if H is complete for the norm $\| \;\|$ (every Cauchy sequence converges).

Let H be a Hilbert space. An orthonormal family $(a_\alpha)_{\alpha\in\Lambda}$ of vectors $a_\alpha \in H$ (which means that $\langle a_\alpha, a_\beta\rangle = 0$ for all $\alpha \neq \beta$, and $\langle a_\alpha, a_\alpha\rangle = 1$, for all $\alpha, \beta \in \Lambda$) is a *Hilbert basis* of H if the subspace spanned by $(a_\alpha)_{\alpha\in\Lambda}$ (the set of all *finite* linear combinations of vectors in $(a_\alpha)_{\alpha\in\Lambda}$) is dense in H. Every Hilbert space admits a Hilbert basis, and the cardinality of the index set Λ is the same for any two Hilbert bases; see Section D.2, Rudin [79] (Chapter 4) and Schwartz [83] (Chapter XXIII).

A Hilbert space is *separable* if it has a countable Hilbert basis.

Definition 11.1. Given an algebra A with involution and a Hilbert space H, a *representation of A in H*[1] is an algebra homomorphism $U\colon A \to \mathcal{L}(H)$ from A to the involutive algebra $\mathcal{L}(H)$ of continuous linear maps from H to itself, which means that U satisfies the conditions

$$\begin{aligned} U(s+t) &= U(s) + U(t) \\ U(\lambda s) &= \lambda U(s) \\ U(st) &= U(s) \circ U(t) \\ U(s^*) &= (U(s))^*, \end{aligned}$$

for all $s, t \in A$ and all $\lambda \in \mathbb{C}$. If A is unital with identity element e, we require that

$$U(e) = \mathrm{id}_H.$$

The Hilbert space H is called the *representation space*. The representation U is *faithful* if the homomorphism $s \mapsto U(s)$ is injective, which means that $U(s)(x) = 0$ for all $x \in H$ implies that $U(s) = 0$.

Following common practice, the composition $U(s) \circ U(t)$ is abbreviated as $U(s)U(t)$. To simplify notation, we often write U_s instead of $U(s)$.

Remark: Folland [33] uses the terminology $*$-*representation* for a representation of an involutive algebra. When different representations $U\colon A \to \mathcal{L}(H)$ of the same algebra A arise, it is sometimes convenient to denote the representation space by H_U. Although a representation of an algebra A consists of a homomorphism U and of a Hilbert space H, by abuse of language, we often refer to a representation as U.

Example 11.1. Let $A = \mathrm{M}_n(\mathbb{C})$ be the algebra of $n \times n$ complex matrices with involution $X \mapsto X^*$ (the conjugate transpose). The map

$$\langle X, Y\rangle \mapsto \langle X, Y\rangle = \mathrm{tr}(Y^* X)$$

is a Hermitian inner product on A which makes A into a Hilbert space of finite dimension denoted H. The linear maps in $\mathrm{Hom}(H, H)$ are automatically continuous, so $\mathrm{Hom}(H, H) = \mathcal{L}(H)$. The map $U\colon A \to \mathcal{L}(H)$ given by

$$U(X)(Y) = XY, \quad X, Y \in \mathrm{M}_n(\mathbb{C})$$

is a representation of A. The only property that is not obvious is the property $U(X^*) = U(X)^*$. But by definition the adjoint $V(X) = U(X)^*$ of the linear map $U(X)$ is the unique linear map $V(X)$ such that

$$\langle U(X)(Y), Z\rangle = \langle Y, V(X)(Z)\rangle \quad \text{for all } Y, Z \in \mathrm{M}_n(\mathbb{C}),$$

[1] Technically, we are defining *unitary representations*, presumably because the definition of equivalence of representations (Definition 11.2) uses isometries betweeen Hilbert spaces, but since we shall not discuss other types of representations, we shall suppress the word "unitary."

that is,

$$\mathrm{tr}(Z^*XY) = \mathrm{tr}((V(X)(Z))^*Y) \quad \text{for all } Y, Z \in \mathrm{M}_n(\mathbb{C}),$$

which implies $V(X)(Z) = (Z^*X)^* = X^*Z$. Thus $U(X)^*(Z) = V(X)(Z) = X^*Z = U(X^*)(Z)$, that is,

$$U(X)^* = U(X^*),$$

as desired.

Example 11.2. Let $A = \mathrm{M}_n(\mathbb{C})$ be the algebra of $n \times n$ complex matrices with involution $X \mapsto X^*$ (the conjugate transpose), and let $H = \mathbb{C}^n$, with the standard hermitian inner product given by $\langle x, y\rangle = y^*x$, where $x, y \in \mathbb{C}^n$. Since H is finite-dimensional, it is a Hilbert space, and the linear maps in $\mathrm{Hom}(H, H)$ are automatically continuous, so $\mathrm{Hom}(H, H) = \mathcal{L}(H)$. The map $U_1 \colon A \to \mathcal{L}(H)$ given by

$$U_1(X)(y) = Xy, \quad X \in \mathrm{M}_n(\mathbb{C}), y \in \mathbb{C}^n$$

is a representation of A. The only property that is not obvious is the property $U_1(X^*) = U_1(X)^*$. But by definition the adjoint $V(X) = U_1(X)^*$ of the linear map $U_1(X)$ is the unique linear map $V(X)$ such that

$$\langle U_1(X)(y), z\rangle = \langle y, V(X)(z)\rangle \quad \text{for all } y, z \in \mathbb{C}^n,$$

that is,

$$z^*Xy = (V(X)(z))^*y, \quad \text{for all } y, z \in \mathbb{C}^n,$$

thus $U_1(X)^*(z) = V(X)(z) = X^*z = U_1(X^*)(z)$, namely

$$U_1(X)^* = U_1(X^*).$$

Observe that $H = \mathbb{C}^n$ is isomorphic to the subspace $\mathfrak{b}$ of A consisting of all $n \times n$ complex matrices whose last $n-1$ columns are zero. The subspace $\mathfrak{b}$ is a left ideal in A, and in fact a minimal left ideal. The map $U_2 \colon A \to \mathcal{L}(\mathfrak{b})$ given by

$$U_2(X)(Y) = XY, \quad X \in \mathrm{M}_n(\mathbb{C}), Y \in \mathfrak{b}$$

is also a representation of A. Since $\mathbb{C}^n$ and $\mathfrak{b}$ are isomorphic Hilbert spaces, we say that U_1 and U_2 are equivalent representations; see Definition 11.2.

A generalization of this example occurs in Proposition 11.18.

Example 11.3. If G is a metrizable locally compact group, the space $A = \mathrm{L}^1(G)$ is an algebra under convolution (denoted $*$) and $H = \mathrm{L}^2(G)$ is a Hilbert space. It is shown in Section 12.3 that the map $\mathbf{R}_{\mathrm{ext}} \colon \mathrm{L}^1(G) \to \mathcal{L}(\mathrm{L}^2(G))$ given by

$$(\mathbf{R}_{\mathrm{ext}}(f))(g) = f * g, \qquad f \in \mathrm{L}^1(G),\, g \in \mathrm{L}^2(G)$$

is a representation of $\mathrm{L}^1(G)$ in $\mathrm{L}^2(G)$ (called *left regular representation*).

Definition 11.1 implies that if s is self-adjoint ($s^* = s$), then $U(s)$ is self-adjoint. Observe that $U(s)$ is *not* necessarily invertible. Also, if A is a normed algebra, then the map $U\colon A \to \mathcal{L}(H)$ is *not* necessarily continuous. However, if A is a unital Banach algebra with involution, then by the next proposition the map $U\colon A \to \mathcal{L}(H)$ is continuous.

Proposition 11.1. *If A is a unital Banach algebra with involution, then every representation $U\colon A \to \mathcal{L}(H)$ satisfies the condition $\|U(s)\| \leq \|s\|$, which implies that U is a continuous mapping from A to $\mathcal{L}(H)$.*

Proof. Recall that $\mathcal{L}(H)$ is a C^*-algebra, so by Proposition 9.31 $\rho(T) = \|T\|$ for all normal linear maps $T \in \mathcal{L}(H)$, so if we let $T = U(s)^* U(s)$, which is obviously self-adjoint, we have

$$\|U(s)\|^2 = \|U(s)^* U(s)\| = \rho(U(s)^* U(s)).$$

By Property (5) just after Definition 9.6,

$$\sigma(U(s^* s)) \subseteq \sigma(s^* s),$$

by Proposition 9.17,

$$\rho(U(s^* s)) \leq \rho(s^* s),$$

and $U(s)^* U(s) = U(s^* s)$, so we have

$$\rho(U(s)^* U(s)) = \rho(U(s^* s)) \leq \rho(s^* s) \leq \|s^* s\| \leq \|s^*\| \, \|s\| = \|s\|^2 ,$$

which proves our result. $\square$

Recall that if $(H_1, \langle -, - \rangle_1)$ and $(H_2, \langle -, - \rangle_2)$ are two Hilbert spaces, a Hilbert space isomorphism is a continuous linear map $T\colon H_1 \to H_2$ whose inverse is also continuous, and T is an isometry, which means that

$$\langle T(x), T(y) \rangle_2 = \langle x, y \rangle_1 \qquad \text{for all } x, y \in H.$$

If $H_1 = H_2$, then a Hilbert space automorphism $T\colon H \to H$ is an invertible element in $\mathcal{L}(H)$ such that

$$TT^* = T^* T = \mathrm{id},$$

where T^* is the adjoint of T (defined by the property that

$$\langle Tx, y \rangle = \langle x, T^* y \rangle \qquad \text{for all } x, y \in H).$$

Such maps are called *unitary*.

Definition 11.2. If A is an algebra (as above) and if H_1 and H_2 are two Hilbert spaces, two representations $U_1\colon A \to \mathcal{L}(H_1)$ and $U_2\colon A \to \mathcal{L}(H_2)$ are *equivalent* if there a Hilbert space isomorphism $T\colon H_1 \to H_2$ such that

$$U_2(s) = T U_1(s) T^{-1} \qquad \text{for all } s \in A,$$

as illustrated by the following diagram:

$$\begin{tikzcd} H_1 \arrow[r, "U_1(s)"] & H_1 \arrow[d, "T"] \\ H_2 \arrow[u, "T^{-1}"] \arrow[r, "U_2(s)"'] & H_2. \end{tikzcd}$$

Example 11.4. The representations $U_1 \colon A \to \mathcal{L}(\mathbb{C}^n)$ and $U_2 \colon A \to \mathcal{L}(\mathfrak{b})$ of Example 11.2 are equivalent under the obvious isomorphism from $\mathbb{C}^n$ to $\mathfrak{b}$.

It is often useful to make a new representation from old ones using the process of constructing a Hilbert sum. We begin with the simplest case involving two Hilbert spaces. Later we generalize this construction to an arbitrary family of Hilbert spaces.

Let H_1 and H_2 be two Hilbert spaces, and let $U_1 \colon A \to \mathcal{L}(H_1)$ and $U_2 \colon A \to \mathcal{L}(H_2)$ be two representations. The *Hilbert sum* H of H_1 and H_2 is the direct sum $H_1 \oplus H_2$ of H_1 and H_2 with the hermitian product given by

$$\langle x_1 + x_2, y_1 + y_2 \rangle = \langle x_1, y_1 \rangle_1 + \langle x_2, y_2 \rangle_2,$$

for all $x_1, y_1 \in H_1$ and all $x_2, y_2 \in H_2$, We define the representation $U \colon A \to H$ by

$$U(s)(x_1 + x_2) = U_1(s)(x_1) + U_2(s)(x_2),$$

for all $x_1 \in H_1$ and all $x_2 \in H_2$. It is immediately verified that $U(s) \in \mathcal{L}(H)$ for all $s \in A$, and that U is a representation of A.

Definition 11.3. The representation U constructed as above from two representations $U_1 \colon A \to \mathcal{L}(H_1)$ and $U_2 \colon A \to \mathcal{L}(H_2)$ is called the *Hilbert sum* of U_1 and U_2.

We now generalize the construction of Hilbert sum to any arbitrary family of Hilbert spaces. The generalization to representations will be made in the next section.

Let $(H_\alpha, \langle -, - \rangle_\alpha)_{\alpha \in \Lambda}$ be a family of Hilbert spaces indexed by some index set Λ. In most applications, $\Lambda = \mathbb{N}$, so for simplicity the reader may assume this. We define the set H as the set of all sequences $(x_\alpha)_{\alpha \in \Lambda}$ with $x_\alpha \in H_\alpha$, such that $\sum_{\alpha \in \Lambda} \|x_\alpha\|_{H_\alpha}^2 < \infty$. Since the index set Λ may not be countable, what we are asserting is that the family $(\|x_\alpha\|_{H_\alpha}^2)_{\alpha \in \Lambda}$ is summable; see Definition D.6 (in particular, this implies that only countably many elements x_α are nonzero). We define a vector space structure on H by defining

$$\begin{aligned} (x_\alpha) + (y_\alpha) &= (x_\alpha + y_\alpha) \\ \lambda(x_\alpha) &= (\lambda x_\alpha), \end{aligned}$$

with $x_\alpha, y_\alpha \in H_\alpha$. It is easy to check that these operations make H into a vector space. We define the inner product $\langle -, - \rangle$ on H by

$$\langle (x_\alpha), (y_\alpha) \rangle = \sum_{\alpha \in \Lambda} \langle x_\alpha, y_\alpha \rangle_\alpha.$$

It can be verified that $\langle -, - \rangle$ is a Hermitian inner product on H. It can also be shown that H is complete, so it is a Hilbert space. For details, see Dieudonné [25], (Chapter VI, Section 4) and Schwartz [83] (Chapter XXIII, Theorem 1 and Theorem 2).

Definition 11.4. Let $(H_\alpha, \langle -, - \rangle_\alpha)_{\alpha \in \Lambda}$ be a family of Hilbert spaces indexed by some index set Λ. The space H constructed as above is called the *Hilbert sum* of the sequence of Hilbert spaces (H_α), and is denoted by

$$H = \bigoplus_{\alpha \in \Lambda} H_\alpha.$$

We define continuous injections $j_\alpha \colon H_\alpha \to H$ such that $j_\alpha(x_\alpha) = (0, \ldots, 0, x_\alpha, 0, \ldots)$, with the αth term being x_α. Each j_α is an isomorphism of H_α onto a closed subspace of H denoted H'_α. By definition of the inner product on H, $\langle j_\alpha(x_\alpha), j_\beta(x_\beta) \rangle = 0$ for all $\alpha \neq \beta$, all $x_\alpha \in H_\alpha$, and all $x_\beta \in H_\beta$. A very useful fact is that the direct sum $\bigoplus_{\alpha \in \Lambda} H'_\alpha$ is dense in H (recall that $\bigoplus_{\alpha \in \Lambda} H'_\alpha$ consists of all sequences $(x_\alpha)_{\alpha \in \Lambda}$ such that $x_\alpha = 0$ for all but finitely many indices α).

Proposition 11.2. *Let H be the direct sum of a family $(H_\alpha, \langle -, - \rangle_\alpha)_{\alpha \in \Lambda}$ of Hilbert spaces indexed by some index set Λ. For every $x = (x_\alpha)_{\alpha \in \Lambda} \in H$, the family $(j_\alpha(x_\alpha))_{\alpha \in \Lambda}$ of vectors in $\bigoplus_{\alpha \in \Lambda} H'_\alpha$ is summable and we have*

$$x = \sum_{\alpha \in \Lambda} j_\alpha(x_\alpha)$$

(the convergence is not necessarily uniform). Consequently, $\bigoplus_{\alpha \in \Lambda} H'_\alpha$ is dense in H.

Proof. The fact that $\sum_{\alpha \in \Lambda} \|x_\alpha\|^2_{H_\alpha} < \infty$ implies that for every ϵ, there is some finite subset J of Λ such that

$$\sum_{\alpha \in \Lambda - J} \|x_\alpha\|^2_{H_\alpha} = \sum_{\alpha \in \Lambda - J} \|j_\alpha(x_\alpha)\|^2_{H'_\alpha} \leq \epsilon;$$

this is the usual Cauchy property. Then for every subset K (finite or not) such that $J \subseteq K \subseteq \Lambda$, the family $(j_\alpha(x_\alpha))_{\alpha \in K}$ is summable in H, and if we let

$$x_K = \sum_{\alpha \in K} j_\alpha(x_\alpha),$$

we have

$$\|x - x_K\|_H = \left(\sum_{\alpha \in \Lambda - K} \|x_\alpha\|^2 \right)^{1/2} = \left(\sum_{\alpha \in \Lambda - K} \|j_\alpha(x_\alpha)\|^2 \right)^{1/2} \leq \left(\sum_{\alpha \in \Lambda - J} \|j_\alpha(x_\alpha)\|^2 \right)^{1/2} \leq \epsilon,$$

which, by Definition D.6, proves that $(j_\alpha(x_\alpha))_{\alpha \in \Lambda}$ is summable in H and that its sum is x. $\square$

Remark: By picking $\epsilon = 1/(n+1)$, we can define a sequence of *finite* subsets $K_n \subseteq K_{n+1}$ such that $\sum_{\alpha \in K_n} j_\alpha(x_\alpha)$ converges to x in H.

We often identify H_α and H'_α. The above construction defines what we might call an external Hilbert sum.

Unfortunately, the notation

$$H = \bigoplus_{\alpha \in \Lambda} H_\alpha$$

for the Hilbert sum of a family $(H_\alpha, \langle -, - \rangle_\alpha)_{\alpha \in \Lambda}$ of Hilbert spaces clashes with the notion of algebraic *direct sum*

$$\bigoplus_{\alpha \in \Lambda} H_\alpha$$

of vector spaces. The second definition refers to the subspace of sequences $(x_\alpha)_{\alpha \in \Lambda}$ such that $x_\alpha = 0$ *for all but finitely many indices* α. If we temporarily denote the Hilbert sum by

$$H = \bigoplus_{\alpha \in \Lambda}^{\mathrm{H}} H_\alpha,$$

then we see that if the index set Λ is finite, then the two notions agree. But if Λ is infinite, then the algebraic direct sum is a proper subspace of the Hilbert sum, because the Hilbert sum consists of sequences $(x_\alpha)_{\alpha \in \Lambda}$ such that $\sum_{\alpha \in \Lambda} \|x_\alpha\|^2 < \infty$, which may contain a countably infinite number of nonzero x_α. The direct sum is a dense subspace of the Hilbert sum. It is usually clear from the context which sum of spaces is intended (typically, when we refer to Hilbert spaces, we mean a Hilbert sum), so we will not use the heavier notation $\bigoplus^{\mathrm{H}}$ for Hilbert sums.

We also have the following proposition proven in Dieudonné [25], (Chapter VI, Section 4) and Schwartz [83] (Chapter XXIII, Theorem 4 and its corollaries), which gives the definition of an internal Hilbert sum (in the sense that the H_α are subspaces of an already given space H).

Proposition 11.3. *Let H be a Hilbert space, and let $(H_\alpha)_{\alpha \in \Lambda}$ be a family of closed subspaces of H satisfying the following conditions:*

(1) For all $\alpha \neq \beta$, the subspaces H_α and H_β are orthogonal.

(2) The direct sum $\bigoplus_{\alpha \in \Lambda} H_\alpha$ is dense in H.

If E is the Hilbert sum of the sequence (H_α), then there is a unique isomorphism of H onto E which, on each H_α, coincides with the injection j_α of H_α into E.

If Λ is finite, (2) is equivalent to the fact that H is the direct sum of the H_α (because, if E is any closed subspace of a Hilbert space, then $H = E \oplus E^\perp$ as a direct sum, so by picking $E = H_{\alpha_1}$, we see that the direct sum $\bigoplus_{j=2}^n H_{\alpha_j}$ is dense in $E^\perp$, and we finish by induction).

Two key notions of representation theory are invariant subspaces and (topologically) irreducible representations.

11.2 Invariant Subspaces and Irreducible Representations

Definition 11.5. Let $U\colon A \to \mathcal{L}(H)$ be a representation of an algebra A. A subspace E of H is *invariant* (or *stable*) under the representation U if $U(s)(x) \in E$ for all $s \in A$ and all $x \in E$. The representation $U_E\colon A \to \mathcal{L}(E)$ given by $U_E(s)(x) = U(s)(x)$ for all $x \in E$, is called a *subrepresentation* of A in E.

Example 11.5. In Example 11.2, the subspace $\mathfrak{b}$ of $A = \mathrm{M}_n(\mathbb{C})$ is invariant under the representation $U\colon A \to \mathcal{L}(H)$ of Example 11.1 and $U_1\colon A \to \mathcal{L}(\mathfrak{b})$ is a subrepresentation of A in $\mathfrak{b}$.

Observe a small abuse of language: if E is not a closed subspace of H, then E is not a Hilbert space, and so $U_E\colon A \to \mathcal{L}(E)$ is not a representation. Thus the notion of subrepresentation should be defined for *closed* invariant subspaces of H. However, Proposition 11.4 shows that the closure $\overline{E}$ of an invariant subspace E is invariant, so we can define the subrepresentation $U_E\colon A \to \mathcal{L}(\overline{E})$.

The following facts hold.

Proposition 11.4. *Let $U\colon A \to \mathcal{L}(H)$ be a representation of an algebra A.*

(1) If the subspace E of H is invariant under U, then its closure $\overline{E}$ is also invariant under U.

(2) Let E be a closed subspace of H invariant under U. If $E^\perp$ is the orthogonal complement of E in H, then $E^\perp$ is invariant under U. If $U_1(s)$ and $U_2(s)$ are the restrictions of $U(s)$ to E and $E^\perp$, then the representation U is the Hilbert sum of the representations U_1 and U_2.

Proof. Part (1) is easy to prove and follows from the continuity of $U(s)$; see Dieudonné [25], (Chapter III, Section 11). For Part (2), let $x \in E$ and $y \in E^\perp$. For any $s \in A$ we have

$$\langle x, U(s)(y)\rangle = \langle (U(s))^*(x), y\rangle = \langle U(s^*)(x), y\rangle = 0,$$

since E is invariant under U, so $U(s^*)(x) \in E$, and since $E^\perp$ is the orthogonal complement of E and $y \in E^\perp$. Then $U(s)(y)$ is orthogonal to all $x \in E$, which means that $U(s)(y) \in E^\perp$, so $E^\perp$ is invariant under U. The last property is obvious because as E is closed, H is the (algebraic) direct sum $H = E \oplus E^\perp$. $\square$

The notion of Hilbert sum of representations is generalized to arbitrary Hilbert sums as follows.

Definition 11.6. Assume a Hilbert space H is the Hilbert sum of a sequence $(H_\alpha)_{\alpha\in\Lambda}$ of subspaces invariant under a representation U of A. For every $s \in A$, let $U_\alpha(s)$ be the

restriction of U to H_α, so that the map $s \mapsto U_\alpha(s)$ is a representation of A in H_α. By abuse of language, we say that U is the *Hilbert sum* of the representations U_α. For each $s \in A$ and each $x = \sum_\alpha x_\alpha \in H$, where $x_\alpha \in H_\alpha$, we have

$$U(s)(x) = \sum_\alpha U_\alpha(s)(x_\alpha),$$

and

$$\sum_\alpha \|U_\alpha(s)(x_\alpha)\|^2 = \|U(s)(x)\|^2 .$$

Recall Definition D.2 of the orthogonal projection p_V of a Hilbert space E onto a closed subspace V. Such a map is linear and continuous. It is also called an *orthogonal projector*. In this chapter we denote p_V by P_V to conform to Dieudonné.

The following result is not hard to prove; see Dieudonné [24], (Chapter XV, Section 5).

Proposition 11.5. *Let H be a Hilbert space. A continuous linear P map on H is an orthogonal projector iff it is idempotent ($P^2 = P \circ P = P$) and hermitian ($P^* = P$).*

Here is a convenient way to characterize when a closed subspace is invariant.

Proposition 11.6. *Let $U\colon A \to \mathcal{L}(H)$ be a representation of an algebra A. A closed subspace E of H is invariant under U iff $P_E U(s) = U(s) P_E$ for all $s \in A$, as illustrated in the diagram below*

$$\begin{tikzcd} H \arrow[r, "U(s)"] \arrow[d, "P_E"'] & H \arrow[d, "P_E"] \\ E \arrow[r, "U(s)"'] & E, \end{tikzcd}$$

where $P_E\colon H \to E$ is the orthogonal projection of H onto E.

The proof of Proposition 11.6 is identical to the proof for group representations; see Proposition 12.7 and its proof. Proposition 11.6 is also proven in Dieudonné [24], (Chapter XV, Section 5, Proposition 15.5.3).

The notion of a topologically irreducible representation is similar in spirit to the notion of prime number. Namely, a topologically irreducible representation cannot be decomposed into simpler representations. It is one of the most important concepts in representation theory.

Definition 11.7. A representation $U\colon A \to \mathcal{L}(H)$ of A in H is *topologically irreducible* if $H \neq (0)$ and if there is no *closed* subspace E of H other than $\{0\}$ and H which is invariant under U.

Example 11.6. The representation $U\colon A \to \mathcal{L}(H)$ of Example 11.1 is reducible because the proper nonzero subspace $\mathfrak{b}$ of H is invariant under U. On the other hand, by Theorem 11.34, the representation $U_2\colon A \to \mathcal{L}(\mathfrak{b})$ of Example 11.2 is topologically irreducible (in finite dimension, a subspace is automatically closed).

Proposition 11.7. *Let $U\colon A \to \mathcal{L}(H)$ be a representation of A in H, let E be the closure of the subspace spanned by the set*

$$\{U(s)(x) \mid s \in A,\ x \in H\},$$

and let E' be the set

$$E' = \{x \in H \mid U(s)(x) = 0,\ \textit{for all } s \in A\}.$$

Then E and E' are invariant under U, and E' is the orthogonal complement of E in H (that is, $E' = E^{\perp}$ and $H = E \oplus E'$).

Proof. Since $U(rs) = U(r)U(s)$ for all $r, s \in A$, it is clear that E and E' are invariant under U. Let us prove that $E^{\perp} \subseteq E'$, where $E^{\perp}$ is the orthogonal complement of E in H. We already know from Proposition 11.4 that $E^{\perp}$ is invariant under U so for any $x \in E^{\perp}$ we have $U(s)(x) \in E^{\perp}$ for all $s \in A$. But by definition of E, we have $U(s)(x) \in E$, so $U(s)(x) \in E \cap E^{\perp} = (0)$, which means that $U(s)(x) = 0$ for all $s \in A$, that is, $x \in E'$. Therefore $E^{\perp} \subseteq E'$.

Next we prove that $E' \subseteq E^{\perp}$. If $x \in E'$, for any $s \in A$ and any $y \in H$ we have

$$\langle x, U(s)(y) \rangle = \langle U(s^*)(x), y \rangle = 0,$$

since $x \in E'$ means that $U(s)(x) = 0$ for all $s \in A$, and since s and y are arbitrary, by definition of E, this means that x is orthogonal to E, that is, $x \in E^{\perp}$, and thus $E' \subseteq E^{\perp}$. In conclusion, $E' = E^{\perp}$. □

Definition 11.8. Let $U\colon A \to \mathcal{L}(H)$ be a representation of A in H. The subspace E defined in Proposition 11.7 is called the *essential subspace* for U. If $E' = (0)$, then we say that the representation is *nondegenerate*.

Although very easy to prove, the following result is important and often used.

Proposition 11.8. *A representation U of A in H is nondegenerate iff the subspace spanned by set $\{U(s)(x) \mid s \in A,\ x \in H\}$ is dense in H. If A is unital, since $U(e) = \mathrm{id}$, this is always the case.*

Definition 11.9. Let $U\colon A \to \mathcal{L}(H)$ be a representation of A in H. A vector $x_0 \in H$ is called a *totalizer* or *totalizing vector* (or *cyclic vector*) for the representation U if the subspace of H spanned by the set $\{U(s)(x_0) \mid s \in A\}$ is dense in H. Equivalently if $\mathcal{M}_{x_0}$ denotes the closure of the set $\{U(s)(x_0) \mid s \in A\}$, called the *cyclic subspace* generated by x_0, which is invariant under U, then x_0 is a totalizer (a cyclic vector) if $\mathcal{M}_{x_0} = H$. A representation which admits a totalizer is said to be *topologically cyclic*.

The following fact follows immediately from the definitions: a representation U is topologically irreducible iff every nonzero vector $x_0 \in H$ is a totalizer. The importance of totalizers stems from the following result.

Proposition 11.9. *Assume that the algebra A is unital, and let $U: A \to \mathcal{L}(H)$ be a representation of A in $H \neq (0)$ (which must be nondegenerate). Then H is the Hilbert sum of a sequence $(H_\alpha)_{\alpha\in\Lambda}$ of closed subspaces $H_\alpha \neq (0)$ of H invariant under U, and such that the restriction of U to each H_α is topologically cyclic. If H is separable, the family Λ is countable (possibly finite).*

Proof. We prove the proposition in the separable case, following Dieudonné [24] (Chapter XV, Proposition 15.5.6). The general case uses Zorn's lemma; see Folland [33] (Chapter 3, Proposition 3.3). Let (x_n) be a dense sequence (finite or countably infinite) in H. We define the sequence (H_n) by induction as follows. Let H_1 be the subspace spanned by the set of vectors $\{H(s)(x_1) \mid s \in A\}$. Since A is unital, $x_1 \in H_1$. Assuming that $H_1, \ldots, H_n$ have been defined, either H is the direct sum of the H_i, in which case we are done, or else we proceed as follows. Let $L \neq (0)$ be the orthogonal complement of the direct sum $H_1\oplus\cdots\oplus H_n$ in H. Then let $p(n+1)$ be the smallest index such that if y_{n+1} is the orthogonal projection of $x_{p(n+1)}$ on L, then the subspace H'_{n+1} of L generated by the subset $\{U(s)(y_{n+1}) \mid s \in A\}$ is not the trivial subspace (0). Since A is unital, $y_{n+1} \in \{U(s)(y_{n+1}) \mid s \in A\}$ because $U(\mathbf{1}) = \mathrm{id}$, so such an index must exist. By definition of $p(n+1)$, we have $x_1, \ldots, x_{p(n+1)-1} \in H_1 \oplus \cdots \oplus H_n$. Let H_{n+1} be the closure of H'_{n+1} in H. Since $H_{n+1} \subseteq L$ and L is the orthogonal complement of the direct sum $H_1\oplus\cdots\oplus H_n$ in H, the fact that $y_{n+1} \in H_{n+1}$ implies that $x_{p(n+1)} \in H_1\oplus\cdots\oplus H_{n+1}$, so the direct sum of the H_k contains the dense sequence (x_n), and by Proposition 11.3, the space H is indeed the Hilbert sum of the H_k. $\square$

11.3 Positive Linear Forms and Positive Hilbert Forms

The Peter–Weyl theorem can be obtained from a structure theorem about certain kinds of algebras with a hermitian inner product satisfying special conditions. Such inner products are bitraces, which are special kinds of positive Hilbert forms. A good method for producing positive Hilbert forms is to use positive linear forms.

Definition 11.10. Let A be an involutive algebra (not necessarily commutative, unital, and not necessarily normed). A linear form $f: A \to \mathbb{C}$ is *positive* if

$$f(s^*s) \geq 0 \qquad \text{for all } s \in A.$$

Positive linear forms arise from representations as follows.

Definition 11.11. Let $U: A \to \mathcal{L}(H)$ be a representation of A in H. For any $x_0 \in H$, define the map $f_{x_0}: A \to \mathbb{C}$ by

$$f_{x_0}(s) = \langle U(s)(x_0), x_0\rangle, \quad s \in A.$$

Proposition 11.10. *The map $f_{x_0}: A \to \mathbb{C}$ is a positive linear form.*

Proof. It is clear that f_{x_0} is a linear form, and since $U(s^*s) = U(s^*)U(s) = U(s)^*U(s)$, we have

$$\begin{aligned} f_{x_0}(s^*s) &= \langle (U(s^*s)(x_0), x_0\rangle = \langle (U(s)^*U(s))(x_0), x_0\rangle \\ &= \langle U(s)(x_0), U(s)(x_0)\rangle = \|U(s)(x_0)\|^2 \geq 0. \end{aligned}$$ ☐

A positive linear form also defines a positive hermitian form as follows.

Proposition 11.11. *Given any positive linear form f on an involutive algebra A, let $g\colon A \times A \to \mathbb{C}$ be the map given by*

$$g(x, y) = f(y^*x), \qquad \text{for all } x, y \in A.$$

Then g is a positive hermitian form, and the following properties hold:

(1) For all $x, y \in A$, we have
$$f(x^*y) = \overline{f(y^*x)}.$$

(2) For all $x, y \in A$, we have
$$|f(y^*x)|^2 \leq f(x^*x)f(y^*y).$$

(3) If A is unital, then
$$f(x^*) = \overline{f(x)}, \qquad |f(x)|^2 \leq f(e)f(x^*x).$$

Proof. To prove (1), since f is linear, we have

$$\begin{aligned} g(x+y, x+y) &= f((x+y)^*(x+y)) \\ &= f(x^*x + x^*y + y^*x + y^*y) \\ &= f(x^*x) + f(x^*y) + f(y^*x) + f(y^*y) \\ &= g(x, x) + g(y, x) + g(x, y) + g(y, y). \end{aligned}$$

Since $g(x + y, x + y), g(x, x)$, and $g(y, y)$ are real (and nonnegative), we must have

$$\Im(g(y, x)) = -\Im(g(x, y)).$$

If we replace x by ix, this becomes

$$\Re(g(y, x)) = \Re(g(x, y)).$$

Therefore,
$$g(y, x) = f(x^*y) = \overline{f(y^*x)} = \overline{g(x, y)},$$

as claimed. Part (2) is the Cauchy–Schwartz inequality, and (3) follows from (1) and (2) by replacing y by e and the fact that $e^* = e$. ☐

The hermitian form g obtained from the positive linear form f is not arbitrary since it satisfies the condition

$$g(xy, z) = f(z^*xy) = f((x^*z)^*y) = g(y, x^*z)$$

for all $x, y, z \in A$.

This motivates the following definition.

Definition 11.12. A *positive Hilbert form* on an involutive algebra A is a positive hermitian form g satisfying the condition

$$g(xy, z) = g(y, x^*z) \qquad \text{for all } x, y, z \in A.$$

Proposition 11.12. *If A is unital with unit e, then every positive Hilbert form g comes from the positive linear form f given by $f(s) = g(s, e)$ for all $s \in A$.*

Proof. Indeed, the positive Hilbert form g' induced by f is given by

$$g'(s, t) = f(t^*s) = g(t^*s, e) = g(s, t),$$

by setting $x = t^*, y = s$, and $z = e$ in the equation of Definition 11.12. $\square$

Given a representation $U\colon A \to \mathcal{L}(H)$, observe that the positive Hilbert form g_{x_0} associated with the positive linear form f_{x_0} is given by

$$g_{x_0}(s, t) = f_{x_0}(t^*s) = \langle U(s)(x_0), U(t)(x_0)\rangle, \qquad s, t \in A.$$

Remarkably, every topologically cyclic representation arises from a positive Hilbert form, but we won't need this fact until Section 11.8, so we postpone discussing this matter.

11.4 Traces, Bitraces, Hilbert Algebras

If A is an involutive algebra and if f is a positive linear form on A, in general $f(st) \neq f(ts)$.

Definition 11.13. Let A be an involutive algebra. A *trace* on A is a positive linear form $f\colon A \to \mathbb{C}$ such that

$$f(st) = f(ts) \qquad \text{for all } s, t \in A.$$

A *bitrace* is a positive Hilbert form $g\colon A \times A \to \mathbb{C}$ such that

$$g(t^*, s^*) = g(s, t), \qquad \text{for all } s, t \in A.$$

Example 11.7. If H is a finite-dimensional vector space of dimension n with a hermitian inner product, and if $A = \mathcal{L}(H)$, the algebra of linear maps from H to itself, for any orthonormal basis $(e_1, \ldots, e_n)$ of H, then for any linear map $T \in \mathcal{L}(H)$, the linear form

$$\mathrm{Tr}(T) = \sum_{i=1}^{n} \langle T(e_i), e_i \rangle$$

is a trace. In fact, $\mathrm{Tr}(T) = \sum_{i=1}^{n} a_{ii}$, the trace of the matrix (a_{ij}) representing T over the basis $(e_1, \ldots, e_n)$. If H is an infinite-dimensional Hilbert space, it can be shown that there exists no trace on $\mathcal{L}(H)$.

Proposition 11.13. *Let $f\colon A \to \mathbb{C}$ be a positive linear form on an involutive algebra A. If*

$$f(ss^*) = f(s^*s), \qquad \textit{for all } s \in A,$$

then f is a trace.

Proof. By replacing s by $s + t$, using Proposition 11.11(1), we have

$$\begin{aligned} f((s+t)(s+t)^*) &= f(ss^* + st^* + ts^* + tt^*) \\ &= f(ss^*) + f(st^*) + f(ts^*) + f(tt^*) \\ &= f(s^*s) + f(st^*) + \overline{f(st^*)} + f(t^*t), \end{aligned}$$

and similarly

$$f((s+t)^*(s+t)) = f(s^*s) + \overline{f(t^*s)} + f(t^*s) + f(t^*t)$$

and since $f((s+t)(s+t)^*) = f((s+t)^*(s+t))$, we get

$$f(st^*) + \overline{f(st^*)} = f(t^*s) + \overline{f(t^*s)},$$

that is

$$\Re(f(t^*s)) = \Re(f(st^*)).$$

If we replace s by is, we get

$$\Im(f(t^*s)) = \Im(f(st^*)),$$

so that

$$f(t^*s) = f(st^*),$$

for all $s, t \in A$, as claimed. ∎

Proposition 11.14. *If the positive Hilbert form g on an involutive algebra A arising from a positive linear form f as $g(s,t) = f(t^*s)$ is a bitrace, then f is a trace. Conversely, if f is a trace, then g is a bitrace.*

Proof. Expressing that g is bitrace says that

$$f(t^*s) = g(s,t) = g(t^*, s^*) = f((s^*)^*t^*) = f(st^*),$$

namely that f is a trace. The same computation shows that if f is trace, then g is a bitrace. □

One of the most important example of a bitrace arises when G is a compact group. In this case, the inner product on the involutive Banach algebra $\mathrm{L}^2(G)$ is a bitrace. In fact, this bitrace satisfies two more properties that makes $\mathrm{L}^2(G)$ into a Hilbert algebra, defined next.

Definition 11.14. An involutive algebra A is a *Hilbert algebra* if its underlying vector space is a hermitian space whose hermitian inner product $\langle -, - \rangle$ is a bitrace satisfying two extra conditions (U) and (N). Specifically, the conditions for being a bitrace hold

$$\langle y^*, x^* \rangle = \langle x, y \rangle \tag{1}$$
$$\langle xy, z \rangle = \langle y, x^*z \rangle, \tag{2}$$

and the following two conditions hold: for every $x \in A$, there is some $M_x \geq 0$ such that

$$\langle xy, xy \rangle \leq M_x \langle y, y \rangle, \quad \text{for all } y \in A, \tag{U}$$

and

$$\text{the subspace spanned by the set } \{xy \mid x, y \in A\} \text{ is dense in } A. \tag{N}$$

From Conditions (1) and (2) and the hermitian property, we get

$$\langle yx, z \rangle = \langle z^*, x^*y^* \rangle = \overline{\langle x^*y^*, z^* \rangle} = \overline{\langle y^*, xz^* \rangle} = \langle xz^*, y^* \rangle = \langle y, zx^* \rangle,$$

so

$$\langle yx, z \rangle = \langle y, zx^* \rangle. \tag{2'}$$

By (1) and (U), for every $x \in A$, we also have

$$\langle yx, yx \rangle = \langle x^*y^*, x^*y^* \rangle \leq M_{x^*} \langle y^*, y^* \rangle = M_{x*} \langle y, y \rangle,$$

namely

$$\langle yx, yx \rangle \leq M_{x^*} \langle y, y \rangle, \quad \text{for all } y \in A. \tag{U'}$$

The inequality (U) says that for x fixed, the linear map $y \mapsto xy$ is continuous, and the inequality (U') says that for y fixed, the linear map $x \mapsto xy$ is continuous.

Beware that in general, this *does not imply* that the map $(x, y) \mapsto xy$ is continuous. Thus, in general, a Hilbert algebra is *not* a normed algebra in the sense of Definition 9.4, because in a normed algebra, the map $(x, y) \mapsto xy$ is continuous. However, if the Hilbert algebra A is *complete*, then using Baire's theorem, it can be shown that the map $(x, y) \mapsto xy$ is continuous; see Dieudonné [24] (Chapter XII, Section 16, Problem 8(c)). As a consequence, we can show that a complete Hilbert algebra is normable, and thus a Banach algebra.

Proposition 11.15. *If A is a complete Hilbert algebra, then there is a norm $\| \|_b$ equivalent to the norm $x \mapsto \|x\| = \sqrt{\langle x, x\rangle}$ induced by the inner product $\langle -, -\rangle$ on A, such that A is a Banach algebra with the norm $\| \|_b$.*

Proof. First, assume that A is not unital. Since A is complete, as we said earlier, the bilinear map $(x, y) \mapsto xy$ is continuous, so there is some constant $c > 0$ such that $\|xy\| \leq c\|x\| \|y\|$ for all $x, y \in A$. Since $c\|xy\| \leq c\|x\| c\|y\|$, if we let $\|x\|_b = c\|x\|$, we obtain a norm equivalent to $\| \|$ such that

$$\|xy\|_b \leq \|x\|_b \|y\|_b,$$

and with this norm, A is a Banach algebra.

If A has a multiplicative unit e, then the norm $\| \|_b$ also needs to satisfy the condition $\|e\|_b = 1$, so we need a different construction. For every $x \in A$, let $L_x : A \to A$ be the linear map given by

$$L_x(y) = xy, \qquad y \in A.$$

Since the bilinear map $(x, y) \mapsto xy$ is continuous, the linear map L_x is continuous. We check immediately that

$$\begin{aligned} L_{x+y} &= L_x + L_y \\ L_{\alpha x} &= \alpha L_x \\ L_{xy} &= L_x \circ L_y. \end{aligned}$$

Therefore, the map $L: A \to \mathcal{L}(A)$ given by $x \mapsto L_x$ is an algebra homomorphism. The homomorphism L is injective, because if $L_x = 0$, then $L_x(e) = xe = x = 0$. We claim that there is a constant $c > 0$ such that $\|L_x\| \leq c\|x\|$. Recall that for the operator norm $\|L_x\|$, we have

$$\|L_x\| = \sup\{\|L_x(y)\| \mid \|y\| = 1\} = \sup\{\|xy\| \mid \|y\| = 1\},$$

and since the bilinear map $(x, y) \mapsto xy$ is continuous

$$\|L_x\| = \sup\{\|xy\| \mid \|y\| = 1\} \leq \sup\{c\|x\| \|y\| \mid \|y\| = 1\} = c\|x\|.$$

On the other hand, since $\|L_x\|$ is the operator norm,

$$\|x\| = \|xe\| = \|L_x(e)\| \leq \|L_x\| \|e\|.$$

Therefore, if we let $\|x\|_b = \|L_x\|$, we obtain a norm on A equivalent to the norm $\| \|$, and A is a normed algebra with the norm $\| \|_b$, since $\|L_{xy}\| = \|L_x \circ L_y\| \leq \|L_x\| \|L_y\|$. ☐

Remark: The proof of Proposition 11.15 shows that if A is an algebra whose topology is defined by a norm $\| \|$ and if the bilinear map $(x, y) \mapsto xy$ is continuous, then there is a norm $\| \|_b$ equivalent to the norm $\| \|$, such that A is a normed algebra with the norm $\| \|_b$.

The following result will be needed to prove that if A is a Hilbert algebra and if $x \in A$ with $x \neq 0$, then $Ax \neq (0)$.

Proposition 11.16. *Let A be a Hilbert algebra. For every $x \in A$, if $x^*x = 0$, then $x = 0$.*

Proof. Since $x^*x = 0$, from Property (2) we have

$$\langle xy, xy\rangle = \langle x^*xy, y\rangle = 0,$$

hence $xy = 0$ for all $y \in A$ (since the hermitian product is positive definite). In particular, $xz^* = 0$. By (2'), we get

$$\langle x, yz\rangle = \langle xz^*, y\rangle = 0$$

for all $y, z \in A$, and since by (N) the subspace spanned by the set $\{yz \mid y, z \in A\}$ is dense in A, we conclude that $x = 0$ (since the hermitian product is positive definite). ☐

The following example is the first of two important instances of Hilbert algebras.

Example 11.8. Let H be a separable Hilbert space. Recall that this means that H has a countable Hilbert basis, that is, a countable orthonormal basis $(a_i)_{i\geq 1}$ such that the subspace spanned by $(a_i)_{i\geq 1}$ is dense in H (for every vector $x \in H$, there is some sequence (x_n), with x_n a linear combination $\sum_{k\in I_n} \lambda_k a_k$ where I_n a finite set, and x_n converges to x).

A linear map $u \in \mathcal{L}(H)$ is a *Hilbert–Schmidt operator* if the series $\sum_{n=1}^{\infty} \left\|u(a_n)\right\|^2$ converges, that is, $\sum_{n=1}^{\infty} \left\|u(a_n)\right\|^2 < \infty$. It can be shown that the quantity $\sum_{n=1}^{\infty} \left\|u(a_n)\right\|^2$ is independent of the Hilbert basis (a_n). The set of Hilbert–Schmidt operators is denoted by $\mathcal{L}_2(H)$. Then we define the map $u \mapsto \|u\|_{\mathrm{HS}}$ on the set $\mathcal{L}_2(H)$ by

$$\|u\|_{\mathrm{HS}}^2 = \sum_{n=1}^{\infty} \left\|u(a_n)\right\|^2.$$

It can be shown (using the Hilbert basis (a_n) and Parseval) that $\|u\|_{\mathrm{HS}} = \|u^*\|_{\mathrm{HS}}$. For any two Hilbert–Schmidt operators $u, v \in \mathcal{L}_2(H)$, it can also be shown that

$$\|u \circ v\|_{\mathrm{HS}} \leq \|u\|_{\infty} \|v\|_{\mathrm{HS}},$$

and if $u \in \mathcal{L}_2(H)$, then

$$\|u\|_{\infty} \leq \|u\|_{\mathrm{HS}}.$$

Then it can be shown that with the norm $\| \; \|_{\mathrm{HS}}$, the space $\mathcal{L}_2(H)$ of Hilbert–Schmidt operators is an involutive Banach algebra under composition, with the involution given by $u \mapsto u^*$ (where u^* is the adjoint of u). The space $\mathcal{L}_2(H)$ is a self-adjoint two-sided ideal in the involutive Banach algebra $\mathcal{L}(H)$, but in general, it is not closed in $\mathcal{L}(H)$.

The space $\mathcal{L}_2(H)$ contains the continuous linear maps of finite rank. The algebra $\mathcal{L}_2(H)$ is not unital because the identity map is not a Hilbert–Schmidt operator, and it is not a C^*-algebra.

If E and F are two normed vector spaces, a linear map $u\colon E \to F$ is a *compact operator* if the closure of $f(B)$ is compact for every bounded subset B of E. A compact operator

is continuous. Every Hilbert–Schmidt operator is a compact operator, but the converse is false. The above facts are proven in Dieudonné [24] (Chapter XV, Section 4).

For any two Hilbert–Schmidt operators $u, v \in \mathcal{L}_2(H)$, it can be shown that the quantity

$$g(u, v) = \sum_{n \geq 1} \langle u(a_n), v(a_n) \rangle$$

is defined and independent of the Hilbert basis (a_n). Then it can be shown that g is a hermitian inner product which is a bitrace such that $g(u, u) = \|u\|_{\mathrm{HS}}$; see Dieudonné [24] (Chapter XV, Section 7).

If H is a finite-dimensional Hilbert space of dimension n, then for every linear map $u \colon H \to H$, for every orthonormal basis $(a_1, \ldots, a_n)$, the quantity $\|u\|_{\mathrm{HS}}^2 = \sum_{i=1}^n \|u(a_n)\|^2$ is defined. Since $\|u(a_i)\|^2 = \langle u(a_i), u(a_i) \rangle = \langle (u^* \circ u)(a_i), a_i \rangle$, we have

$$\|u\|_{\mathrm{HS}}^2 = \sum_{i=1}^n \langle (u^* \circ u)(a_i), a_i \rangle = \mathrm{tr}(u^* \circ u) = \mathrm{tr}(u \circ u^*),$$

which is just the *Frobenius norm* (also called *Hilbert–Schmidt norm*). Then

$$g(u, v) = \sum_{i=1}^n \langle u(a_n), v(a_n) \rangle = \sum_{i=1}^n \langle (v^* \circ u)(a_n), a_n \rangle = \mathrm{tr}(v^* \circ u)$$

is the corresponding inner product, denoted by $\langle u, v \rangle_{\mathrm{HS}}$.

If $\mathcal{L}_2(H)$ is the involutive algebra of Hilbert–Schmidt operator of Example 11.8, then it can be shown that the bitrace g satisfies the properties (U) and (N). Consequently, $\mathcal{L}_2(H)$ *is a Hilbert algebra*; see Dieudonné [24] (Chapter XV, Section 7).

Recall from Proposition A.47 that if a topological space is metrizable and compact, then it is separable (which means that it contains a countable dense subset). Here is our second most important example of a Hilbert algebra.

Proposition 11.17. *If G is a compact metrizable group, then the involutive (complex) Banach algebra $\mathrm{L}^2(G)$ is a separable Hilbert algebra.*

Proof sketch. Proposition 11.17 is proven in Dieudonné [24] (Chapter XXI, Section 2). We may assume that G is equipped with a Haar measure λ such that $\lambda(G) = 1$. Condition (1) follows from the definition of the inner product on $\mathrm{L}^2(G)$. Since G is compact, $\mathcal{C}_0(G; \mathbb{C}) = \mathcal{K}_{\mathbb{C}}(G) \subseteq \mathrm{L}^2(G)$. By Proposition 8.49, $f * g \in \mathcal{C}_0(G; \mathbb{C}) \subseteq \mathrm{L}^2(G)$, and $\|f * g\|_\infty \leq \|f\|_2 \|g\|_2$ (where $\| \, \|_2$ is the L^2 semi-norm on $\mathrm{L}^2(G)$). By Proposition 5.24(2) and since $\lambda(G) = 1$, we also have $\|f * g\|_2 \leq \|f * g\|_\infty$. Consequently, $\|f * g\|_2 \leq \|f\|_2 \|g\|_2$, so Condition (U) follows. Condition (N) is a corollary of regularization (see Section 8.14). Finally, Condition (2), namely

$$\langle f * g, h \rangle = \langle g, f^* * h \rangle$$

is easily shown if f is real and continuous. In this case, $f^* = \check{f}$. We obtain the formula in general by continuity and using the fact that $\mathcal{K}_{\mathbb{R}}(G)$ is dense in $\mathrm{L}^2(G)$. $\square$

In the next sections we consider the special case in which a Hilbert algebra is complete.

11.5 Complete Separable Hilbert Algebras

We now consider the case where the Hilbert algebra A is complete, which means that it is a Hilbert space. Although this is not obvious, as said in the previous section, it can be shown that the map $(x, y) \mapsto xy$ from $A \times A$ to A is continuous. Consequently, by Proposition 11.15, a complete Hilbert algebra is normable, and thus a Banach algebra.

Our main goal is to show that every separable complete Hilbert algebra is the (countable) Hilbert sum of two-sided ideals $\mathfrak{a}_k$, where each ideal $\mathfrak{a}_k$ is the (countable) Hilbert sum of minimal left ideals all isomorphic to a common left ideal (Theorem 11.31). This is a beautiful and powerful result which is one of the main steps in proving the Peter–Weyl theorem.

Following Dieudonné, we only deal with the case where the Hilbert algebras A are separable, because then only countable Hilbert sums are needed. This implies that when we consider the Hilbert algebra $A = \mathrm{L}^2(G)$, the group G is metrizable and compact. This is not a serious restriction, because every Lie group, being a secound-countable manifold, is metrizable. The reader who wishes to see an exposition of the Peter–Weyl theorem in the general case of an arbitrary compact group is invited to consult Folland [33] (Chapter 5). Hilbert sums indexed by arbitrary (possibly uncountable) index sets arise.

We are led to the study of minimal left ideals and to irreducible self-adjoint idempotents which generate them. Recall that an element e of an algebra A is *idempotent* if $e^2 = e$, *self-adjoint* if $e = e^*$.[2]

Complete proofs are provided in Dieudonné [24] (Chapter XV, Section 8). There are many tedious technical details so to make it easier on the reader, we decided to only state most results (except the most important ones) without proof.

First observe that every closed self-adjoint subalgebra B (which means that $B^* = B$) of a complete Hilbert algebra A is a complete Hilbert algebra. Certain representations play a crucial role.

Proposition 11.18. *Let A be a complete Hilbert algebra. For any closed left ideal $\mathfrak{b}$ in A, let $U_{\mathfrak{b}}$ be the map from A to $\mathcal{L}(\mathfrak{b})$ given by*

$$U_{\mathfrak{b}}(x)(y) = xy, \qquad x \in A,\ y \in \mathfrak{b}.$$

Observe that $U_{\mathfrak{b}}(x)$ is left multiplication by $x \in A$. Then $U_{\mathfrak{b}}$ is a representation of A in $\mathfrak{b}$.

Proof. For every $x \in A$, the linear map $U_{\mathfrak{b}}(x)$ from $\mathfrak{b}$ to itself is continuous because of Condition (U). The verification that $U_{\mathfrak{b}}(x_1 x_2) = U_{\mathfrak{b}}(x_1) \circ U_{\mathfrak{b}}(x_2)$ is immediate. For all

[2]Here e does not denote the unit of A if A is unital. To avoid confusion, we denote the unit of A by $\mathbf{1}$ if A is unital.

$y, z \in \mathfrak{b}$, by Condition (2), we have

$$\begin{aligned}\langle U_{\mathfrak{b}}(x)^*(y), z\rangle &= \langle y, U_{\mathfrak{b}}(x)(z)\rangle \\ &= \langle y, xz\rangle \\ &= \langle x^*y, z\rangle \\ &= \langle U_{\mathfrak{b}}(x^*)(y), z\rangle,\end{aligned}$$

which implies that $U_{\mathfrak{b}}(x)^* = U_{\mathfrak{b}}(x^*)$. If A has the unit element $\mathbf{1}$, then $U_{\mathfrak{b}}(\mathbf{1})$ is the identity transformation. $\square$

Definition 11.15. Let A be a complete Hilbert algebra. For any closed left ideal $\mathfrak{b}$ in A, let $U_{\mathfrak{b}}\colon A \to \mathcal{L}(\mathfrak{b})$ be the representation given by

$$U_{\mathfrak{b}}(x)(y) = xy, \qquad x \in A,\ y \in \mathfrak{b}.$$

The representation $U_A\colon A \to \mathcal{L}(A)$ is called the *regular representation* of A.

When $\mathfrak{b} = A$, we usually write $U(x)$ instead of $U_A(x)$. By Proposition 11.16, the representation U_A is faithful. Moreover, since the map $(x, y) \mapsto xy$ from $A \times A$ to A is continuous, the map $x \mapsto U_A(x)$ from A to $\mathcal{L}(A)$ is also continuous.

Let A be a complete Hilbert algebra. For each left ideal $\mathfrak{b}$ of A, the set $\mathfrak{b}^* = \{x^* \mid x \in \mathfrak{b}\}$ is a right ideal in A. We now start a fairly long chain of definitions and results leading to our main result (Theorem 11.31).

In the sequel A denotes a complete Hilbert algebra. Two keys concepts are the notion of an irreducible self-adjoint idempotent and of a minimal left ideal.

Proposition 11.19. *For every left ideal $\mathfrak{l}$ of A, the orthogonal complement $\overline{\mathfrak{l}}^{\perp}$ of the closure of $\mathfrak{l}$ is a left ideal.*

Proposition 11.19 is proven in Dieudonné [24] (Chapter XV, Proposition 15.8.2). The next proposition gives useful properties of idempotents.

Proposition 11.20. *Let $e \neq 0$ be an idempotent element in A ($e^2 = e$). Then the following properties hold:*

(1) $\|e\| \geq 1$;

(2) e^* is idempotent;

(3) The left ideal Ae is equal to the set $\{x \in A \mid x = xe\}$, and is closed in A.

Proposition 11.20 is proven in Dieudonné [24] (Chapter XV, Proposition 15.8.3). The next proposition gives orthogonality properties of self-adjoint idempotents.

Proposition 11.21. *If e_1 and e_2 are self-adjoint idempotents in A, then the following properties are equivalent:*

(a) $\langle e_1, e_2 \rangle = 0$;

(b) $e_1 e_2 = 0$;

(c) $e_2 e_1 = 0$.

The proof of Proposition 11.21 makes use of Property (2) and Property (2') of Definition 11.14; see Dieudonné [24] (Chapter XV, Proposition 15.8.4).

The following proposition shows that there are plenty nonzero self-adjoint idempotents. Since Proposition 11.22 establishes a very important fact we supply a proof.

Proposition 11.22. *Every left ideal $\mathfrak{l} \neq (0)$ in A contains a nonzero self-adjoint idempotent.*

Proof. Let x be any nonzero element in $\mathfrak{l}$. By Proposition 11.16, we have $x^*x \neq 0$. Let $z = x^*x$. Then z is a self-adjoint element of $\mathfrak{l}$, but in general it is not idempotent. Consider the representation U_A. By rescaling z we may assume that $\|U_A(z)\| = 1$. Since $z = z^*$, the linear map $U_A(z)$ is self-adjoint, and we have $U_A(z^2) = U_A(z) \circ U_A(z) = (U_A(z))^2$. By Proposition D.11, we have

$$\left\|U_A(z^2)\right\| = \left\|(U_A(z))^2\right\| = \left\|U_A(z)\right\|^2 = 1. \tag{$*_1$}$$

We claim that the sequence (z^{2k}) is a Cauchy sequence whose limit e is a nonzero self-adjoint idempotent in $\mathfrak{l}$.

We are led to investigate bounds on $\|z^{2m} - z^{2n}\|$, with $m = n + p$, $n, p > 0$ and the proof uses the following steps.

Step 1. First we prove that $\left\|U_A(z^k)\right\| = 1$ for all $k \geq 1$.

From Equation $(*_1)$, by induction we obtain

$$\left\|U_A(z^{2k})\right\| = 1, \quad k \geq 1. \tag{$*_2$}$$

On the other hand, since $\|U_A(z)\| = 1$, we have

$$\left\|U_A(z^{k+1})\right\| = \left\|U_A(z) \circ U_A(z^k)\right\| \leq \|U_A(z)\| \left\|U_A(z^k)\right\| = \left\|U_A(z^k)\right\|.$$

Thus the sequence $(\left\|U_A(z^k)\right\|)$ is nonincreasing, and since it contains infinitely many terms (of the form $\|U_A(z^{2i})\|$) equal to 1, we must indeed have

$$\left\|U_A(z^k)\right\| = 1, \quad \text{for all } k \geq 1. \tag{$*_3$}$$

Since we are using the operator norm, we deduce that

$$1 = \left\|U_A(z^k)\right\| \leq \|U_A\| \left\|z^k\right\|,$$

which implies that

$$\left\|z^k\right\| \geq 1/\left\|U_A\right\|, \quad k \geq 1. \tag{$*_4$}$$

Here, $\|U_A\|$ is the operator norm of U_A as a continuous linear map from A to $\mathcal{L}(A)$.

Step 2. Next we show that the sequence $(\left\|z^{2k}\right\|^2)$ is nonincreasing, and since it is bounded from below by $1/\left\|U_A\right\|^2$, it has a limit $a > 0$.

Since z is self-adjoint and $U_A(x)(y) = xy$, using Property (2) of Definition 11.14, we have (recall that $m = n + p$)

$$\begin{aligned}\langle z^{2m}, z^{2n}\rangle &= \langle z^p z^{p+2n}, z^{2n}\rangle = \langle z^{p+2n}, z^{p+2n}\rangle = \left\|U_A(z^p)(z^{2n})\right\|^2 \\ &\leq \left\|U_A(z^p)\right\|^2 \left\|z^{2n}\right\|^2 = \left\|z^{2n}\right\|^2 = \langle z^{2n}, z^{2n}\rangle,\end{aligned}$$

and

$$\begin{aligned}\langle z^{2m}, z^{2m}\rangle &= \left\|U(z^p)(z^{p+2n})\right\|^2 \leq \left\|U(z^p)\right\|^2 \left\|z^{p+2n}\right\|^2 = \left\|z^{p+2n}\right\|^2 \\ &= \langle z^{p+2n}, z^{p+2n}\rangle = \langle z^{2m}, z^{2n}\rangle,\end{aligned}$$

so that for all $m > n$, we have

$$1/\left\|U_A\right\|^2 \leq \langle z^{2m}, z^{2m}\rangle \leq \langle z^{2m}, z^{2n}\rangle \leq \langle z^{2n}, z^{2n}\rangle. \tag{$*_5$}$$

Then $(*_5)$ shows that the sequence $(\left\|z^{2k}\right\|^2)$ is nonincreasing, and since it is bounded from below by $1/\left\|U_A\right\|^2$, it has a limit $a > 0$.

Step 3. Using $(*_5)$, we also have

$$\begin{aligned}\left\|z^{2m} - z^{2n}\right\|^2 &= \langle z^{2m}, z^{2m}\rangle - 2\langle z^{2m}, z^{2n}\rangle + \langle z^{2n}, z^{2n}\rangle \\ &\leq \langle z^{2n}, z^{2n}\rangle - \langle z^{2m}, z^{2n}\rangle \\ &\leq \left\|z^{2n}\right\|^2 - a.\end{aligned}$$

Since the sequence $(\left\|z^{2k}\right\|^2)$ has limit a, the sequence (z^{2k}) is a Cauchy sequence, as asserted earlier.

Step 4. Let e be the limit of the Cauchy sequence (z^{2k}). By continuity,

$$e^2 = \lim_{k\mapsto\infty} z^{4k} = e,$$

and

$$e^* = \lim_{k\mapsto\infty} (z^*)^{4k} = \lim_{k\mapsto\infty} z^{4k} = e.$$

We also have

$$ez^2 = \lim_{k\mapsto\infty} z^{2k+2} = e,$$

and since $z \in \mathfrak{l}$ and $\mathfrak{l}$ is a left ideal, we see that $e \in \mathfrak{l}$. Finally, since by $(*_4)$ we have

$$\left\|z^k\right\| \geq 1/\left\|U_A\right\|, \quad k \geq 1,$$

and we deduce that $\|e\| > 0$, so $e \neq 0$. $\square$

Definition 11.16. A self-adjoint idempotent $e \neq 0$ is *reducible* if there exist two orthogonal nonzero self-adjoint idempotents e_1 and e_2 such that $e = e_1 + e_2$. If a self-adjoint idempotent $e \neq 0$ is not reducible, we say that it is *irreducible*.

By Proposition 11.21, if $e = e_1 + e_2$ is reducible, then $ee_1 = e_1 e = e_1$ and $ee_2 = e_2 e = e_2$.

Proposition 11.23 shows that irreducible self-adjoint idempotents are the building blocks for self-adjoint idempotents.

Proposition 11.23. *Every self-adjoint idempotent $e \neq 0$ is the sum of a finite number of irreducible self-adjoint idempotents in Ae. Every left ideal $\mathfrak{l} \neq (0)$ contains an irreducible self-adjoint idempotent.*

Proof. The second statement follows from Proposition 11.22 and the first statement. Here is the proof of the first statement.

Let $e \neq 0$ be a self adjoint idempotent. Since $ee = e$, obviously $e \in Ae$. We claim that if $\|e\|^2 < 2$, then e is irreducible. Otherwise, $e = e_1 + e_2$ for two orthogonal nonzero self-adjoint idempotents e_1 and e_2, so

$$\|e\|^2 = \|e_1\|^2 + \|e_2\|^2 .$$

By Proposition 11.20, since $e_1, e_2 \neq 0$, we have $\|e_1\|, \|e_2\| \geq 1$, so $\|e\|^2 \geq 2$, a contradiction.

If $\|e\|^2 \geq 2$, we prove by complete induction on the smallest natural number n such that $\|e\|^2 < n$ that e is is the sum of a finite number of irreducible self-adjoint idempotents in Ae.

If e is reducible, then $e = e_1 + e_2$, where $e_1, e_2 \neq 0$ are orthogonal self-adjoint idempotents. By Proposition 11.21, we have $e_1 = e_1 e$ and $e_2 = e_2 e$, which implies that $e_1, e_2 \in Ae$. Since $\|e\|^2 = \|e_1\|^2 + \|e_2\|^2$, we have

$$\|e_1\|^2 = \|e\|^2 - \|e_2\|^2 \leq \|e\|^2 - 1 < n - 1$$

and similarly

$$\|e_2\|^2 < n - 1.$$

Therefore we can apply the induction hypothesis to e_1 and e_2, and this finishes the proof. $\square$

Definition 11.17. A left ideal $\mathfrak{l}$ is *minimal* if $\mathfrak{l} \neq (0)$ and if there exists no nonzero left ideal $\mathfrak{l}' \neq \mathfrak{l}$ such that $\mathfrak{l}' \subseteq \mathfrak{l}$. A similar definition applies to minimal right ideals.

Here is the first significant result which gives the structure of minimal left ideals in terms of irreducible self-adjoint idempotents.

Theorem 11.24. *A left ideal $\mathfrak{l}$ in A is minimal if and only if it is of the form $\mathfrak{l} = Ae$, where $e \neq 0$ and e is an irreducible self-adjoint idempotent.*

Proof. First assume that $\mathfrak{l}$ is a minimal left ideal. By Proposition 11.23, the ideal $\mathfrak{l}$ contains an irreducible self-adjoint idempotent $e \neq 0$. Since $e \in \mathfrak{l}$ and $\mathfrak{l}$ is a left ideal, we have $Ae \subseteq \mathfrak{l}$ and $e = e^2 \in Ae$. This shows that Ae is a nonzero ideal contained in the minimal ideal $\mathfrak{l}$, which implies that $\mathfrak{l} = Ae$.

Conversely, let $e \neq 0$ be an irreducible self-adjoint idempotent. We need to prove that $\mathfrak{l} = Ae$ is a minimal ideal.

Suppose by contradiction that $\mathfrak{l}$ contains a left ideal $\mathfrak{l}' \neq (0)$ such that $\mathfrak{l}' \neq \mathfrak{l}$. By Proposition 11.22, there is some self-adjoint idempotent $e' \neq 0$ that belongs to $\mathfrak{l}'$.

If we let $e_1 = e - ee'$, and $e_2 = ee'$, then $e = e_1 + e_2$, and we claim that e_1 and e_2 are orthogonal nonzero self-adjoint idempotents. If so, this contradicts the fact that e is irreducible, and finishes the proof by contradiction.

Since $e' \in Ae$ and $ee = e$, we have

$$e' = e'e.$$

Consequently,

$$\begin{aligned} e_2^2 &= ee'ee' = ee'e' = ee' = e_2 \\ ee_2 &= eee' = ee' = e_2 \\ e_2e &= ee'e = ee' = e_2, \end{aligned}$$

and thus

$$\begin{aligned} e_1e_2 &= (e - e_2)e_2 = ee_2 - e_2^2 = e_2 - e_2 = 0 \\ e_2e_1 &= e_2(e - e_2) = e_2e - e_2^2 = e_2 - e_2 = 0 \\ e_1^2 &= (e - e_2)^2 = e^2 - ee_2 - e_2e + e_2^2 = e - e_2 - e_2 + e_2 = e - e_2 = e_1. \end{aligned}$$

We also have

$$e_2^* = (ee'e)^* = e^*(e')^*e^* = ee'e = e_2$$

and so

$$e_1^* = (e - e_2)^* = e^* - e_2^* = e - e_2 = e_1.$$

In summary, we proved that e_1, e_2 are orthogonal self-adjoint idempotents. It remains to prove that $e_1 \neq 0$ and $e_2 \neq 0$.

Since $e = e_1 + e_2 = e_1 + ee'$, if $e_1 = 0$, then $e = ee' \in \mathfrak{l}'$ (since $e' \in \mathfrak{l}'$), hence $\mathfrak{l}' = \mathfrak{l}$, contradicting the hypothesis that $\mathfrak{l}' \neq \mathfrak{l}$. Since $e_2 = ee'$, we have

$$e'e_2 = e'ee' = e'e' = e' \neq 0,$$

which implies that $e_2 \neq 0$. Finally we showed that $e = e_1 + e_2$ is a reducible decomposition of e, contradicting the hypothesis that e is irreducible. $\square$

Example 11.9. Consider the algebra $A = \mathrm{M}_n(\mathbb{C})$ from Example 11.1 consisting of $n \times n$ complex matrices with involution $X \mapsto X^*$ (the conjugate transpose). It is immediately verified that the map

$$\langle X, Y\rangle \mapsto \langle X, Y\rangle = \mathrm{tr}(Y^* X)$$

is a Hermitian inner product on A which makes A into a complete Hilbert algebra, which is obviously separable. We immediately check that the $n \times n$ matrices E_i defined such that $(E_i)_{jk} = 1$ iff $j = k = i$, else $(E_i)_{jk} = 0$, are irreducible self-adjoint idempotents in A. Then the subspaces $\mathfrak{l}_i = AE_i$ $(1 \leq i \leq n)$ are minimal left ideals in A. Observe that $\mathfrak{l}_i = AE_i$ consists of those $n \times n$ matrices whose columns of index $1, \ldots, i-1, i+1, \ldots, n$ are zero columns. Observe that

$$A = \bigoplus_{j=1}^{n} \mathfrak{l}_j = \bigoplus_{j=1}^{n} AE_j.$$

Also note that

$$\mathfrak{l}_j = \bigoplus_{k=1}^{n} E_k A E_j,$$

where $E_k A E_j$ is the one-dimensional subspace of A consisting of the $n \times n$ matrices whose only nonzero entry, if any, is the entry of index (k, j). This example is an illustration of Theorem 11.32.

Theorem 11.24 and Proposition 11.21 immediately imply the following second significant result which shows that the minimal left ideals are the building blocks of left ideals.

Theorem 11.25. *Every left ideal in A contains a minimal left ideal. Every minimal left ideal is closed.*

The next two results are technical lemmas that are needed to prove Theorem 11.28. They build special kinds of orthogonal systems from self-adjoint idempotents.

Proposition 11.26. *The following properties hold.*

(1) *If e and e' are two orthogonal self-adjoint idempotents, then the left ideals Ae and Ae' are orthogonal.*

(2) *Let $(e_i)_{1 \leq i \leq n}$ be a finite family of pairwise orthogonal, self-adjoint idempotents. Then for every $x \in A$, the element $x - \sum_{i=1}^{n} x e_i$ is orthogonal to each Ae_j $(1 \leq j \leq n)$.*

Proposition 11.26 is proven in Dieudonné [24] (Chapter XV, Proposition 15.8.9). The proof is quite simple. On the other hand, the proof of Proposition 11.27 makes use of a fairly complicated inductive construction which we omit. See Dieudonné [24] (Chapter XV, Proposition 15.8.10) for complete details.

Proposition 11.27. *For every $x \in A$ there exists a finite or countably infinite sequence (e_n) of pairwise orthogonal irreducible self-adjoint idempotents belonging to the closure $\mathfrak{l}$ of the ideal Ax, such that $x = \sum_n x e_n$ (this series being convergent in A), and $\|x\|^2 = \sum_n \|x e_n\|^2$.*

Form now on *we assume that the complete Hilbert algebra A is separable*. We have the following important result.

Theorem 11.28. *Suppose A is separable. Then every closed left ideal $\mathfrak{b}$ is the Hilbert sum of a finite or countably infinite sequence of minimal left ideals $\mathfrak{l}_n = Ae_n$, where e_n is an irreducible self-adjoint idempotent. For every $x \in \mathfrak{b}$, we have $x = \sum_n xe_n$, and for all $x, y \in \mathfrak{b}$ we have $\langle x, y\rangle = \sum_n \langle xe_n, ye_n\rangle$.*

Proof. We follow Dieudonné's proof from [24] (Chapter XV, Section 8, Theorem 15.8.11). By Proposition 11.26 and by the properties of Hilbert sums (see Proposition 11.3), the vector xe_n is the orthogonal projection of x onto Ae_n, so the second and the third assertions follow.

To prove the first assertion, let $(x_n)_{n\geq 1}$ be a dense sequence in $\mathfrak{b}$, which exists since A is separable. We define inductively, for each n, a finite or countably infinite sequence $(e_{n,i})_{i\in I_n}$ of irreducible self-adjoint idempotents as follows. For $n = 1$, by Proposition 11.27 applied to $x_1 \in \mathfrak{b}$, since $\overline{Ax_1} \subseteq \mathfrak{b}$, there is a finite or countably infinite sequence $(e_{1,i})_{i\in I_1}$ of pairwise orthogonal irreducible self-adjoint idempotents belonging to $\mathfrak{b}$ such that $x_1 = \sum_{i\in I_1} x_1 e_{1,i}$. Suppose that the $e_{m,i}$ have been defined for all $m \leq n$ in such a way that they are pairwise orthogonal and belong to $\mathfrak{b}$, and are such that the x_m with $m \leq n$ belong to the closure $\mathfrak{a}_n \subseteq \mathfrak{b}$ of the left ideal which is the sum of the ideals $Ae_{m,i}$ for all $m \leq n$ and all $i \in I_m$,

$$\mathfrak{a}_n = \overline{\bigoplus_{m\leq n,\, i\in I_m} Ae_{m,i}}.$$

Let x'_{n+1} be the orthogonal projection of x_{n+1} onto $\mathfrak{a}_n^{\perp} \cap \mathfrak{b}$. Using Proposition 11.27 applied to $x'_{n+1} \in \mathfrak{a}_n^{\perp} \cap \mathfrak{b}$, since $\overline{Ax'_{n+1}} \subseteq \mathfrak{a}_n^{\perp} \cap \mathfrak{b}$, there is a finite or countably infinite sequence $(e_{n+1,i})_{i\in I_{n+1}}$ of pairwise orthogonal irreducible self-adjoint idempotents which belong to $\mathfrak{a}_n^{\perp} \cap \mathfrak{b}$ and are such that $x'_{n+1} = \sum_{i\in I_{n+1}} x'_{n+1} e_{n+1,i}$; see Figure 11.1. Since each family $(e_{n,i})_{i\in I_n}$ is countable and since there are countably many of these families, the union of these families is also countable, so it can be listed as a single sequence. Using Proposition 11.26 and the properties of Hilbert sums, we leave it as an exercise to check that this sequence has the desired properties. $\square$

Theorem 11.28 applies in particular when $\mathfrak{b} = A$. In this case we get a decomposition of A as a Hilbert sum of minimal left ideals. However, in general, there are infinitely many such decompositions. More precisely we have the following result.

Proposition 11.29. *Suppose A is separable, and let $\mathfrak{l}$ be a minimal left ideal of A. Then there exists a decomposition of A as a Hilbert sum of minimal left ideals $\mathfrak{l}_n$ such that $\mathfrak{l}_1 = \mathfrak{l}$.*

The following technical result is needed in the proof of the main theorem (Theorem 11.31).

Theorem 11.30. *Let e and e' be two irreducible self-adjoint idempotents, and let $\mathfrak{l} = Ae$, and $\mathfrak{l}' = Ae'$ be the corresponding minimal left ideals. The following properties hold.*

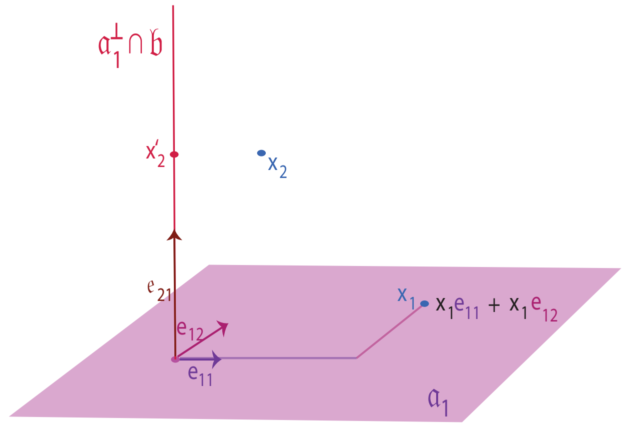


Figure 11.1: A schematic illustration of the construction used in the proof of Theorem 11.28. Let $\mathfrak{a}_1 = \overline{Ae_{1,1} \oplus Ae_{1,2}}$ be the horizontal purple plane. The vertical red line is $\mathfrak{a}_1^{\perp} \cap \mathfrak{b}$. Then $\mathfrak{a}_2 = \overline{Ae_{1,1} \oplus Ae_{1,2} \oplus Ae_{2,1}}$.

(1) Every homomorphism of the A-module $\mathfrak{l}$ into the A-module $\mathfrak{l}'$ is a map f_a of the form $f_a(x) = xa$ $(x \in \mathfrak{l})$, for some $a \in eAe' = eA \cap Ae'$; it is either zero or bijective, and the map $a \mapsto f_a$ is an isomorphism of the complex vector space eAe' onto $\mathrm{Hom}_A(\mathfrak{l}, \mathfrak{l}')$ such that $f_{ab} = f_b \circ f_a$.

(2) The $\mathbb{C}$-algebra eAe, isomorphic to $\mathrm{End}_A(\mathfrak{l}) = \mathrm{Hom}_A(\mathfrak{l}, \mathfrak{l})$ (the space of endomorphisms of $\mathfrak{l}$), is a field equal to $\mathbb{C}e$ (and therefore isomorphic to $\mathbb{C}$).

(3) If $\mathfrak{l}$ and $\mathfrak{l}'$ are not isomorphic as A-modules, then e and e' (and consequently $\mathfrak{l}$ and $\mathfrak{l}'$) are orthogonal, and $\mathfrak{l}\mathfrak{l}' = \mathfrak{l}'\mathfrak{l} = (0)$. If $\mathfrak{l}$ and $\mathfrak{l}'$ are isomorphic as A-modules, then eAe' is a complex vector space of dimension 1, and $\mathfrak{l}\mathfrak{l}' = \mathfrak{l}'$.

(4) If $x \in A$, then $\mathfrak{l}x$ is a left ideal which is either (0) or isomorphic (as A-module) to $\mathfrak{l}$.

Proof. We follow Dieudonné's proof from [24] (Chapter XV, Section 8, Theorem 15.8.12). (1) Let $g\colon \mathfrak{l} \to \mathfrak{l}'$ be an A-module homomorphism, and let $a = g(e)$. For every $x \in \mathfrak{l} = Ae$, since $e^2 = e$, we have

$$g(x) = g(xe) = xg(e) = xa,$$

so $g = f_a$. Since $a \in g(\mathfrak{l}) \subseteq \mathfrak{l}'$, $\mathfrak{l}' = Ae'$ and $(e')^2 = e'$, we have $a = ae'$. On the other hand,

$$a = g(e) = g(e^2) = eg(e) = ea,$$

so $a = eae' \in eAe'$.

By definition of eAe', we have $eAe' \subseteq eA \cap Ae'$. Conversely, if $y \in eA \cap Ae'$, as $e^2 = e$ and $(e')^2 = e'$, we have $y = ye'$ and $y = ey$, so that $y = eye' \in eAe'$. Therefore, $eAe' = eA \cap Ae'$.

The image $g(\mathfrak{l})$ of g is a left ideal contained in $\mathfrak{l}'$, and since $\mathfrak{l}'$ is a minimal left ideal, either $g(\mathfrak{l}) = (0)$ or $g(\mathfrak{l}) = \mathfrak{l}'$. Likewise, the kernel $\mathrm{Ker}\, g$ of g is a left ideal contained in $\mathfrak{l}$, and since $\mathfrak{l}$ is a minimal left ideal, either $\mathrm{Ker}\, g = (0)$ or $\mathrm{Ker}\, g = \mathfrak{l}$. If $\mathrm{Ker}\, g = \mathfrak{l}$, then $g(\mathfrak{l}) = (0)$ and g is the zero map. If $\mathrm{Ker}\, g = (0)$, then $g(\mathfrak{l}) \neq (0)$, so we must have $g(\mathfrak{l}) = \mathfrak{l}'$, and g is bijective.

If $g = f_a = 0$, then $f_a(e) = ea = 0$; but $a \in eAe'$, so $ea = a$, and consequently $a = 0$, which shows that the map $a \mapsto f_a$ is an isomorphism.

(2) By Proposition 11.20(3), the $\mathbb{C}$-algebra eAe is a closed subalgebra of A. In (1), we saw that every element of $\mathrm{End}_A(\mathfrak{l})$ is either zero or invertible, so $\mathrm{End}_A(\mathfrak{l})$ is a (possibly noncomutative) field, and since eAe isomorphic to $\mathrm{End}_A(\mathfrak{l})$, it is also a field. Clearly e is a unit in eAe, and since A is a complete Hilbert algebra, it is a Banach algebra, and since eAe is closed in A, it is also a Banach algebra. By the Gelfand–Mazur theorem (Theorem 9.14), $eAe = \mathbb{C}e \cong \mathbb{C}$.

(3) If $\mathfrak{l}$ and $\mathfrak{l}'$ are not isomorphic, by (1) we have $eAe' = (0)$, and in particular, $ee' = 0$. By Proposition 11.21, e and e' are orthogonal. By Proposition 11.26, the left ideals $\mathfrak{l}$ and $\mathfrak{l}'$ are orthogonal, and $\mathfrak{l}\mathfrak{l}' = AeAe' = A(0) = (0)$. Similarly $e'Ae = (0)$, so $\mathfrak{l}'\mathfrak{l} = (0)$. If $\mathfrak{l}$ and $\mathfrak{l}'$ are isomorphic, and if $g\colon \mathfrak{l} \to \mathfrak{l}'$ is an isomorphism from $\mathfrak{l}$ to $\mathfrak{l}'$, then every homomorphism $h\colon \mathfrak{l} \to \mathfrak{l}'$ is of the form $h = g \circ u$, where $u \in \mathrm{End}_A(\mathfrak{l})$, which means that $\mathrm{End}_A(\mathfrak{l})$ and $\mathrm{Hom}_A(\mathfrak{l}, \mathfrak{l}')$ are isomorphic. By (2), the space $\mathrm{End}_A(\mathfrak{l})$ is one-dimensional, so $\mathrm{Hom}_A(\mathfrak{l}, \mathfrak{l}')$ is also one-dimensional, and by (1), the space $\mathrm{Hom}_A(\mathfrak{l}, \mathfrak{l}')$ is isomorphic to eAe', so we deduce that eAe' is complex vector space of dimension 1. The ideal $\mathfrak{l}\mathfrak{l}'$ is obviously contained in $\mathfrak{l}'$, and contains $eAe' \neq (0)$. Since $\mathfrak{l}'$ is a minimal left ideal, we must have $\mathfrak{l}\mathfrak{l}' = \mathfrak{l}'$.

(4) Since $\mathfrak{l}x$ is the image of $\mathfrak{l}$ under the homomorphism $\varphi\colon \mathfrak{l} \to A$ given by $\varphi(y) = yx$, it is a left ideal isomorphic to $\mathfrak{l}/\mathrm{Ker}\,\varphi$. But since $\mathrm{Ker}\,\varphi$ is a left ideal contained in the minimal left ideal $\mathfrak{l}$, either $\mathrm{Ker}\,\varphi = (0)$ or $\mathrm{Ker}\,\varphi = \mathfrak{l}$, which means that φ is either injective or zero, and so $\mathfrak{l}x$ is a left ideal which is either (0) or isomorphic (as A-module) to $\mathfrak{l}$. $\square$

Finally, we come to the main theorems of this chapter.

11.6 The Structure of Complete Separable Hilbert Algebras

Theorem 11.31. *(Master decomposition theorem) Let A be a complete, separable Hilbert algebra. The following properties hold.*

(1) There exists a finite or countably infinite sequence $(\mathfrak{l}_k)_{k\in J}$ of minimal left ideals, no pair of which are isomorphic, such that every minimal left ideal of A is isomorphic (as an A-module) to some $\mathfrak{l}_k$.

(2) For each index $k \in J$, the closure of the sum of all the minimal left ideals of A which are isomorphic to $\mathfrak{l}_k$ is a self-adjoint two-sided ideal $\mathfrak{a}_k$. Every minimal left ideal of

the Hilbert algebra $\mathfrak{a}_k$ is a minimal left ideal of A, isomorphic to $\mathfrak{l}_k$, and the algebra $\mathfrak{a}_k$ contains no closed two-sided ideals other than (0) and $\mathfrak{a}_k$.

(3) Each of the algebras $\mathfrak{a}_k$ is a Hilbert sum

$$\mathfrak{a}_k = \bigoplus_{j \in I_k} \mathfrak{l}'_j,$$

with I_k finite or countably infinite, and where each $\mathfrak{l}'_j$ is a minimal left ideal isomorphic to $\mathfrak{l}_k$. The algebra A is the Hilbert sum

$$A = \bigoplus_{k \in J} \mathfrak{a}_k,$$

and $\mathfrak{a}_h \mathfrak{a}_k = (0)$ for all $h \neq k$.

Proof. We reproduce Dieudonné's proof from [24] (Chapter XV, Section 8, Theorem 15.8.13). By Theorem 11.28, we obtain A as the Hilbert sum

$$A = \bigoplus_{n \in L} \mathfrak{l}'_n$$

of minimal left ideals $\mathfrak{l}'_n$ of A, where L is finite or countably infinite. We define inductively the index set J and the sequence $(\mathfrak{l}_k)_{k \in J}$ of minimal left ideals $\mathfrak{l}_k$, no pair of which are isomorphic, as follows. Start with $\mathfrak{l}_1 = \mathfrak{l}'_1$. Having defined $\mathfrak{l}_1, \ldots, \mathfrak{l}_k$, let $\mathfrak{l}_{k+1}$ be the equal to $\mathfrak{l}'_m$, where m is the smallest integer such that $\mathfrak{l}'_m$ is not isomorphic to any of the ideals $\mathfrak{l}_1, \ldots, \mathfrak{l}_k$. If all the $\mathfrak{l}'_n$ are isomorphic to one of the ideals $\mathfrak{l}_1, \ldots, \mathfrak{l}_k$, then stop. Let J be the finite or countably infinite sequence of indices k so obtained, and for every $k \in J$, let I_k be the sequence of integers n such that $\mathfrak{l}'_n$ is isomorphic to $\mathfrak{l}_k$. If J is infinite, then each I_k is finite. Otherwise, $I_1, \ldots, I_{k-1}$ are finite, and I_k is finite or countably infinite. Define $\mathfrak{a}_k$ as the Hilbert sum

$$\mathfrak{a}_k = \bigoplus_{j \in I_k} \mathfrak{l}'_j.$$

By construction, it is clear that each $\mathfrak{a}_k$ is a left ideal and that H is the Hilbert sum

$$A = \bigoplus_{k \in J} \mathfrak{a}_k.$$

Let $\mathfrak{l}$ be any minimal left ideal in A. Then $\mathfrak{l}$ must be isomorphic to one of the $\mathfrak{l}_k$, for otherwise by Theorem 11.30(3), it would be orthogonal to all of the $\mathfrak{l}'_n$, and hence orthogonal to A itself, a contradiction. The same argument shows that $\mathfrak{l}$ is orthogonal to all the $\mathfrak{a}_h$ with $h \neq k$. Hence, since $\mathfrak{a}_k$ is the orthogonal complement of the Hilbert sum

$$\bigoplus_{h \in J - \{k\}} \mathfrak{a}_h,$$

we must have $\mathfrak{l} \subseteq \mathfrak{a}_k$. This implies that $\mathfrak{a}_k$ is the closure of the sum of all the minimal left ideals of A which are isomorphic to $\mathfrak{l}_k$, and therefore $\mathfrak{a}_k$ is independent of the decomposition of A as the Hilbert sum of the $\mathfrak{l}'_n$ from which we started. Moreover, for every $x \in A$, and every $n \in I_k$, by Theorem 11.30(4), we know that $\mathfrak{l}'_n x$ is a left ideal which is either (0) or isomorphic to $\mathfrak{l}'_n$, hence contained in $\mathfrak{a}_k$. This proves that $\mathfrak{a}_k$ is a two-sided ideal. If $\mathfrak{l}'_n = Ae'_n$, where e'_n is an irreducible self-adjoint idempotent, then $(\mathfrak{l}'_n)^* = e'_n A$, hence $\mathfrak{a}_k^* = \mathfrak{a}_k$.

Let $\mathfrak{l}''$ be a minimal left ideal of the Hilbert algebra $\mathfrak{a}_k$. By Theorem 11.24, we have $\mathfrak{l}'' = \mathfrak{a}_k e''$, where e'' is a self-adjoint idempotent, and $e'_n e''$ cannot vanish for all $n \in I_k$, otherwise $\mathfrak{l}''$ would be orthogonal to all of the $\mathfrak{l}'_n$ with $n \in I_k$, and therefore to the closure of their sum, namely to $\mathfrak{a}_k$, which is impossible because $\mathfrak{l}'' \neq (0)$. Hence there exists at least one index $n \in I_k$ such that $\mathfrak{l}'_n\mathfrak{l}'' \neq (0)$; since $\mathfrak{l}'_n\mathfrak{l}''$ is a left ideal in $\mathfrak{a}_k$, we must have $\mathfrak{l}'_n\mathfrak{l}'' = \mathfrak{l}''$. By Theorem 11.30(3), $\mathfrak{l}''$ is a minimal left ideal of A necessarily isomorphic to $\mathfrak{l}'_n$, and therefore to $\mathfrak{l}_k$.

If $\mathfrak{b}$ is a nonzero two-sided ideal of the algebra $\mathfrak{a}_k$, by Theorem 11.25, it contains at least one minimal left ideal $\mathfrak{l}''$ of this algebra, hence also contains all the $\mathfrak{l}''\mathfrak{l}'_n$ $(n \in I_k)$. But by Theorem 11.30(3), we have $\mathfrak{l}''\mathfrak{l}'_n = \mathfrak{l}'_n$, and therefore $\mathfrak{b}$ contains the sum of all the $\mathfrak{l}'_n$ (with $n \in I_k$). If $\mathfrak{b}$ is closed, it follows that $\mathfrak{b} = \mathfrak{a}_k$. Finally, $\mathfrak{a}_h\mathfrak{a}_k \subseteq \mathfrak{a}_h \cap \mathfrak{a}_k = (0)$ if $h \neq k$, because $\mathfrak{a}_h$ and $\mathfrak{a}_k$ are two-sided ideals. $\square$

The proof of the theorem shows that if J is infinite, then all the index sets I_k are finite. Also observe that Theorem 11.30(3) plays a crucial role in proving that $\mathfrak{a}_h\mathfrak{a}_k = (0)$ for all $h \neq k$.

Roughly speaking, Theorem 11.31 says that if A is a complete separable Hilbert algebra, then there is an irredundant list $(\mathfrak{l}_k)_{k\in J}$ of the minimal left ideals of A, and A is the Hilbert sum of two-sided ideals $\mathfrak{a}_k$, where each $\mathfrak{a}_k$ is the Hilbert sum obtained by picking a certain number of copies of the minimal left ideal $\mathfrak{l}_k$ of A.

Example 11.10. As an aid to help the reader process the indexing for the master decomposition of Theorem 11.31, suppose A is a complete separable Hilbert algebra with the following finite Hilbert sum decomposition:

$$A = \mathfrak{l}'_1 \oplus \mathfrak{l}'_2 \oplus \mathfrak{l}'_3 \oplus \mathfrak{l}'_4 \oplus \mathfrak{l}'_5 \oplus \mathfrak{l}'_6,$$

where

$$\mathfrak{l}'_1 \cong \mathfrak{l}'_3 \cong \mathfrak{l}'_4, \qquad \mathfrak{l}'_2 \cong \mathfrak{l}_6.$$

Set

$$\mathfrak{l}_1 := \mathfrak{l}'_1, \qquad \mathfrak{l}_2 := \mathfrak{l}'_2, \qquad \mathfrak{l}_3 := \mathfrak{l}'_5.$$

Then $J = (1, 2, 3)$, and

$$I_1 = (1, 3, 4), \qquad I_2 = (2, 6), \qquad I_3 = (5).$$

Also, note that

$$\begin{aligned}
\mathfrak{a}_1 &:= \bigoplus_{j\in I_1} \mathfrak{l}'_j = \mathfrak{l}'_1 \oplus \mathfrak{l}'_3 \oplus \mathfrak{l}'_4 \\
\mathfrak{a}_2 &:= \bigoplus_{j\in I_2} \mathfrak{l}'_j = \mathfrak{l}'_2 \oplus \mathfrak{l}'_6 \\
\mathfrak{a}_3 &:= \bigoplus_{j\in I_3} \mathfrak{l}'_j = \mathfrak{l}'_5,
\end{aligned}$$

which in turn implies

$$A = \mathfrak{a}_1 \oplus \mathfrak{a}_2 \oplus \mathfrak{a}_3 = (\mathfrak{l}'_1 \oplus \mathfrak{l}'_3 \oplus \mathfrak{l}'_4) \oplus (\mathfrak{l}'_2 \oplus \mathfrak{l}'_6) \oplus \mathfrak{l}'_5.$$

The ideals $\mathfrak{a}_k$ have properties analogous to those of simple algebras.

Definition 11.18. A complete Hilbert algebra A is *topologically simple* if it contains no *closed* two-sided ideal other than (0) and A.

Theorem 11.31 shows that the complete Hilbert algebras $\mathfrak{a}_k$ are topologically simple. Theorem 11.31 also shows that the study of complete separable Hilbert algebras reduces to the study of the topologically simple ones. We have the following theorem that gives us a sharper image of the structure of topologically simple complete separable Hilbert algebras.

Theorem 11.32. *(Structure of a topologically simple Hilbert algebra) Let A be topologically simple, complete, separable Hilbert algebra. For any minimal left ideal $\mathfrak{l}$ of A, the representation $U_{\mathfrak{l}}\colon A \to \mathcal{L}(\mathfrak{l})$ of A in the Hilbert space $\mathfrak{l}$ is faithful.*

If A is infinite-dimensional, then so is $\mathfrak{l}$. The image of A under $U_{\mathfrak{l}}$ is the algebra $\mathcal{L}_2(\mathfrak{l})$ of Hilbert–Schmidt operators on $\mathfrak{l}$, and there exists a constant $c > 0$ such that

$$c\langle x, y\rangle_A = \langle U_{\mathfrak{l}}(x), U_{\mathfrak{l}}(y)\rangle_{\mathrm{HS}} \qquad \text{for all } x, y \in A. \tag{$*$}$$

The inner product on the right-hand side is the inner product defined in Example 11.8.

If A is finite-dimensional, then the image of A under $U_{\mathfrak{l}}$ is the algebra $\mathrm{End}_{\mathbb{C}}(\mathfrak{l})$ *of all endomorphisms of the vector space $\mathfrak{l}$, and $(*)$ remains valid (the inner product on the right-hand side is also the inner product defined in Example 11.8). In fact,*

$$A = \bigoplus_{j=1}^{n} \mathfrak{l}_j, \qquad \mathfrak{l}_j = Ae_j = \bigoplus_{i=1}^{n} e_iAe_j,$$

where $e_1, \ldots, e_n$ are irreducible self-adjoint idempotents in A, the $\mathfrak{l}_j = Ae_j$ are isomorphic minimal left ideals, each space e_iAe_j is one-dimensional, and $e_1 + \cdots + e_n$ is the unit of A.

Proof. We follow Dieudonné's proof from [24] (Chapter XV, Section 8, Theorem 15.8.14). By Proposition 11.29 and by Theorem 11.31, we may assume that A is the Hilbert sum of a finite or countably infinite sequence of minimal left ideals $\mathfrak{l}_n = Ae_n$, where $\mathfrak{l} = \mathfrak{l}_1$ and all the $\mathfrak{l}_n$ are isomorphic. We begin by observing that for any $x \in A$ such that $x \neq 0$, we have $Ax \neq (0)$. Indeed, if $x \neq 0$ and if $Ax = (0)$, then for $x^* \in A$ we have $x^*x = 0$, but by Proposition 11.16 this implies that $x = 0$, a contradiction.

First we prove that the representation $U_{\mathfrak{l}}$ is faithful. If $U_{\mathfrak{l}}(x) = 0$ for some $x \neq 0$ in A, that is, if $x\mathfrak{l} = (0)$, then we should have $(Ax)\mathfrak{l} = (0)$. Since the ideal Ax is nonzero, by Theorem 11.25, the ideal Ax contains a minimal left ideal $\mathfrak{l}'$, and by Theorem 11.31, the left ideal $\mathfrak{l}'$ is isomorphic to $\mathfrak{l}$; hence $\mathfrak{l}'\mathfrak{l} = (0)$, contrary to Theorem 11.30(3). Therefore the representation $U_{\mathfrak{l}}$ is faithful.

Next, we claim that $\mathfrak{l}$ is the Hilbert sum of the subspaces e_nAe_1,

$$\mathfrak{l} = \bigoplus_n e_nAe_1,$$

where the spaces e_nAe_1 are one-dimensional.

Let $P_n = U_{\mathfrak{l}}(e_n)$. By Theorem 11.30(3), since the $\mathfrak{l}_i$ are isomorphic, the subspace e_nAe_1 is one-dimensional, and we shall show that P_n is the orthogonal projection of $\mathfrak{l} = Ae_1$ onto e_nAe_1. Since e_n is a self-adjoint idempotent and since an arbitrary element of Ae_1 is of the form xe_1 for some $x \in A$, and $U_{\mathfrak{l}}(e_n)(xe_1) = e_nxe_1$, we have

$$\langle xe_1 - e_nxe_1, e_nye_1\rangle = \langle e_nxe_1 - e_n^2xe_1, ye_1\rangle = \langle e_nxe_1 - e_nxe_1, ye_1\rangle = 0.$$

Since $e_me_n = 0$ whenever $m \neq n$, we have $P_mP_n = 0$, and therefore the subspaces e_nAe_1 are orthogonal in pairs. Moreover, $\mathfrak{l}$ is the Hilbert sum of these subspaces. Here, we use a standard fact of Hilbert space theory, which is that if (z_i) is an orthonormal family in $\mathfrak{l}$, and for every w, if w is orthogonal to all the z_i, then $w = 0$, then (z_i) is dense in $\mathfrak{l}$. Otherwise, $B = \overline{\bigoplus_i \mathbb{C}z_i}$ is a closed proper subspace of $\mathfrak{l}$, and thus its orthogonal complement $B^{\perp}$ is nonempty, so there is a nonzero $w \in B^{\perp}$ orthogonal to all the z_i. Since the subspaces e_iAe_1 are one-dimensional we can choose z_i to be some nonzero vector in e_iAe_1. Now if xe_1 is orthogonal to all the subspaces e_nAe_1 (which are one-dimensional), then $P_n(xe_1) = 0$ for all n, so that $e_nxe_1 = 0$ for all n, and thus $y = xe_1$ belongs to the right annihilator of A. Proposition 11.16 implies that this right annihilator is equal to (0), because $y^* \in A$, so $y^*y = 0$, and by Proposition 11.16 we must have $y = xe_1 = 0$. A similar proof shows that $\mathfrak{l}_j$ is the Hilbert sum

$$\mathfrak{l}_j = \bigoplus_k e_kAe_j, \tag{†}$$

where the spaces e_kAe_j are one-dimensional.

Equation (†) shows that the sequence $(\mathfrak{l}_k)$ is finite iff $\mathfrak{l}$ (and thus each $\mathfrak{l}_k$) is finite-dimensional over $\mathbb{C}$, or equivalently iff A is finite-dimensional. If A is finite-dimensional,

then

$$A = \bigoplus_{j=1}^{n} \mathfrak{l}_j, \qquad \mathfrak{l}_j = Ae_j = \bigoplus_{i=1}^{n} e_i A e_j,$$

where $e_1, \ldots, e_n$ are irreducible self-adjoint idempotents in A, the $\mathfrak{l}_j = Ae_j$ are isomorphic minimal left ideals, and each space e_iAe_j is one-dimensional. By Proposition 11.21, since the subspaces Ae_j are pairwise orthogonal, $e_ie_j = 0$ whenever $i \neq j$, so $e_1 + \cdots + e_n$ is a unit for e_iAe_j, and thus for A.

Let (a_n) be an orthonormal basis of $\mathfrak{l}$ such that $a_n \in e_nAe_1$ for each n. Since the e_n are self-adjoint idempotents, we have

$$a_ne_1 = a_n, \qquad e_na_n = a_n, \qquad a_na_n^* \in e_nAe_n, \qquad a_n^*a_n \in e_1Ae_1,$$

and by Theorem 11.30(2), we must have $a_na_n^* = \alpha_ne_n$ for some nonzero $\alpha_n \in \mathbb{C}$. Similarly, $a_n^*a_n \in e_1Ae_1$, and we must have $a_n^*a_n = \alpha_n'e_1$ for some nonzero $\alpha_n' \in \mathbb{C}$.

We claim that $\alpha_n = \alpha_n'$ for all n.

On the one hand,

$$a_na_n^*a_na_n^* = \alpha_ne_n\alpha_ne_n = \alpha_n^2e_n^2 = \alpha_n^2e_n,$$

and on the other hand, since $a_ne_1 = a_n$, we have

$$a_na_n^*a_na_n^* = a_n\alpha_n'e_1a_n^* = \alpha_n'a_ne_1a_n^* = \alpha_n'a_na_n^* = \alpha_n'\alpha_ne_n.$$

Consequently, $\alpha_n^2e_n = \alpha_n'\alpha_ne_n$, and since $\alpha_n', \alpha_n \neq 0$, we conclude that $\alpha_n = \alpha_n'$ for all n.

We also have

$$\langle e_n, e_n\rangle = \langle e_1, e_1\rangle, \qquad \text{for all } n \geq 1.$$

Indeed, we have

$$1 = \langle a_n, a_n\rangle = \langle a_n, e_na_n\rangle = \langle a_na_n^*, e_n\rangle = \alpha_n\langle e_n, e_n\rangle,$$

and

$$1 = \langle a_n, a_n\rangle = \langle a_n, a_ne_1\rangle = \langle a_n^*a_n, e_1\rangle = \alpha_n\langle e_1, e_1\rangle,$$

and since $\alpha_n \neq 0$, we obtain $\langle e_n, e_n\rangle = \langle e_1, e_1\rangle$ for all $n \geq 1$.

Let $c = 1/\langle e_1, e_1\rangle = \alpha_n$. Then $a_na_n^* = ce_n$, and for all $x, y \in A$, we have

$$\langle xa_n, ya_n\rangle = \langle y^*x, a_na_n^*\rangle = \langle y^*x, ce_n\rangle = c\langle xe_n, ye_n\rangle.$$

By Theorem 11.28, the series with general term $\langle xe_n, ye_n\rangle$ is absolutely convergent with sum $\langle x, y\rangle$, and since $\langle xa_n, ya_n\rangle = c\langle xe_n, ye_n\rangle$, if A is infinite-dimensional, then $\sum_n \langle xa_n, xa_n\rangle = \sum_n \|U_{\mathfrak{l}}(x)(a_n)\|^2$ converges, so $U_{\mathfrak{l}}(x)$ is a Hilbert–Schmidt operator, and Equation $(*)$ holds. Since A is a Hilbert space, so is its image under $U_{\mathfrak{l}}$, and to show that this image is the whole of the Hilbert space $\mathcal{L}_2(\mathfrak{l})$ (see Example 11.8), it suffices to show that $U_{\mathfrak{l}}(A)$ is dense in $\mathcal{L}_2(\mathfrak{l})$.

Now for all m, n with $m \neq n$, by Theorem 11.30(3), we have

$$(e_m A e_n)(e_n A e_1) = e_m(Ae_n)(Ae_1) = e_m A e_1,$$

and since $e_n A e_1 = \mathbb{C} a_n$, $a_n \in e_n A e_1$, and $a_m \in e_m A e_1$, it follows that there exists $e_{mn} \in e_m A e_n$ such that $e_{mn} a_n = a_m$ (which implies that $e_{mn} = c^{-1} a_m a_n^*$, since $a_n a_n^* = c e_n$ and $e_{mn} \in e_m A e_n$), and clearly, $e_{mn} a_p = 0$ if $p \neq n$. We conclude from this that $E_{mn} = U_{\mathfrak{l}}(e_{mn})$ is the continuous endomorphism of the Hilbert space $\mathfrak{l}$ such that $E_{mn}(a_n) = a_m$ and $E_{mn}(a_p) = 0$ if $p \neq n$. Our assertion now follows from the fact that it is not hard to show that the finite linear combinations of the E_{mn} are dense in $\mathcal{L}_2(\mathfrak{l})$.

We already took care of the case where A is finite-dimensional. $\square$

Example 11.11. Let us apply the construction of Theorem 11.32 to $\mathfrak{a}_1$ of Example 11.10. In other words let

$$A := \mathfrak{a}_1 = \mathfrak{l}_1' \oplus \mathfrak{l}_3' \oplus \mathfrak{l}_4' = \mathfrak{l}_1 \oplus \mathfrak{l}_2 \oplus \mathfrak{l}_3,$$

where

$$\mathfrak{l}_1 := \mathfrak{l}_1' \text{ and } \mathfrak{l}_1 = Ae_1, \quad \mathfrak{l}_2 := \mathfrak{l}_3' \text{ and } \mathfrak{l}_2 = Ae_2, \quad \mathfrak{l}_3 := \mathfrak{l}_4' \text{ and } \mathfrak{l}_3 = Ae_3.$$

Then

$$A = Ae_1 \oplus Ae_2 \oplus Ae_3.$$

Now set $\mathfrak{l} := \mathfrak{l}_1 = Ae_1$ and decompose each of the above $\mathfrak{l}_j$, where $1 \leq j \leq 3$, as

$$\begin{aligned}
\mathfrak{l} &= e_1 A e_1 \oplus e_2 A e_1 \oplus e_3 A e_1, \\
\mathfrak{l}_2 &= e_1 A e_2 \oplus e_2 A e_2 \oplus e_3 A e_2, \\
\mathfrak{l}_3 &= e_1 A e_3 \oplus e_2 A e_3 \oplus e_3 A e_3.
\end{aligned}$$

See Figure 11.2. We now scale the directions of $e_1 A e_1$, $e_2 A e_1$, and $e_3 A e_1$ to form the or-

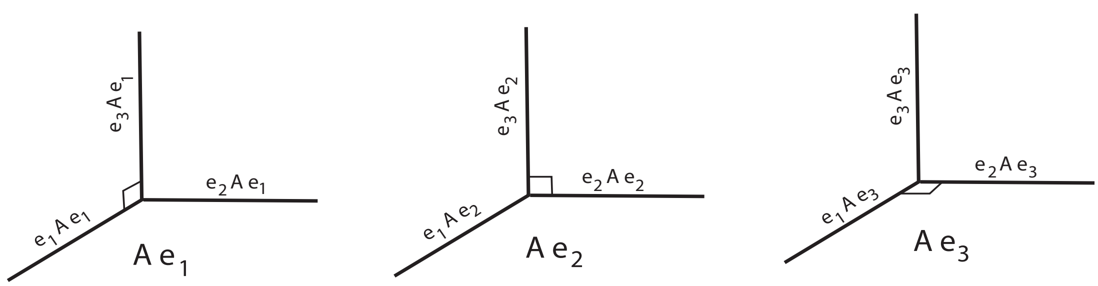


Figure 11.2: An illustration of the one-dimensional perpendicular directions within each $\mathfrak{l}_i$; i.e. the decomposition $Ae_i = e_1 A e_i \oplus e_2 A e_i \oplus e_3 A e_i$, for $1 \leq i \leq 3$.

thonormal basis $(a_i)_{1 \leq i \leq 3}$ of $\mathfrak{l}$, where

$$a_1 \in e_1 A e_1, \qquad a_2 \in e_2 A e_1, \qquad a_3 \in e_3 A e_1.$$

We also define elements six elements $e_{mn} \in e_m A e_n$, where $m, n \in \{1, 2, 3\}$ and $m \neq n$, i.e.

$$\begin{array}{lll} e_{21} \in e_2 A e_1, & e_{31} \in e_3 A e_1, & e_{12} \in e_1 A e_2, \\ e_{32} \in e_3 A e_2, & e_{13} \in e_1 A e_3, & e_{23} \in e_2 A e_3, \end{array}$$

such that

$$\begin{array}{lll} e_{21} a_1 = a_2, & e_{12} a_1 = a_1, & e_{13} a_3 = a_1 \\ e_{31} a_1 = a_3, & e_{32} a_2 = a_3, & e_{23} a_3 = a_2. \end{array}$$

See Figure 11.3.

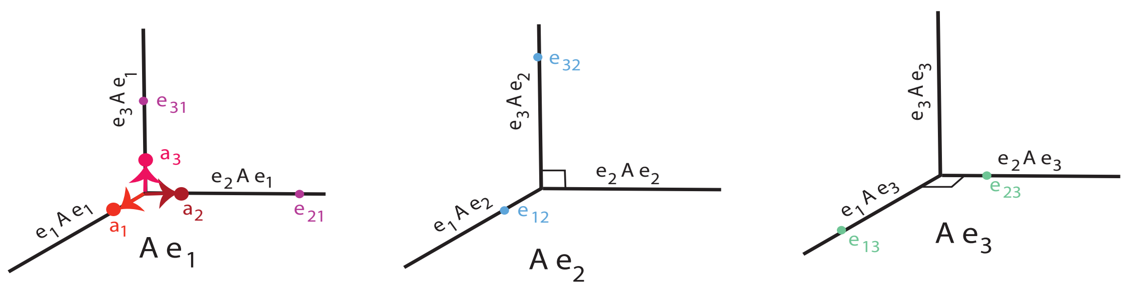


Figure 11.3: The orthonormal basis (a_1, a_2, a_3) and the elements $e_{mn} \in e_m A e_n$ whenever $m \neq n$.

Theorem 11.32 implies that in the Hilbert sum

$$A = \bigoplus_{k \in J} \mathfrak{a}_k$$

given by Theorem 11.31, the Hilbert algebra $\mathfrak{a}_k$, which is a building block of the decomposition, is either isomorphic to the algebra $\mathcal{L}_2(\mathfrak{l}_k)$ of Hilbert–Schmidt operators on $\mathfrak{l}_k$, or to the finite-dimensional algebra $\mathrm{End}_{\mathbb{C}}(\mathfrak{l}_k)$ of all endomorphisms of the vector space $\mathfrak{l}_k$. If G is a metrizable compact group and $A = \mathrm{L}^2(G)$, then every $\mathfrak{a}_k$ in the Hilbert sum for A is isomorphic to the finite-dimensional algebra $\mathrm{End}_{\mathbb{C}}(\mathfrak{l}_k)$.

Observe that in all cases, the spaces $U_{\mathfrak{l}}(\mathfrak{l}_n)$ consist of endomorphisms of the form $U(x) \circ P_n$ (where the $P_n = U_{\mathfrak{l}}(e_n)$ are the orthogonal projections defined in the proof of Theorem 11.32), and therefore consist of endomorphisms of rank 1 of $\mathfrak{l}$.

The following result gives a sufficient condition for a Hilbert algebra satisfying the hypotheses of Theorem 11.32 to be finite. Recall that the center $Z(A)$ of an algebra A is the set $Z(A) = \{x \in A \mid xy = yx, \text{ for all } y \in A\}$.

Proposition 11.33. *Let A be topologically simple, complete, separable Hilbert algebra. If there is an element $c \neq 0$ in the center of A, then A is finite-dimensional. In that case, the center of A is $\mathbb{C}\mathbf{1}$, where $\mathbf{1}$ is the identity element of A.*

Proof. If c belongs to the center of A, then for any irreducible self-adjoint idempotent e in A (which exists by Theorem 11.24 and Theorem 11.25), $u = U_{\mathfrak{l}}(c)$ is an endomorphism of the A-module $\mathfrak{l} = Ae$, and by Proposition 11.30(1), it is a map of the form $x \mapsto \alpha x$ for all $x \in \mathfrak{l}$, with $\alpha \in \mathbb{C}$. However, such a map cannot be a Hilbert-Schmidt operator on an infinite-dimensional space unless it is the zero map ($\sum_{n=1}^{\infty} \|u(a_n)\|^2$ can't be finite, where (a_i) is a countable Hilbert basis of $\mathfrak{l}$). Therefore $\mathfrak{l}$ is finite-dimensional, so by Theorem 11.32 A is also finite-dimensional. In this case, by Theorem 11.32(2), the algebra A is isomorphic to an algebra of $n \times n$ matrices, and it is a well-known fact of linear algebra that the matrices that commute with all $n \times n$ matrices are of the form λI_n, with $\lambda \in \mathbb{C}$ (where I_n is the $n \times n$ identity matrix). ☐

Theorem 11.31 will be applied to the complete, separable, Hilbert algebra $\mathrm{L}^2(G)$ (G metrizable and compact), which is then a Hilbert sum of two-sided ideals $\mathfrak{a}_k$. It turns out that Proposition 11.33 applies to the Hilbert algebras $\mathfrak{a}_k$, so they are finite-dimensional. This will yield the first part of the Peter–Weyl theorem.

We conclude with a result about representations of complete Hilbert algebras that makes use of Theorem 11.31.

Theorem 11.34. *(Master decomposition for nondegenerate representations) Let A be a separable, complete Hilbert algebra and let $\mathfrak{a}_{k\in J}$ be the topologically simple Hilbert algebras which are the Hilbert summands as in Theorem 11.31(3). For every $k \in J$, let $\mathfrak{l}_k$ be a minimal left ideal of $\mathfrak{a}_k$. Let $V\colon A \to \mathcal{L}(H)$ be a nondegenerate representation of A in a separable Hilbert space H, such that $V\colon A \to \mathcal{L}(H)$ is continuous. The following properties hold.*

(1) The Hilbert space H is the Hilbert sum of subspaces H_k ($k \in J$) invariant under V, such that if V_k is the restriction of V to H_k, we have $V_k(s) = 0$ for all $s \in \mathfrak{a}_h$ and for all h with $h \neq k$; thus V_k can be considered as a representation of $\mathfrak{a}_k$ in H_k.

(2) The representation V_k is the Hilbert sum of a finite or countably infinite sequence of irreducible representations, each equivalent to the representation $U_{\mathfrak{l}_k}$ of $\mathfrak{a}_k$ as in Theorem 11.32. If $\mathfrak{a}_k$ is finite-dimensional, then so is H_k.

Proof. We follow Dieudonné's proof from [24] (Chapter XV, Section 8, Theorem 15.8.16). Let H_k be the closure of the subspace spanned by the set of vectors

$$\{V(s_k)(x) \mid s_k \in \mathfrak{a}_k,\ x \in H\}.$$

Since every $s \in A$ can be written as $s = \sum_k s_k$ for some $s_k \in \mathfrak{a}_k$ and since V is continuous, we have

$$V(s)(x) = \sum_k V(s_k)(x),$$

which implies that H is the closure of the sum of the H_k. Also, if $h \neq k$ and if $s_h \in \mathfrak{a}_h$, $s_k \in \mathfrak{a}_k$, we have

$$\langle V(s_h)(x), V(s_k)(y)\rangle = \langle V^*(s_k)V(s_h)(x), y\rangle = \langle V(s_k^* s_h)(x), y\rangle = \langle V(0)(x), y\rangle = \langle 0, y\rangle = 0,$$

because $\mathfrak{a}_k$ is self-adjoint and $\mathfrak{a}_k\mathfrak{a}_h = (0)$ since $h \neq k$, so $s_k^* s_h = 0$. Thus (1) holds.

Now assume that A is topologically simple and finite-dimensional over $\mathbb{C}$, and thus has a unit element. This means that A is equal to some $\mathfrak{a}_k$ in the master decomposition theorem (Theorem 11.31), V is equal to V_k, and H is equal to H_k. Our representation is $V_k \colon \mathfrak{a}_k \to \mathcal{L}(H_k)$, which we also denote by $V \colon A \to \mathcal{L}(H)$. The case where A is infinite-dimensional can also be handled but it is a bit more complicated; see Dieudonné's [24] (Chapter XV, Section 8, Problem 1). If A is finite-dimensional, then $A = \bigoplus_{j=1}^n \mathfrak{l}_j$ is the sum of a finite number of isomorphic minimal left ideals $\mathfrak{l}_j$ in A. By Proposition 11.9, the representation $V \colon A \to \mathcal{L}(H)$ is the Hilbert sum (finite or countable) of topologically cyclic representations $V'_k \colon A \to \mathcal{L}(H'_k)$. Thus it suffices to prove that each topologically cyclic representation $V'_k \colon A \to \mathcal{L}(H'_k)$ is a Hilbert sum (finite or countable) of irreducible representations $V'_{k,j} \colon A \to \mathcal{L}(H'_{k,j})$, each equivalent to the representation $U_{\mathfrak{l}}$, where $\mathfrak{l}$ is some minimal left ideal in A. Then $V \colon A \to \mathcal{L}(H)$ is the Hilbert sum of the family (finite or countable) of irreducible representations $V'_{k,j} \colon A \to \mathcal{L}(H'_{k,j})$.

For simplicity of notation we may assume that $H = H'_k$, and let x_0 be a cyclic vector in H. The subspace of H spanned by the set of vectors $\{V(s)(x_0) \mid s \in A\}$ is finite-dimensional, and thus closed and dense in H, so it is equal to H. As a consequence, H is finite-dimensional and we can argue by induction on the dimension of H.

Since $A = \bigoplus_{j=1}^n \mathfrak{l}_j$ is the sum of a finite number of isomorphic minimal left ideals $\mathfrak{l}_j$, there is at least one minimal ideal, say $\mathfrak{l}$, such that the subspace $E = V(\mathfrak{l})(x_0) \subseteq H$ is nonzero. The surjection $\varphi \colon \mathfrak{l} \to E$ given by $\varphi(s) = V(s)(x_0)$ is then an A-module homomorphism, and since its kernel is a left ideal $\mathfrak{l}'$ contained in $\mathfrak{l}$ and distinct from $\mathfrak{l}$, since $\mathfrak{l}$ is minimal, we must have $\mathfrak{l}' = (0)$. Hence $\varphi \colon \mathfrak{l} \to E$ is a linear isomorphism and E is a subspace of H invariant under V. By Theorem 11.32, the representation $U_{\mathfrak{l}} \colon A \to \mathcal{L}(\mathfrak{l})$ is faithful and it is an isomorphism between A and $\mathrm{End}_{\mathbb{C}}(\mathfrak{l})$, with $\mathfrak{l}$ finite-dimensional, so if $\mathfrak{l}'$ is a proper linear subspace of $\mathfrak{l}$ invariant under $U_{\mathfrak{l}}$, as $U_{\mathfrak{l}}$ remains a faithful representation in $\mathfrak{l}'$ (because any linear map on $\mathfrak{l}'$ can be extended by zero to a linear map on $\mathfrak{l}$), A would be isomorphic to $\mathrm{End}_{\mathbb{C}}(\mathfrak{l}')$ whose dimension is strictly smaller than the dimension of $\mathrm{End}_{\mathbb{C}}(\mathfrak{l})$, a contradiction. Therefore, the representation $U_{\mathfrak{l}} \colon A \to \mathcal{L}(\mathfrak{l})$ is irreducible, and $\varphi \colon \mathfrak{l} \to E$ establishes an equivalence between the representations $U_{\mathfrak{l}} \colon A \to \mathcal{L}(\mathfrak{l})$ and $V \colon A \to \mathcal{L}(E)$. Since the orthogonal complement $E^\perp$ of E in H is also invariant under V and has dimension strictly smaller than the dimension of H, we can apply the induction hypothesis to the representation $V \colon A \to \mathcal{L}(E^\perp)$ to complete the proof. □

Example 11.12. Let us apply the proof techniques of Theorem 11.34 to $A = \mathfrak{a}_1 \oplus \mathfrak{a}_2 \oplus \mathfrak{a}_3$ where

$$\mathfrak{a}_1 = \mathfrak{l}_1 \oplus \mathfrak{l}_3 \oplus \mathfrak{l}_4, \qquad \mathfrak{a}_2 = \mathfrak{l}_2 \oplus \mathfrak{l}_6, \qquad \mathfrak{a}_3 = \mathfrak{l}_5;$$

see Example 11.10. Note that we have removed the primes so as to match the notation of the theorem. We are given a continuous representation $V\colon A \to \mathcal{L}(H)$. Define

$$\begin{aligned} H_1 &:= \overline{\mathrm{span}\{V(s_1)(x) \mid s_1 \in \mathfrak{a}_1,\ x \in H\}}, \\ H_2 &:= \overline{\mathrm{span}\{V(s_2)(x) \mid s_2 \in \mathfrak{a}_2,\ x \in H\}}, \\ H_3 &:= \overline{\mathrm{span}\{V(s_3)(x) \mid s_3 \in \mathfrak{a}_3,\ x \in H\}}. \end{aligned}$$

Then $H = H_1 \oplus H_2 \oplus H_3$, and since each H_i, for $1 \leq i \leq 3$, is invariant under V, we may define the restricted representations $V_1\colon \mathfrak{a}_1 \to \mathcal{L}(H_1)$, $V_2\colon \mathfrak{a}_2 \to \mathcal{L}(H_2)$, and $V_3\colon \mathfrak{a}_3 \to \mathcal{L}(H_3)$ such that if $s = s_1 + s_2 + s_3 \in \mathfrak{a}$, where $s_i \in \mathfrak{a}_i$ for $1 \leq i \leq 3$, and if $h = h_1 + h_2 + h_3 \in H$, where $h_i \in H_i$, for $1 \leq i \leq 3$, then

$$\begin{aligned} V(s)(h) &= V(s_1 + s_2 + s_3)(h_1 + h_2 + h_3) \\ &= V(s_1 + s_2 + s_3)(h_1) + V(s_1 + s_2 + s_3)(h_2) + V(s_1 + s_2 + s_3)(h_3) \\ &= V(s_1)(h_1) + V(s_2)(h_2) + V(s_3)(h_3) = V_1(s_1)(h_1) + V_2(s_2)(h_2) + V_3(s_3)(h_3). \end{aligned}$$

The above decomposition is a schematic example of Part 1 of Theorem 11.34.

To illustrate Part 2 of Theorem 11.34 we now focus on $V_1\colon \mathfrak{a}_1 \to \mathcal{L}(H_1)$. We also assume V_1 is a topologically cyclic representation with cyclic vector $x_0 \in H_1$, i.e.

$$\overline{\{V_1(s_1)(x_0) \mid s_1 \in \mathfrak{a}_1\}} = H_1.$$

Since $\mathfrak{a}_1 = \mathfrak{l}_1 \oplus \mathfrak{l}_3 \oplus \mathfrak{l}_4$, we will let $\mathfrak{l} := \mathfrak{l}_1$ be a minimal left ideal such that $E := V_1(\mathfrak{l}_1)(x_0)$ is a nonzero subspace of H_1 invariant under V_1. This means $H_1 = E \oplus E^{\perp}$ and V_1 is the Hilbert sum of the two restricted representations $V_{1,1}\colon \mathfrak{a}_1 \to \mathcal{L}(E)$ and $V_{1,2}\colon \mathfrak{a}_1 \to \mathcal{L}(E^{\perp})$. In other words, given $h_1 \in H_1$, if $h_1 = h_{1,1} + h_{1,2}$ with $h_{1,1} \in E$ and $h_{1,2} \in E^{\perp}$, then

$$V_1(s_1)(h_1) = V_1(s_1)(h_{1,1} + h_{1,2}) = V_1(s_1)(h_{1,1}) + V_1(s_1)(h_{1,2}) = V_{1,1}(s_1)(h_1) + V_{1,2}(s_1)(h_{1,2}).$$

Furthermore, $V_{1,1}$ is an *irreducible* representation which is *equivalent* to $U_{\mathfrak{l}_1}\colon \mathfrak{a}_1 \to \mathcal{L}(\mathfrak{l}_1)$. The equivalence is provided by the isomorphism $\varphi\colon \mathfrak{l}_1 \to E$, where $\varphi(s_{1,1}) = V_1(s_{1,1})(x_0)$ with $s_{1,1} \in \mathfrak{l}_1 \subseteq \mathfrak{a}_1$.

Theorem 11.34 will be used to prove another part of the Peter–Weyl theorem. In fact, we will only use Part (2) of Theorem 11.34 when $\mathfrak{a}_k$ is finite-dimensional, and we gave a proof in this case.

11.7 Positive Hilbert Forms And Representations

Our next goal is to state the Plancherel–Godement theorem, which will be used later in discussing harmonic analysis on the space induced by a Gelfand pair. A related theorem is the Bochner–Godement theorem. These theorems apply to a *commutative* Hilbert algebra

(not necessarily complete) arising from the quotient of a commutative Hilbert algebra by a left ideal induced by a bitrace satisfying two additional conditions. Therefore we go back to positive Hilbert forms (on an involutive algebra which is not necessarily commutative) to describe the construction of a certain representation. Commutativity is only required for the Plancherel–Godement theorem and the Bochner–Godement theorem. The first step is the following proposition.

The idea is that if g is a positive Hilbert form on an involutive (not necessarily commutative) algebra A, it almost defines an inner product, but in general it fails to be positive definite because there may be nonzero elements $s \in A$ such that $g(s,s) = 0$. However, if we take the quotient of A by the set $\mathfrak{n} = \{s \in A \mid g(s,s) = 0\}$, which is a left ideal because g is a positive Hilbert form, then we can define an inner product on the quotient vector space $A/\mathfrak{n}$. If g is a bitrace, then $A/\mathfrak{n}$ is an involutive algebra.

Proposition 11.35. *Let g be a positive Hilbert form on an involutive algebra A. The set*

$$\mathfrak{n} = \{s \in A \mid g(s,s) = 0\}$$

is a left ideal in A, and

$$\mathfrak{n} = \{s \in A \mid g(s,t) = 0 \quad \textit{for all } t \in A\}.$$

If $\pi\colon A \to A/\mathfrak{n}$ is the quotient map, then there exists a hermitian inner product $\langle -, - \rangle$ on $A/\mathfrak{n}$ such that

$$\langle \pi(s), \pi(t) \rangle = g(s,t) \qquad \textit{for all } s,t \in A$$

Furthermore, if g is a bitrace, then $\mathfrak{n}$ is a self-adjoint two-sided ideal, so $A/\mathfrak{n}$ is an algebra. The involution $s \mapsto s^*$ on A induces an involution on $A/\mathfrak{n}$ given by $(\pi(s))^* = \pi(s^*)$, and the hermitian inner product $\langle -, - \rangle$ on $A/\mathfrak{n}$ is a bitrace on $A/\mathfrak{n}$.

Proof. By the Cauchy–Schwarz inequality

$$|g(s,t)|^2 \leq g(s,s)g(t,t),$$

we see that

$$\mathfrak{n} = \{s \in A \mid g(s,t) = 0 \quad \text{for all } t \in A\}. \tag{$*_1$}$$

Since g is a positive Hilbert form, we have

$$g(xy,z) = g(y, x^*z) \qquad \text{for all } x,y \in A,$$

so if $s \in \mathfrak{n}$, that is, $g(s,s) = 0$, then by $(*_1)$,

$$g(ts,ts) = g(s, t^*st) = 0,$$

so $\mathfrak{n}$ is a left ideal. Since g is hermitian, $g(t,s) = \overline{g(s,t)}$, and we also have

$$\mathfrak{n} = \{s \in A \mid g(t,s) = 0 \quad \text{for all } t \in A\}.$$

Consequently, if $s - s' \in \mathfrak{n}$ and if $t - t' \in \mathfrak{n}$ (that is, $\pi(s) = \pi(s')$ and $\pi(t) = \pi(t')$), since

$$g(s, t - t') + g(s - s', t') = g(s, t) - g(s, t') + g(s, t') - g(s', t') = g(s, t) - g(s', t')$$

and since $s, t' \in A$, $t - t', s - s' \in \mathfrak{n}$, we have $g(s, t - t') = g(s - s', t') = 0$, and thus $g(s, t) - g(s', t') = 0$, that is,

$$g(s, t) = g(s', t').$$

We can now define the function $\langle -, - \rangle$ on $A/\mathfrak{n}$ by

$$\langle \pi(s), \pi(t) \rangle = g(s, t) \qquad \text{for all } s, t \in A,$$

and it is well-defined since $\pi(s) = \pi(s')$ and $\pi(t) = \pi(t')$ imply that $g(s, t) = g(s', t')$. It is immeditely verified that $\langle -, - \rangle$ is a hermitian inner product on $A/\mathfrak{n}$.

If g is a bitrace, then

$$g(y^*, x^*) = g(x, y) \qquad \text{for all } x, y \in A,$$

so

$$g(s^*, s^*) = g(s, s) \qquad \text{for all } s \in A,$$

which shows that $\mathfrak{n}^* = \mathfrak{n}$. If $s \in \mathfrak{n}$, then $s^* \in \mathfrak{n}$, and since $(st)^* = t^* s^*$ and $\mathfrak{n}$ is a left ideal,

$$g(st, st) = g((st)^*, (st)^*) = g(t^* s^*, t^* s^*) = 0,$$

which proves that $\mathfrak{n}$ is also a right ideal. It follows that $A/\mathfrak{n}$ is an algebra.

Since $\mathfrak{n} = \mathfrak{n}^*$, the map on $A/\mathfrak{n}$ given by

$$\pi(s) = s + \mathfrak{n} \mapsto (\pi(s))^* = s^* + \mathfrak{n}^* = s^* + \mathfrak{n} = \pi(s^*)$$

is an involution, and we have $(\pi(s))^* = \pi(s^*)$.

Finally, since by definition

$$\langle \pi(s), \pi(t) \rangle = g(s, t),$$

we get

$$\langle (\pi(t))^*, (\pi(s))^* \rangle = \langle \pi(t^*), \pi(s^*) \rangle = g(t^*, s^*) = g(s, t) = \langle \pi(s), \pi(t) \rangle,$$

which shows that $\langle -, - \rangle$ is a bitrace on $A/\mathfrak{n}$. $\square$

We are now going to show that if we add one more condition to a positive Hilbert form that insures that certain linear maps on $A/\mathfrak{n}$ are continuous, then we can define a representation of A into a Hilbert space which is the completion of $A/\mathfrak{n}$.

However, let us first observe that if g arises from a positive linear form f_{x_0} induced by a topologically cyclic representation $U \colon A \to \mathcal{L}(H)$ with cyclic vector $x_0 \in H$, where

$$f_{x_0}(s) = \langle U(s)(x_0), x_0 \rangle, \quad s \in A,$$

then the hermitian space $H_0 = \{U(s)(x_0) \mid s \in A\}$, which is dense in H, is determined by g_{x_0}, and in fact, the representation U is determined by g_{x_0}.

Indeed, since $g(s,t) = g_{x_0}(s,t) = f_{x_0}(t^*s)$, we have

$$g_{x_0}(s,t) = \langle U(s)(x_0), U(t)(x_0)\rangle, \tag{$*_2$}$$

and we see that

$$g_{x_0}(s,s) = \|U(s)(x_0)\|^2 .$$

Hence $\mathfrak{n}$ is the kernel of the linear map $h\colon A \to H_0$ given by

$$h(s) = U(s)(x_0).$$

Equation $(*_2)$ shows that the quotient map $\widehat{h}\colon A/\mathfrak{n} \to H_0$ is a bijective isometry. This shows that H_0, and therefore the Hilbert space H (since H_0 is dense in H), is determined by g_{x_0} up to isomorphism.

Since $h = \widehat{h} \circ \pi$, for every $s \in A$, the map $U(s)$ is completely determined by $\widehat{h}$ on H_0, because

$$U(s)(U(t)(x_0)) = U(st)(x_0) = h(st) = \widehat{h}(\pi(st)).$$

Since H_0 is dense in H, the continuous map $U(s)$ extends uniquely to H. Therefore, the representation U is determined by g_{x_0}.

We also have the following uniqueness result up to equivalence.

Proposition 11.36. *If $U_1\colon A \to \mathcal{L}(H_1)$ and $U_2\colon A \to \mathcal{L}(H_2)$ are two topologically cyclic representations with respective cyclic vectors x_0 and x_0', and if $f_{x_0} = f_{x_0'}$, that is, $\langle U_1(s)(x_0), x_0\rangle = \langle U_2(s)(x_0'), x_0'\rangle$ for all $s \in A$, then the representations U_1 and U_2 are equivalent.*

Proof. We follow Dieudonné [24] (Chapter XV, Section 6, Theorem 15.6.7). For all $s,t \in A$ we have

$$\begin{aligned}\langle U_1(s)(x_0), U_1(t)(x_0)\rangle &= \langle U_1(t^*s)(x_0), x_0\rangle\\ &= \langle U_2(t^*s)(x_0'), x_0'\rangle\\ &= \langle U_2(s)(x_0'), U_2(t)(x_0')\rangle;\end{aligned}$$

that is,

$$\langle U_1(s)(x_0), U_1(t)(x_0)\rangle = \langle U_2(s)(x_0'), U_2(t)(x_0')\rangle, \qquad \text{for all } s,t \in A. \tag{$*_3$}$$

Since the vectors $U_1(t)(x_0)$ (resp. $U_2(t)(x_0')$) form a dense subset H_1^0 (resp. H_2^0) of H_1 (resp. H_2) and since the inner product is continuous in each argument, we deduce that $U_1(s)(x_0)$ is orthogonal to H_1 iff $U_2(s)(x_0')$ is orthogonal to H_2, so $U_1(s)(x_0) = 0$ iff $U_2(s)(x_0') = 0$ for all $s \in A$. As a consequence, for every $z \in H_1^0$ and for all $s,t \in A$ such that $U_1(s)(x_0) = U_1(t)(x_0) = z$, since $U_1(s-t)(x_0) = 0$ iff $U_2(s-t)(x_0') = 0$, we have $U_2(s)(x_0') = U_2(t)(x_0')$,

which means that the vector $U_2(s)(x_0')$ has the same value $z' \in H_2^0$ for all $s \in A$ such that $U_1(s)(x_0) = z$, so we can define the map $T\colon H_1^0 \to H_2^0$ by $T(z) = z'$, or equivalently,

$$T(U_1(s)(x_0)) = U_2(s)(x_0'), \qquad s \in A.$$

It is immediately verified that T is a surjective linear map, and by $(*_3)$, it is an isometry of the hermitian space H_1^0 onto the hermitian space H_2^0. But then this isomorphism extends uniquely to an isomorphism, also denoted T, between the Hilbert spaces H_1 and H_2. It remains to show that T induces an equivalence of the representations U_1 and U_2. Since H_1^0 is dense in H_1 and H_2^0 is dense in H_2, it suffices to prove that

$$T(U_1(s)(z)) = U_2(s)(T(z))$$

for all z of the form $z = U_1(t)(x_0)$ $(t \in A)$. Since

$$U_1(s)(U_1(t)(x_0)) = U_1(st)(x_0),$$

by definition of T, we have

$$\begin{aligned} T(U_1(s)(z)) &= T(U_1(s)(U_1(t)(x_0))) \\ &= T(U_1(st)(x_0)) \\ &= U_2(st)(x_0') \\ &= U_2(s)(U_2(t)(x_0')) = U_2(s)(T(z)), \end{aligned}$$

as claimed. $\square$

We now go back to our positive Hilbert form g, and we assume that it satisfies the analog of Condition (U) of Definition 11.14: for every $s \in A$, there is some $M_s \geq 0$ such that

$$g(st, st) \leq M_s g(t, t), \quad \text{for all } t \in A, \tag{U}$$

If g arises from a representation $U\colon A \to \mathcal{L}(H)$ as

$$g(s, t) = \langle U(s)(x_0), U(t)(x_0)\rangle, \qquad s, t \in A,$$

then it is easy to see that g satisfies Property (U).

If the positive Hilbert form g in Proposition 11.35 satisfies Condition (U), then we check immediately that the inner product $\langle -, -\rangle$ on $A/\mathfrak{n}$ given by

$$\langle \pi(s), \pi(t)\rangle = g(s, t)$$

also satisfies Property (U).

The following proposition shows how to construct a representation of A from g.

Proposition 11.37. *Let g be a positive Hilbert form on an involutive algebra A satisfying Condition (U), let $\mathfrak{n}_g$ be the left ideal given by*

$$\mathfrak{n}_g = \{s \in A \mid g(s, s) = 0\},$$

and suppose that the hermitian space $A/\mathfrak{n}_g$ constructed in Proposition 11.35 is separable. If so, let H_g be the Hilbert space which is the completion of $A/\mathfrak{n}_g$, so that $A/\mathfrak{n}_g$ can be identified with a dense subspace H_0 of the separable Hilbert space H_g. If $\pi_g \colon A \to A/\mathfrak{n}_g$ denotes the quotient map, for every $s \in A$, the linear map $U_g(s) \colon A/\mathfrak{n}_g \to A/\mathfrak{n}_g$ given by

$$U_g(s)(\pi_g(t)) = \pi_g(st)$$

extends to a continuous linear map $U_g(s) \colon H_g \to H_g$, and the map $s \mapsto U_g(s)$ is a representation of A in H_g. If g is a bitrace, then $A/\mathfrak{n}_g$ is an involutive algebra, and the inner product $\langle -, - \rangle_g$ on $A/\mathfrak{n}_g$ given by

$$\langle \pi_g(s), \pi_g(t) \rangle_g = g(s, t)$$

is a bitrace that satisfies Property (U).

Proof. If $\pi_g(t) = \pi_g(t')$, then $\pi_g(st) = \pi_g(st')$ because $\mathfrak{n}_g$ is a left ideal. Hence for any fixed $s \in A$, the endomorphism of H_0 given by $\pi_g(t) \mapsto \pi_g(st)$ is well-defined. The definition of the inner product $\langle -, - \rangle_g$ on $A/\mathfrak{n}_g$ and Condition (U) ensure that this map is continuous. Since H_0 is dense in H_g and the map $\pi_g(t) \mapsto U_g(s)(\pi_g(t)) = \pi_g(st)$ is continuous, it extends to a continuous map $U_g(s) \colon H_g \to H_g$.

Since $\pi_g((ss')t) = \pi_g(s(s't))$, we have $U_g(ss') = U_g(s) \circ U_g(s')$. Since g is a positive Hilbert form, we have

$$g(xy, z) = g(y, x^* z) \quad \text{for all } x, y, z \in A$$

and since g is hermitian, we have by Proposition 11.35 that

$$\begin{aligned} \langle U_g(s^*)(\pi_g(t)), \pi_g(t') \rangle_g &= \langle \pi_g(s^* t), \pi_g(t') \rangle = g(s^* t, t') \\ &= g(t, st') \\ &= \overline{g(st', t)} = \overline{\langle \pi_g(st'), \pi_g(t) \rangle} \\ &= \overline{\langle U_g(s)(\pi_g(t')), \pi_g(t) \rangle_g} \\ &= \langle \pi_g(t), U_g(s)(\pi_g(t')) \rangle_g, \end{aligned}$$

which shows that $U_g(s^*) = (U_g(s))^*$. If A has a unit element e then $U_g(e) = \mathrm{id}$. Therefore U_g is a representation of A in H_g.

When g is bitrace, the ideal $\mathfrak{n}_g$ is a two-sided ideal, in which case $A/\mathfrak{n}_g$ is an algebra and $\pi_g(s)\pi_g(t) = \pi_g(st)$. $\square$

In general, the representation $U_g \colon A \to \mathcal{L}(H_g)$ given by Proposition 11.37 may be degenerate. It is nondegenerate if and only if the following condition holds:

$$\text{the subspace spanned by the set } \{\pi_g(st) \mid s, t \in A\} \text{ is dense in } A/\mathfrak{n}_g. \tag{N}$$

Observe that the above condition is the generalization of Condition (N) of Definition 11.14 to $A/\mathfrak{n}_g$. In Definition 11.14, the bitrace g is an inner product so $\mathfrak{n}_g = (0)$. If A has a unit element, Condition (N) holds trivially.

If the positive Hilbert form g in Proposition 11.35 satisfies Condition (N), then we check immediately that the inner product $\langle -, - \rangle$ on $A/\mathfrak{n}$ given by

$$\langle \pi(s), \pi(t) \rangle = g(s, t)$$

also satisfies Property (N).

It can be shown that if A is a unital Banach algebra with involution, then of course Condition (N) is automatically satisfied, but Condition (U) also holds. This follows from the following result proven in Dieudonné [24] (Chapter XV, Section 7, Theorem 15.6.11).

Proposition 11.38. *Let A be a unital Banach algebra with involution and unit element $e \neq 0$. For any positive linear form f, the following properties hold.*

(1) f is continuous and $\|f\| = f(e)$.

(2) $|f(y^*xy)| \leq \|x\| \, f(y^*y)$, for all $x, y \in A$.

Recall that since A is unital, every positive Hilbert form g arises from the positive linear form f given by $f(s) = g(s, e)$. Then for every $s \in A$,

$$g(st, st) = f((st)^*st) = f(t^*s^*st) \leq \|s^*s\| \, f(t^*t) = \|s^*s\| \, g(t, t),$$

which is Condition (U) with $M_s = \|s^*s\|$.

The proof of Proposition 11.38 makes use of the following result of independent interest also proven in Dieudonné [24] (Chapter XV, Section 7, Theorem 15.6.11.1).

Proposition 11.39. *Let A be a unital Banach algebra with involution and unit element $e \neq 0$. If $x \in A$ is self-adjoint and $\|x\| < 1$, then there exists a self-adjoint element $y \in A$ such that $y^2 = e + x$.*

Here is a fairly general situation where a positive Hilbert form satisfies conditions (U) and (N).

Proposition 11.40. *Let A be a unital separable Banach algebra with involution and unit element $e \neq 0$. For any positive linear form f on A, let g be the corresponding positive Hilbert form given by $g(x, y) = f(y^*x)$ for all $x, y \in A$. Let $\mathcal{A}_g$ be the unital Banach algebra which is the closure of $U_g(A)$ in $\mathcal{L}(H_g)$. Then g satisfies Property (U), and the hermitian space $A/\mathfrak{n}_g$ and the unital Banach algebra $\mathcal{A}_g \subseteq \mathcal{L}(H_g)$ are separable.*

Proof. By Proposition 11.38, the positive Hilbert form g induced by the positive linear form f satisfies Property (U). By Proposition 11.37, we have

$$\|\pi_g(x)\|^2 = \langle \pi_g(x), \pi_g(x)\rangle_g = g(x,x) = f(x^*x).$$

By Proposition 11.38, since f is continuous we get

$$\|\pi_g(x)\|^2 = f(x^*x) \leq \|f\| \, \|x^*x\| \leq \|f\| \, \|x\|^2 ,$$

which shows that π_g is continuous. Since π_g is also surjective, it is easy to show that the image under π_g of a countable dense set in A is dense in $A/\mathfrak{n}_g$, so $A/\mathfrak{n}_g$ is separable. Since by Proposition 11.1, the map U_g is continuous, it sends a countable dense set into a countable dense set in $\mathcal{A}_g$, so $\mathcal{A}_g$ is also separable. ☐

11.8 The Plancherel–Godement Theorem ⊛

After these preliminaries, we assume that A is a commutative (but not necessarily complete) involutive algebra equipped with a bitrace g satisfying Conditions (U) and (N). We also assume that the hermitian space $A/\mathfrak{n}_g$ is separable, as in Proposition 11.37. Then by Proposition 11.37, the bitrace g induces a representation $U_g \colon A \to \mathcal{L}(H_g)$, where the separable Hilbert space H_g is the completion of $A/\mathfrak{n}_g$, so that $A/\mathfrak{n}_g$ is dense in the Hilbert space H_g. Since Property (N) holds, the representation U_g is nondegenerate. The image of A under U_g is a commutative subalgebra of the C^*-algebra $\mathcal{L}(H_g)$. Let $\mathcal{A}_g$ *be the closure of* $U_g(A)$ *in* $\mathcal{L}(H_g)$, so that $\mathcal{A}_g$ is a commutative C^*-algebra (and thus, consists of normal operators). In general, $\mathcal{A}_g$ is not separable, *but we assume it is separable.*

A particular example of a trace f on A, which gives rise to a bitrace g (such that $g(x,y) = f(y^*x)$) is provided by the hermitian characters of A. These are the characters $\chi \in \mathsf{X}(A)$ such that

$$\chi(x^*) = \overline{\chi(x)} \qquad \text{for all } x \in A.$$

We have $\chi(x^*x) = \chi(x^*)\chi(x) = |\chi(x)|^2$, and

$$g(x,y) = \chi(y^*x) = \chi(y^*)\chi(x) = \overline{\chi(y)}\chi(x).$$

Condition (U) holds because

$$g(st,st) = \overline{\chi(st)}\chi(st) = \overline{\chi(s)}\,\overline{\chi(t)}\chi(s)\chi(t) = |\chi(s)|^2|\chi(t)|^2 = |\chi(s)|^2 g(t,t).$$

The ideal $\mathfrak{n}_g$ is the kernel of χ, so by Proposition 9.12(1) it is a hyperplane in A, thus $A/\mathfrak{n}_g$ is isomorphic to $\mathbb{C}$, and Condition (N) follows immediately from the fact that $\chi(x) \neq 0$ implies that $\chi(x^2) = (\chi(x))^2 \neq 0$. It can be shown that the corresponding representation is irreducible.

Let $\mathbf{H}(A)$ denote the set of hermitian characters in $\mathsf{X}(A)$. The set $\mathbf{H}(A)$ is a subset of the product space $\mathbb{C}^A$ and is closed under the product topology. We give $\mathbf{H}(A)$ the topology

induced by the product topology (the topology of pointwise convergence). When A is a unital commutative Banach algebra with involution, the space $\mathbf{H}(A)$ is a compact subspace of $\mathsf{X}(A)$, since $\mathsf{X}(A)$ is compact, by Theorem 9.19. If A is also separable, then it can be shown that $\mathsf{X}(A)$ is metrizable; see Dieudonné [24] (Chapter XV, Section 3, Theorem 15.3.2). In general, $\mathbf{H}(A) \neq \mathsf{X}(A)$, but if A is a unital C^*-algebra, then $\mathbf{H}(A) = \mathsf{X}(A)$, by Proposition 9.34.

The following theorem shows that the bitraces discussed in this section and the previous one all arise from a positive measure by a process of integration involving the hermitian characters.

Theorem 11.41. *(Plancherel–Godement theorem) Let A be a commutative involutive algebra, and let g be a bitrace on A satisfying Conditions (U) and (N), such that the hermitian space $A/\mathfrak{n}_g$ and the C^*-algebra $\mathcal{A}_g \subseteq \mathcal{L}(H_g)$ are separable.*

(I) We can define canonically: (1) a subspace S_g of $\mathbf{H}(A)$ whose closure in $\mathbb{C}^A$ is either S_g or $S_g \cup \{0\}$ and is metrizable and compact (so that S_g is locally compact, metrizable, and separable); (2) a (positive) Radon measure m_g on S_g, with the following properties:

(i) For each $x \in A$, define the function $\widehat{x}\colon S_g \to \mathbb{C}$ by $\widehat{x}(\chi) = \chi(x)$. Then $\widehat{x} \in \mathcal{L}^2_{m_g}(S_g)$, and we have

$$g(x, y) = \int_{S_g} \chi(xy^*)\, dm_g(\chi) = \int_{S_g} \widehat{x}(\chi)\overline{\widehat{y}(\chi)}\, dm_g(\chi) \qquad \text{for all } x, y \in A. \qquad (\dagger)$$

(ii) As x runs through A, the set of functions $\widehat{x}$ is contained in $\mathcal{C}_0(S_g; \mathbb{C})$ and is dense in this Banach space.

(iii) The map $x \mapsto \widehat{x}$ factors as $T_0 \circ \pi_g$ as illustrated below,

$$\begin{tikzcd} A \arrow[r, "\pi_g"] \arrow[dr, "\widehat{\ }"'] & A/\mathfrak{n}_g \arrow[r] \arrow[d, "T_0"] & H_g \arrow[d, "T"] \\ & \mathrm{L}^2_{m_g}(S_g) \cap \mathcal{C}_0(S_g, \mathbb{C}) \arrow[r] & \mathrm{L}^2_{m_g}(S_g), \end{tikzcd}$$

where the map $T_0\colon A/\mathfrak{n}_g \to \mathrm{L}^2_{m_g}(S_g) \cap \mathcal{C}_0(S_g; \mathbb{C})$ extends to an isomorphism $T\colon H_g \to \mathrm{L}^2_{m_g}(S_g)$, such that for all $x \in A$, we have $U_g(x) = T^{-1}M(\widehat{x})T$, where $M(\widehat{x})$ is multiplication by the class of $\widehat{x}$ in $\mathrm{L}^2_{m_g}(S_g)$, where U_g is the representation $U_g\colon A \to \mathcal{L}(H_g)$.

(iv) We have

$$\left\|\lambda \mathrm{id}_{H_g} + U_g(x)\right\| = \left\|\lambda \mathrm{id}_{S_g} + \widehat{x}\right\|$$

for all $\lambda \in \mathbb{C}$ and all $x \in A$.

(II) Conversely, let S be a subspace of $\mathbf{H}(A)$ such that $S \cup \{0\}$ is compact and metrizable, and let m be a positive measure on S, such that for all $x \in A$, the function $\widehat{x}\colon S \to \mathbb{C}$ given by $\widehat{x}(\chi) = \chi(x)$ belongs to

$$\mathcal{L}^2_m(S) \cap \mathcal{C}_0(S; \mathbb{C}).$$

Then the map g' given by

$$g'(x, y) = \int_S \widehat{x}(\chi)\overline{\widehat{y}(\chi)}\, dm(\chi) \qquad \text{for all } x, y \in A$$

is a bitrace on A satisfying Conditions (U) and (N), such that $A/\mathfrak{n}_{g'}$ and $\mathcal{A}_{g'}$ are separable, and we have $S_{g'} = S$ and $m_{g'} = m$.

Note that for any $x \in A$, the map $\widehat{x}$ is the restriction of the Gelfand transform to either S_g or S.

Theorem 11.41 is proven in Dieudonné [24] (Chapter XV, Section 9, Theorem 15.9.2). The proof is long and very technical. Among other results, it uses the Gelfand–Naimark theorem (Theorem 9.37). We simply describe the construction of S_g since it will be used in Theorem 17.21.

Let $\mathcal{A}'_g = \mathcal{A}_g \oplus \mathbb{C}\mathrm{id}_{H_g}$. This is a closed subalgebra of $\mathcal{L}(H_g)$ so it is a commutative C^* algebra with a unit element. By Proposition 9.34, every character $\xi' \in \mathsf{X}(\mathcal{A}'_g)$ is hermitian, and since $U_g(a^*) = (U_g(a))^*$, the map $\xi' \circ U_g$ is either the zero map or a hermitian character of A. This means that we have have a map $\omega\colon \mathsf{X}(\mathcal{A}'_g) \to \mathbf{H}(A) \cup \{0\}$ given by $\omega(\xi') = \xi' \circ U_g$.

The map ω is injective because $\xi'(\mathrm{id}_{H_g}) = 1$ for all $\xi' \in \mathsf{X}(\mathcal{A}'_g)$, and since ξ' is continuous on $\mathcal{A}'_g$ and $U_g(A)$ is dense in $\mathcal{A}_g$, the restriction of ξ' to $U_g(A)$ has a unique extension to $\mathcal{A}_g$, so the character ξ' is uniquely determined. The map ω is also continuous with respect to the weak topologies on $\mathsf{X}(\mathcal{A}'_g)$ and $\mathbb{C}^A$. Since $\mathsf{X}(\mathcal{A}'_g)$ is metrizable and compact, the same is true of its image S'_g, with

$$S'_g = \omega(\mathsf{X}(\mathcal{A}'_g)) \subseteq \mathbf{H}(A) \cup \{0\},$$

and ω is a *homeomorphism* of $\mathsf{X}(\mathcal{A}'_g)$ onto S'_g. Then the space S_g is defined as follows.

(1) If $\mathrm{id}_{H_g} \in \mathcal{A}_g$, then $\mathcal{A}'_g = \mathcal{A}_g$ and S'_g does not contain the element $0 \in \mathbb{C}^A$. We set

$$S_g = S'_g.$$

(2) If $\mathrm{id}_{H_g} \notin \mathcal{A}_g$, then $\mathcal{A}_g$ is a closed hyperplane and an ideal in $\mathcal{A}'_g$, hence it is a maximal ideal. In fact, there is a nonzero character ξ'_0 of $\mathcal{A}'_g$ whose kernel is $\mathcal{A}_g$, so $\omega(\xi'_0) = 0 \in \mathbb{C}^A$. We set

$$S_g = S'_g - \{0\}.$$

In both cases, S_g is separable, metrizable, locally compact, and the complements in $S_g \cup \{0\}$ of the compact subsets of S_g are the open sets in $S_g \cup \{0\}$ that contain 0. The Gelfand–Naimark theorem (Theorem 9.37) also shows that the Gelfand transform is an isometry between $\mathcal{A}'_g$ and $\mathcal{C}_0(\mathsf{X}(\mathcal{A}'_g); \mathbb{C})$. Then (ii) follows quite easily.

The construction of the measure m_g is far more involved.

We should mention that Dieudonné uses a theory of integration in which positive Radon functionals are used instead of Borel measures. However, the version in Dieudonné also states that the support of the Radon functional is the whole of S_g, and by Proposition A.49, since S_g is locally compact, metrizable and separable, it is σ-compact, so there is no problem in obtaining the theorem for Radon measures measures by using Radon–Riesz II (Theorem 7.15).

If the bitrace g arises from a positive linear form f, which is a trace since A is commutative, the formula (†) leads us to ask whether we also have

$$f(x) = \int_{S_g} \widehat{x}(\chi)\, dm_g(\chi).$$

The Bochner–Godement theorem provides a partial answer to this question.

Theorem 11.42. *(Bochner–Godement theorem) Let A be a commutative involutive algebra A.*

(1) Let f be a positive linear form such that the bitrace g given by $g(x, y) = f(y^* x)$ satisfies the hypotheses of the Plancherel–Godement theorem (Theorem 11.41). Then if the formula

$$f(x) = \int_{S_g} \widehat{x}(\chi)\, dm_g(\chi) \tag{BG}$$

holds and if the (positive) Radon measure m_g is bounded, then f satisfies the following condition:

$$\text{There is some } M > 0 \text{ such that } |f(x)|^2 \leq M f(xx^*) \text{ for all } x \in A. \tag{B}$$

(2) Conversely, let f be positive linear form on A which satisfies Condition (B), and suppose that the induced bitrace g given by $g(x, y) = f(y^* x)$ satisfies Condition (U), and is such that the hermitian space $A/\mathfrak{n}_g$ and the C^*-algebra $\mathcal{A}_g$ are separable. Then g also satisfies Condition (N), the (positive) Radon measure m_g is bounded, and formula (BG) holds.

Theorem 11.42 is proven in Dieudonné [24] (Chapter XV, Section 9, Theorem 15.9.4). The proof is shorter than the proof of the Plancherel–Godement theorem and also uses the Gelfand–Naimark theorem.

If A is unital, then by Proposition 11.11(3), Condition (B) is satisfied with $M = f(e)$. Also, by Proposition 11.40, if A is a unital separable Banach algebra with involution, then the hermitian space $A/\mathfrak{n}_g$ and the C^*-algebra $\mathcal{A}_g$ are separable. Therefore, if A is commutative unital separable Banach algebra with involution, then the Bochner–Godement theorem Part 2 applies to *any* positive linear form on A.

Here is another situation in which the Bochner–Godement theorem Part 2 applies. Let $U\colon A \to \mathcal{L}(H)$ be a representation of a commutative involutive algebra A into a *separable* Hilbert space H. For any $x_0 \in H$, let f_{x_0} be the positive linear form given by

$$f_{x_0}(s) = \langle U(s)(x_0), x_0\rangle.$$

We claim that f_{x_0} satisfies Condition (B). Indeed, by Cauchy–Schwarz, we have

$$|f_{x_0}(s)|^2 \leq \|U(s)(x_0)\|^2 \, \|x_0\|^2 = \|x_0\|^2 \, f_{x_0}(s^* s).$$

We know that the bitrace g induced by f_{x_0} (given by $g(x,y) = f_{x_0}(y^* x)$) satisfies Condition (U), and because H is separable, the hermitian space $A/\mathfrak{n}_g$ and the C^*-algebra $\mathcal{A}_g$ are separable. *Therefore the Bochner-Godement theorem applies to the positive linear form f_{x_0}, and Equation (BG) shows that f_{x_0} is determined by the hermitian characters of* $\mathsf{X}(A)$.

As shown just after Proposition 11.35, if the topologically cyclic representation $U\colon A \to \mathcal{L}(H)$ has $x_0 \in H$ as cyclic vector, then it is completely determined by f_{x_0}. Also, by Proposition 11.9, if A is unital, then every (nondegenerate) representation of A in a separable Hilbert space is the countable Hilbert sum of topologically cyclic representations. Therefore, if A is commutative unital algebra, then every representation $U\colon A \to \mathcal{L}(H)$ of A in a separable Hilbert space H is completely determined by the hermitian characters of $\mathsf{X}(A)$.

As a nice application of both the Plancherel–Godement theorem and the Bochner–Godement theorem we obtain a characterization of the nondegenerate representation of the algebra $\mathcal{C}(K)$ of continuous functions on a compact metrizable space K.

11.9 Representations of Algebras of Continuous Functions

Let K be a compact metrizable space. Then $\mathcal{C}_{\mathbb{C}}(K)$ is a commutative unital C^*-algebra under pointwise multiplication, see Example 9.1(2) and Example 9.6(2). By Proposition 9.22, the space K is homeomorphic to the space $\mathsf{X}(\mathcal{C}_{\mathbb{C}}(K))$ of characters of $\mathcal{C}_{\mathbb{C}}(K)$ (which are Dirac measures on K), and the Gelfand transform from $\mathcal{C}_{\mathbb{C}}(K)$ to $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{C}_{\mathbb{C}}(K))$ can be viewed as the identity. Furthermore, by Proposition 9.34, the characters in $\mathsf{X}(\mathcal{C}_{\mathbb{C}}(K))$ are hermitian (so $\mathbf{H}(\mathcal{C}_{\mathbb{C}}(K)) = \mathsf{X}(\mathcal{C}_{\mathbb{C}}(K))$). Write $A = \mathcal{C}_{\mathbb{C}}(K)$, and let $U\colon A \to \mathcal{L}(H)$ be a topologically cyclic representation of the commutative C^*-algebra A into a *separable* Hilbert space H. As discussed at the end of Section 11.8, for any cyclic vector $x_0 \in H$, we have the positive linear form f_{x_0} given by

$$f_{x_0}(u) = \langle U(u)(x_0), x_0\rangle, \quad u \in A,$$

and by Bochner-Godement the positive linear form f_{x_0} is completely determined by the space $\mathsf{X}(A)$ of hermitian characters, the representation U is completely determined by f_{x_0}, and thus by $\mathsf{X}(A) \approx K$. If g is the bitrace associated with f_{x_0}, it can be shown (exercise left to the reader) that the subspace $S_g \subseteq \mathbf{H}(A)$ introduced in the Plancherel–Godement

theorem (Theorem 11.41) is actually equal to $\mathbf{H}(A) = \mathsf{X}(A) \approx K$. Therefore we can apply Theorem 11.41(iii) to obtain the following remarkable result.

Theorem 11.43. *Let K be a compact metrizable space. Every topologically cyclic representation $U\colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(H)$ of the commutative unital C^*-algebra $\mathcal{C}_{\mathbb{C}}(K)$ in a separable Hilbert space H is equivalent to a representation $M_\mu\colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(\mathrm{L}^2_\mu(K;\mathbb{C}))$ obtained as follows: for some positive Radon measure μ on K, for every $u \in \mathcal{C}_{\mathbb{C}}(K)$, let $M_\mu(u)\colon \mathrm{L}^2_\mu(K;\mathbb{C}) \to \mathrm{L}^2_\mu(K;\mathbb{C})$ be the continuous linear map induced by multiplication by u; that is, for every $f \in \mathcal{L}^2_\mu(K;\mathbb{C})$, define $M_\mu(u)(\mathbf{f})$ as the equivalence class $\mathbf{uf}$ of uf in $\mathrm{L}^2_\mu(K;\mathbb{C})$. More precisely, there is some unitary map $W\colon H \to \mathcal{L}^2_\mu(K;\mathbb{C})$ such that*

$$WU(u)W^{-1} = M_\mu(u) \quad \textit{for all } u \in \mathcal{C}_{\mathbb{C}}(K).$$

It can be shown that

$$\|M_\mu(u)\| = \|u\|_\infty .$$

If the representation $U\colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(H)$ is not topologically cyclic (it is nondegenerate since $\mathcal{C}_{\mathbb{C}}(K)$ is unital), then by Proposition 11.9, the separable Hilbert space H ($H \neq (0)$) is the Hilbert sum of a sequence $(H_n)_{n\geq 1}$ of closed subspaces $H_n \neq (0)$ of H invariant under U, and such that the restriction U_n of U to each H_n is topologically cyclic. But then we can apply Theorem 11.43 to each topologically cyclic representation U_n, so there is a positive measure μ_n associated with H_n such that H_n is isomorphic to $\mathrm{L}^2_{\mu_n}(K;\mathbb{C})$ and U_n is equivalent to M_{μ_n}.

A particularly interesting case for the space K arises if we consider a commutative unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(H)$. In this case, by the Gelfand–Naimark theorem (Theorem 9.37), the Gelfand transform $\mathcal{G}\colon \mathcal{A} \to \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$ is an isometric isomorphism between $\mathcal{A}$ and $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$. Furthermore, $K = \mathsf{X}(\mathcal{A})$ is compact (see Theorem 9.19). But the inverse Gelfand transform $\mathcal{G}^{-1}\colon \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A})) \to \mathcal{A}$ is a representation of $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$ as continuous operators in $\mathcal{L}(H)$, so the results obtained above apply to the representation $U = \mathcal{G}^{-1}$. We obtain Theorem 11.45 which can be viewed as a generalization of the spectral theorem for normal linear maps. Note that since $\mathcal{A}$ is a unital C^*-subalgebra of $\mathcal{L}(H)$, the continuous linear maps in $\mathcal{A}$ are indeed normal. First we need a technical result from the theory of Hilbert spaces.

Proposition 11.44. *Let E and F be two Hilbert spaces, where E is the Hilbert sum of a countable family $(E_n)_{n\in I}$ of closed subspaces of E and F is the Hilbert sum of a countable family $(F_n)_{n\in I}$ of closed subspaces of F (with the same index family I). For each n, let $T_n\colon E_n \to F_n$ be a continuous linear map, and assume that there is a uniform bound $b > 0$ such that $\|T_n\| \leq b$ for all n. Then there is a unique continuous linear map $T\colon E \to F$ whose restriction to E_n is equal to T_n. Furthermore, the restriction of T^* to E_n is equal to T_n^*. So if the T_n are normal, so is T, and if the T_n are unitary, so is T.*

A proof of Proposition 11.44 can be found in Dieudonné [24] (Chapter XV, Section 10, Theorem 15.10.8.1).

Theorem 11.45. *(Spectral Theorem, I) Let $\mathcal{A}$ be a commutative unital C^*-subalgebra of $\mathcal{L}(H)$, with H a separable Hilbert space. There is a measure space $(\Omega, \mathcal{M}, \mu)$, a unitary map $W\colon H \to \mathrm{L}^2_\mu(\Omega, \mathcal{M}; \mathbb{C})$, and an isometric algebra homomorphism $\varphi\colon \mathcal{A} \to \mathrm{L}^\infty_\mu(\Omega, \mathcal{M}; \mathbb{C})$, such that*

$$WTW^{-1} = M_\mu(\varphi(T)) \quad \text{for all } T \in \mathcal{A}\ .$$

Recall that $M_\mu(\varphi(T))(\mathbf{f})$ is the class of $\varphi(T)f$ in $\mathrm{L}^2_\mu(\Omega, \mathcal{M}; \mathbb{C})$ for every $f \in \mathcal{L}^2_\mu(\Omega, \mathcal{M}; \mathbb{C})$. Furthermore, Ω can be taken as a finite or countably infinite disjoint union of copies of $\mathsf{X}(\mathcal{A})$, in such a way that μ is a positive Radon measure μ_n on each copy and $\varphi(T) = \mathcal{G}_T$ on each copy (where $\mathcal{G}_T$ is the Gelfand transform of T).

Proof. First consider the case where $\mathcal{G}^{-1}$ is topologically cyclic. This means that there is some $x_0 \in H$ such that $\{\mathcal{G}^{-1}(u)(x_0) \mid u \in \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))\}$ is dense in H, and since $\mathcal{G}$ is a bijection between $\mathcal{A}$ and $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$, this is equivalent to saying that $\{Tx_0 \mid T \in \mathcal{A}\}$ is dense in H. In this case Theorem 11.43 applies, so there is a positive Radon measure μ on $\mathsf{X}(\mathcal{A})$ and a unitary map $W\colon H \to \mathrm{L}^2_\mu(\mathsf{X}(\mathcal{A}); \mathbb{C})$ such that

$$W\mathcal{G}^{-1}(u)W^{-1} = M_\mu(u) \quad \text{for all } u \in \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A})). \tag{$\dagger_1$}$$

Since $\mathcal{G}$ is a bijection between $\mathcal{A}$ and $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$, every $u \in \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$ is of the form $u = \mathcal{G}_T$ for a unique $T \in \mathcal{A}$, so ($\dagger_1$) is equivalent to

$$WTW^{-1} = M_\mu(\mathcal{G}_T) \quad \text{for all } T \in \mathcal{A}. \tag{$\dagger_2$}$$

If $\mathcal{G}^{-1}$ is not topologically cyclic (it is nondegenerate since $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$ is unital), then by Proposition 11.9, the separable Hilbert space H ($H \neq (0)$) is the Hilbert sum of a finite or countably infinite sequence $(H_n)_{n\in I}$ of closed subspaces $H_n \neq (0)$ of H invariant under $\mathcal{G}^{-1}$, and such that the restriction U_n of $\mathcal{G}^{-1}$ to each H_n is topologically cyclic. The representation U_n is given by $U_n(u)(x) = \mathcal{G}^{-1}(u)(x)$ for every $u \in \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$ and every $x \in H_n$, but since $\mathcal{G}^{-1}(u) = T$ for a unique $T \in \mathcal{A}$, we see that $U_n(u)(x) = T(x) \in H_n$ for all $x \in H_n$ so the restriction of T to H_n is a continuous linear map $T_n\colon H_n \to H_n$. But then we can apply our previous result to the topologically cyclic representation U_n, so there is a positive Radon measure μ_n associated with H_n and a unitary map $W_n\colon H_n \to \mathcal{L}^2_{\mu_n}(\mathsf{X}(\mathcal{A}); \mathbb{C})$ such that

$$W_nT_nW_n^{-1} = M_{\mu_n}(\mathcal{G}_T) \quad \text{for all } T \in \mathcal{A}, \tag{$\dagger_3$}$$

where $T_n\colon H_n \to H_n$ is the restriction of T to H_n. Let $\Omega = \coprod_{n\in I} \mathsf{X}(\mathcal{A})$ be the disjoint union of copies of $\mathsf{X}(\mathcal{A})$, one for each index $n \in I$, and let us denote the nth copy as $\mathsf{X}(\mathcal{A})_n$. We can combine the measures μ_n and the unitary maps W_n to construct a measure μ and a unitary map W as follows. Let $\mathcal{M}$ be the σ-algebra consisting of all sets $E \subseteq \Omega$ such that $E \cap \mathsf{X}(\mathcal{A})_n$ is a Borel set in $\mathsf{X}(\mathcal{A})_n$, and define the measure μ on $(\Omega, \mathcal{M})$ by

$$\mu(E) = \sum_{n\in I} \mu_n(E \cap \mathsf{X}(\mathcal{A})_n).$$

It is easily verified that μ is a measure on Ω, and obviously it is finite on each copy $\mathsf{X}(\mathcal{A})_n$. it is easy to see that the Hilbert sum of the $\mathrm{L}^2_{\mu_n}(\mathsf{X}(\mathcal{A})_n; \mathbb{C})$ is isomorphic to the Hilbert space $\mathrm{L}^2_\mu(\Omega, \mathcal{M}; \mathbb{C})$. Using Proposition 11.44, since $\|W_n\| = 1$ because W_n is a unitary map, there is a unique unitary map W

$$W = \bigoplus_{n \in I} W_n \colon H \to \mathrm{L}^2_\mu(\Omega, \mathcal{M}; \mathbb{C})$$

whose restriction to $\mathsf{X}(\mathcal{A})_n$ is equal to W_n. The maps $\mathcal{G}_T \colon \mathsf{X}(\mathcal{A}) \to \mathbb{C}$ yield a joint map $\coprod_{n \in I} \mathcal{G}_T \colon \Omega \to \mathbb{C}$ defined such that the restriction of $\coprod_{i \in I} \mathcal{G}_T$ to $\mathsf{X}(\mathcal{A})_i$ is equal to $\mathcal{G}_T$, so we obtain a map $\varphi \colon \mathcal{A} \to \mathrm{L}^\infty_\mu(\Omega, \mathcal{M}; \mathbb{C})$, given by $\varphi(T) = \coprod_{n \in I} \mathcal{G}_T$. By construction, for every $f \in \mathcal{L}^2_\mu(\Omega, \mathcal{M}; \mathbb{C})$, we have $M_\mu(\varphi(T))f = \varphi(T)f$, where on each copy $\mathsf{X}(\mathcal{A})_n$, the function $\varphi(T)f$ is equal to the pointwise product of $\mathcal{G}_T$ and f. It remains to verify that the map φ is an isometric algebra homomorphism. This technical fact is proven in Folland [33] (Section 1.4, Lemma 1.46). $\square$

In the special case where the commutative unital C^*-algebra is generated by T, T^* and I, where T is a continuous normal linear map on a separable Hilbert space H, we can be more precise.

Let T be a normal continuous linear map on a Hilbert space H and let $\mathcal{A}_T$ be the subalgebra of $\mathcal{L}(H)$ generated by T, T^* and I. Since T and T^* commute, $\mathcal{A}_T$ is a commutative unital C^*-algebra. Theorem 9.38 asserts that there is an isometric isomorphism $G \colon \mathcal{A}_T \to \mathcal{C}_{\mathbb{C}}(\sigma(T))$ such that

$$G(T) = \mathrm{id}_{\sigma(T)}.$$

As observed in the discussion following Theorem 9.38, the inverse $G^{-1} \colon \mathcal{C}_{\mathbb{C}}(\sigma(T)) \to \mathcal{A}_T$ of the isomorphism $G \colon \mathcal{A}_T \to \mathcal{C}_{\mathbb{C}}(\sigma(T))$ is a representation of $\mathcal{C}_{\mathbb{C}}(\sigma(T))$ in H. Here we should remind the reader that $\sigma(T)$ is the spectrum of T viewed as an element of the unital C^*-algebra $\mathcal{L}(H)$. As we noted just before stating Theorem 9.38, this spectrum is equal to the spectrum of T viewed as an element of the unital C^*-algebra $\mathcal{A}_T$.

Remark: The representation $G^{-1} \colon \mathcal{C}_{\mathbb{C}}(\sigma(T)) \to \mathcal{A}_T$ is often denoted $f \mapsto f(T)$; Dieudonné, Folland, Lang and Rudin use this notation.

The proof of Theorem 11.45 can be adapted to yield the following result (see Taylor [96], Appendix B, for a proof using a different method).

Theorem 11.46. *(Spectral Theorem for Normal Bounded Operators, I) Let T be a normal continuous linear map on a separable Hilbert space H.*

(1) If the representation $G^{-1} \colon \mathcal{C}_{\mathbb{C}}(\sigma(T)) \to \mathcal{A}_T$ is topologically cyclic, then there is a unitary map $W \colon H \to \mathrm{L}^2_\mu(\sigma(T); \mathbb{C})$, such that

$$WTW^{-1} = M_\mu(\mathrm{id}_{\sigma(T)}).$$

(2) *If G^{-1} is not topologically cyclic, then H is the Hilbert sum of a family $(H_n)_{n\in I}$ of closed subspaces of H for some finite or countably infinite index set I, and if $T_n\colon H_n \to H_n$ is the restriction of T to H_n, then for each $n \in I$ there is a positive Radon measure μ_n on $\sigma(T_n)$ and a unitary map $W_n\colon H_n \to \mathrm{L}^2_{\mu_n}(\sigma(T_n);\mathbb{C})$, such that*

$$W_n T_n W_n^{-1} = M_{\mu_n}(\mathrm{id}_{\sigma(T_n)}).$$

If we let $\Omega = \coprod_{n\in I} \sigma(T_n)$ be the disjoint union of the $\sigma(T_n)$, then we can define a σ-algebra $\mathcal{M}$ on Ω, a measure μ on Ω whose restriction to each $\sigma(T_n)$ is equal to μ_n, and a unitary map $W\colon H \to \mathrm{L}^2_\mu(\Omega,\mathcal{M};\mathbb{C})$ such that

$$WTW^{-1} = M_\mu(\mathrm{id}_\Omega).$$

We simply indicate how to prove Part (1) of Theorem 11.46, leaving the proof of Part (2) as an exercise. Since we are assuming that the representation $G^{-1}\colon \mathcal{C}_{\mathbb{C}}(\sigma(T)) \to \mathcal{A}_T$ is topologically cyclic, Theorem 11.43 applies. Therefore, there is a positive Radon measure μ on $\sigma(T)$ and a unitary map $W\colon H \to \mathrm{L}^2_\mu(\sigma(T);\mathbb{C})$ such that

$$WG^{-1}(u)W^{-1} = M_\mu(u) \quad \text{for all } u \in \mathcal{C}_{\mathbb{C}}(\sigma(T)).$$

Since G is a bijection between $\mathcal{A}_T$ and $\mathcal{C}_{\mathbb{C}}(\sigma(T))$ such that $G(T) = \mathrm{id}_{\sigma(T)}$, for $u = G(T)$ we obtain

$$WTW^{-1} = M_\mu(\mathrm{id}_{\sigma(T)}).$$

The following fact is proven in Dieudonné [24] (Chapter XV, Section 11, Proposition 15.11.5).

Proposition 11.47. *The spectrum $\sigma(T)$ of the normal continuous linear map T as above is the closure in $\mathbb{C}$ of the union $\bigcup_{n\in I} \sigma(T_n)$.*

The measures μ_n also determine which scalars $\lambda \in \sigma(T)$ are eigenvalues of T. Recall that a scalar $\lambda \in \mathbb{C}$ is an eigenvalue of T iff $\mathrm{Ker}\,(\lambda\mathrm{id} - T)$ is nontrivial, equivalently iff $\lambda\mathrm{id} - T$ is not injective. If λ is an eigenvalue of T, then $E(T,\lambda) = \mathrm{Ker}\,(\lambda\mathrm{id} - T)$ is called the eigenspace associated with λ, and the nonzero vectors in $E(T,\lambda)$ are the eigenvectors of T associated with λ. On the other hand, the spectrum of T consists of those $\lambda \in \mathbb{C}$ such that $\lambda\mathrm{id} - T$ is *not* invertible,

If H is finite-dimensional, a linear map is not invertible iff it is not injective, so in this case eigenvalues and spectral values coincide. But if H is infinite-dimensional, a linear map may be injective and yet not invertible because it is not surjective. As a consequence, if $\lambda \in \sigma(T)$, the set of eigenvectors associated with λ may be empty. There are continuous linear maps that have no eigenvalues; see Example 9.2.

The following result proven in Dieudonné [24] (Chapter XV, Section 11, Proposition 15.11.6) gives a necessary and sufficient condition for a spectral value to be an eigenvalue. A

similar result is proven in Rudin [80] (Theorem 12.29) in the framework of projection-valued measures, which will be discussed in Section 11.11.

In what follows we use the notation of Theorem 11.46. First, it is easy to see that $\lambda \in \sigma(T)$ is an eigenvalue of T if there is a nonempty subset $J \subseteq I$ such that $\lambda \in \sigma(T_n)$ is an eigenvalue of T_n for all $n \in J$.

Proposition 11.48. *Let T be a normal continuous linear map on a separable Hilbert space H. Using the notation of Theorem 11.46, a scalar $\lambda \in \sigma(T_n)$ is an eigenvalue of T_n iff $\mu_n(\{\lambda\}) \neq 0$. The space spanned by the eigenvectors of T_n associated with λ is a one-dimensional space D_n which is an orthogonal projection of H_n.*

As a corollary of Proposition 11.48, if $\lambda \in \sigma(T)$ is an eigenvalue of T, then there is a finite or countably infinite index set J such that the eigenspace $E(T, \lambda)$ is the Hilbert sum of the one-dimensional spaces D_n (with $n \in J$). Each D_n is spanned by the eigenvectors of T_n associated with λ. If λ and μ are two distinct eigenvalues of T, then $E(T, \lambda)$ and $E(T, \mu)$ are orthogonal.

We state a few more facts whose proof is left as an exercise. A normal continuous linear map is hermitian iff $\sigma(T) \subseteq \mathbb{R}$, unitary iff $\sigma(T) \subseteq \mathbf{U}(1)$.

Finally, a stronger result is obtained if T is a normal (continuous) linear map which is also compact. Recall that this means that the closure of $T(B)$ is compact if B is bounded. It can be shown that the spectrum $\sigma(T)$ of a compact operator is finite or countably infinite, and that the nonzero spectral values are eigenvalues of T. If H is infinite-dimensional, then $0 \in \sigma(T)$; see Lang [62] (Chapter XVII, Section 3). The spaces $E(T, \lambda_n)$, with $\lambda_n \in \sigma(T) - \{0\}$ are finite-dimensional, and together with $\operatorname{Ker} T$ form a Hilbert sum in H. These subspaces are all pairwise orthogonal; see Dieudonné [24] (Chapter XV, Section 11, no 15.11.14) and Folland [33] (Section 1.4, Theorem 1.52).

11.10 Extending Representations from $\mathcal{C}_{\mathbb{C}}(K)$ to $B(K)$

The next crucial step is to realize that a representation $U \colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(H)$ as above determines certain complex Radon measures $\mu_{u,v}$ on K, and that conversely these measures determine U. Then it is possible to extend the representation U of $\mathcal{C}_{\mathbb{C}}(K)$ to the larger commutative unital C^*-algebra $B(K)$ of bounded Borel measurable functions on K. What we gain in doing so is the fact that we can apply the extended representation U to the characteristic functions χ_E of Borel sets E (on K) (the functions χ_E are not continuous), and such operators $U(\chi_E)$ turn out to be orthogonal projections in $\mathcal{L}(H)$. These families of projections have properties that make them *projection-valued measures* (also called *spectral measures*), and such measures can be used to define representations of $B(K)$ that generalize the notion of integral.

For any fixed $u, v \in H$, consider the functional $\Phi_{u,v}$ on $\mathcal{C}_{\mathbb{C}}(K)$ given by

$$\Phi_{u,v}(f) = \langle U(f)(u), v\rangle, \quad f \in \mathcal{C}_{\mathbb{C}}(K).$$

Recall that from Proposition 11.1, we have $\|U(f)\| \leq \|f\|_\infty$, so by Cauchy-Schwarz, we have

$$|\langle U(f)(u), v\rangle| \leq \|U(f)(u)\| \, \|v\| \leq \|U(f)\| \, \|u\| \, \|v\| \leq \|f\|_\infty \, \|u\| \, \|v\| ,$$

so the functional $\Phi_{u,v}$ is bounded (continuous). By Radon–Riesz III, there is a unique complex Radon measure $\mu_{u,v}$ on K such that

$$\langle U(f)(u), v\rangle = \int_K f \, d\mu_{u,v}, \quad f \in \mathcal{C}_{\mathbb{C}}(K). \tag{$*_1$}$$

The measure $\mu_{u,v}$ is often called a *spectral measure*; see Lang [62] (Chapter XX, Section 1). From the definition we have

$$\|\mu_{u,v}\| \leq \|u\| \, \|v\| .$$

The following properties are easy to prove; see Folland [33] (Section 1.4, Proposition 1.34).

Proposition 11.49. *The map from $H \times H$ to $\mathbb{C}$ given by $(u, v) \mapsto \mu_{u,v}$ is sesquilinear. Moreover, $\mu_{v,u} = \overline{\mu_{u,v}}$, and $\mu_{u,u}$ is a positive measure.*

The next step is to extend U to the commutative unital C^*-algebra $B(K)$ of bounded Borel measurable functions on K. This can be done in two ways.

(1) By using Theorem 11.43 for a topologically cyclic representation and the decomposition of H as a Hilbert sum for an arbitrary representation, as explained above. This is the approach followed by Dieudonné [24] (Chapter XV, Section 10).

(2) A faster way is to use the fact that every function $f \in B(K)$ is limit of a sequence of continuous functions f_n converging to f pointwise almost everywhere with respect to the measure $|\mu_{u,v}|$ such that $\|f_n\| \leq \|f\|_\infty$; see Lang [62] (Chapter XX, Section 1), and then to use the dominated convergence theorem (Theorem 5.34).

Using the second approach, we see that

$$\left| \int_K f \, d\mu_{u,v} \right| \leq \|f\|_\infty \, \|u\| \, \|v\| , \tag{$*_2$}$$

so for any fixed $f \in B(K)$ and any fixed u the map $v \mapsto \int_K f \, d\mu_{u,v}$ is a continuous semi-linear form on H, and by the Riesz representation theorem, there is a unique vector $\widetilde{U}(f)(u) \in H$ such that

$$\langle \widetilde{U}(f)(u), v\rangle = \int_K f \, d\mu_{u,v} \quad \text{for all } v \in H.$$

However, $(*_2)$ shows that the map $u \mapsto \widetilde{U}(f)(u)$ is continuous, so indeed $\widetilde{U}(f) \in \mathcal{L}(H)$. Therefore, $\widetilde{U}(f) \in \mathcal{L}(H)$ is completely determined by the equation

$$\langle \widetilde{U}(f)(u), v\rangle = \int_K f \, d\mu_{u,v} \quad \text{for all } u, v \in H \text{ and for all } f \in B(K). \tag{$*_3$}$$

We also have
$$\left\|\widetilde{U}(f)\right\| \leq \|f\|_\infty .$$

It remains to prove that that $\widetilde{U}$ is a representation of $B(K)$. The proof of Theorem 1.36 in Folland [33] applies immediately because all is needed is Proposition 11.49 and Equation $(*_3)$.

Proposition 11.50. *The map $\widetilde{U}\colon B(K) \to \mathcal{L}(H)$ is a representation. Furthermore, if (f_n) is a uniformly bounded sequence of functions in $B(K)$ which converge pointwise to $f \in B(K)$, then*
$$\lim_{n\mapsto\infty} \langle \widetilde{U}(f_n)(u), v\rangle = \langle \widetilde{U}(f)(u), v\rangle \quad \text{for all } u, v \in H.$$
We say that $\widetilde{U}(f_n)$ converges to $\widetilde{U}(f)$ in the weak operator topology.

For the sake of completeness we define three notions of convergence on $\mathcal{L}(H)$, where H is a Hilbert space.

Definition 11.19. Let H be a Hilbert space with inner product $\langle -, -\rangle$ and corresponding norm $\|u\| = \sqrt{\langle u, u\rangle}$ $(u \in H)$. As usual we have the operator norm on $\mathcal{L}(H)$ defined such that for any $f \in \mathcal{L}(H)$,
$$\|f\| = \sup\{\|f(u)\| \mid \|u\| = 1, u \in H\}.$$

We have three notions of convergence corresponding to the following topologies on $\mathcal{L}(H)$:

(1) The *norm topology* on $\mathcal{L}(H)$ is the topology associated with the operator norm. A sequence (f_n) of continuous linear maps $f_n \in \mathcal{L}(H)$ converges to a continuous linear map $f \in \mathcal{L}(H)$ if $\lim_{n\mapsto\infty} \|f - f_n\| = 0$.

(2) The *topology of pointwise convergence* or *strong operator topology*, defined by the family of semi-norms $p_u(f) = \|f(u)\|$, for all $u \in H$, $f \in \mathcal{L}(H)$. A sequence (f_n) of maps $f_n \in \mathcal{L}(H)$ converges to a map $f \in \mathcal{L}(H)$ if
$$\lim_{n\mapsto\infty} \|f(u) - f_n(u)\| = 0 \quad \text{for all } u \in H.$$
This is *pointwise convergence*.

(3) The *weak operator topology*, defined by the family of semi-norms $p_{u,v}(f) = |\langle f(u), v\rangle|$, for all $u, v \in H$, $f \in \mathcal{L}(H)$. A sequence (f_n) of maps $f_n \in \mathcal{L}(H)$ converges to a map $f \in \mathcal{L}(H)$ if
$$\lim_{n\mapsto\infty} \langle f(u) - f_n(u), v\rangle = 0 \quad \text{for all } u, v \in H.$$

This is *weak pointwise convergence*.

From now on, to simplify notation we usually write U instead of $\widetilde{U}$. If we denote by μ the family of complex Radon measures $(\mu_{u,v})_{(u,v)\in H\times H}$, the usual convention is to write

$$U(f) = \int f\, d\mu.$$

Such integrals are often called *weak integrals*.

As we said just before Theorem 11.45, if $\mathcal{A}$ is a commutative unital C^*-subalgebra of $\mathcal{L}(H)$, then the inverse Gelfand transform $\mathcal{G}^{-1}\colon \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A})) \to \mathcal{A}$ is a representation of $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A}))$ as continuous operators in $\mathcal{L}(H)$, and $K = \mathsf{X}(\mathcal{A})$ is compact. Thus the results obtained above apply to the representation $U = \mathcal{G}^{-1}$. Since the Gelfand transform $\mathcal{G}_T$ belongs to $\mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A})) \subseteq B(\mathsf{X}(\mathcal{A}))$ for every $T \in \mathcal{A}$, and since $U(\mathcal{G}_T) = \mathcal{G}^{-1}(\mathcal{G}_T) = T$, Equation $(*_1)$ says that

$$\langle T(u), v\rangle = \int_{\mathsf{X}(\mathcal{A})} \mathcal{G}_T\, d\mu_{u,v} \quad \text{for all } u, v \in H \text{ and for all } T \in \mathcal{A}, \tag{$*_4$}$$

which is also written as

$$T = \int \mathcal{G}_T\, d\mu.$$

As a consequence we obtain a preliminary version of another spectral theorem for a commutative unital C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(H)$.

Theorem 11.51. *Let $\mathcal{A}$ be a commutative unital C^*-subalgebra of $\mathcal{L}(H)$. The extension $U\colon B(\mathsf{X}(\mathcal{A})) \to \mathcal{L}(H)$ of $\mathcal{G}^{-1}\colon \mathcal{C}_{\mathbb{C}}(\mathsf{X}(\mathcal{A})) \to \mathcal{A}$ is a representation. There is a family of complex Radon measures $(\mu_{u,v})_{(u,v)\in H\times H}$ on $\mathsf{X}(\mathcal{A})$ satisfying the properties of Proposition 11.49 such that the following properties hold:*

$$\langle T(u), v\rangle = \int_{\mathsf{X}(\mathcal{A})} \mathcal{G}_T\, d\mu_{u,v} \quad \textit{for all } u, v \in H \textit{ and for all } T \in \mathcal{A},$$

and

$$\langle U(f)(u), v\rangle = \int_{\mathsf{X}(\mathcal{A})} f\, d\mu_{u,v} \quad \textit{for all } u, v \in H \textit{ and for all } f \in B(\mathsf{X}(\mathcal{A}))),$$

where U is the extension of $\mathcal{G}^{-1}$ to $B(\mathsf{X}(\mathcal{A}))$. In short, we write

$$T = \int \mathcal{G}_T\, d\mu, \quad U(f) = \int f\, d\mu.$$

Remark: For the sake of simplicity we omitted to state another property that should be included in Theorem 11.51. This is the fact that for any $S \in \mathcal{L}(H)$, if S commutes with every $T \in \mathcal{A}$, then S commutes with $U(f)$ for every $f \in B(\mathsf{X}(\mathcal{A}))$; see Folland [33] (Proposition 1.36). This property is used to proof Schur's lemma for irreducible unitary representations.

Remarkably, families $(\mu_{u,v})_{(u,v)\in H\times H}$ of measures as above arise from families of projection-valued measures. Such projection-valued measures are defined by the operators $U(\chi_E)$, which are orthogonal projections in $\mathcal{L}(H)$, where E is a Borel set on K.

Proposition 11.52. *Consider the representation $\widetilde{U}\colon B(K) \to \mathcal{L}(H)$ that extends the representation $U\colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(H)$. The map P defined by $P(E) = \widetilde{U}(\chi_E)$, where E is any Borel set in K, has the following properties:*

(1) Each $P(E)$ is an orthogonal projection in $\mathcal{L}(H)$.

(2) $P(\emptyset) = 0$ and $P(K) = I$.

(3) $P(E \cap F) = P(E) \circ P(F)$.

(4) For any family $(E_i)_{i\geq 1}$ of pairwise disjoint Borel sets, we have

$$P\left(\bigcup_{i\geq 1} E_i\right) = \sum_{i\geq 1} P(E_i),$$

which means that if we define F_n and F as $F_n = \bigcup_{i=1}^n E_i$ and $F = \bigcup_{i\geq 1} E_i$, then $\lim_{n\mapsto\infty} \|P(F)(u) - P(F_n)(u)\| = 0$ for all $u \in H$ (convergence in the strong operator topology).

Proof. Since $\chi_E^2 = \chi_E = \overline{\chi_E}$, we have $P(E)^2 = P(E) = P(E)^*$. The equation $P(E)^2 = P(E)$ says that $P(E)$ is a projection, and since it is well-known from linear algebra that $\mathrm{Ker}(P(E))^{\perp} = \mathrm{Im}(P(E)^*)$, we have $\mathrm{Ker}(P(E))^{\perp} = \mathrm{Im}(P(E))$, which means that $\mathrm{Ker}(P(E))$ is the orthogonal complement of $\mathrm{Im}(P(E))$. Property (2) is obvious, and (3) follows from the fact that $\chi_{E\cap F} = \chi_E \chi_F$.

For a finite family $(E_i)_{i=1}^n$ of pairwise disjoint subsets, since

$$\chi_{\bigcup_{i=1}^n E_i} = \sum_{i=1}^n \chi_{E_i},$$

we have

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i).$$

Otherwise, if we write $F_n = \bigcup_{i=1}^n E_i$ and $F = \bigcup_{i\geq 1} E_i$ as above, since the sequence (χ_{F_n}) is uniformly bounded and converges pointwise to χ_F, by Proposition 11.50,

$$\sum_{i=1}^n P(E_i) = P(F_n) \quad \text{converges weakly to} \quad P(F).$$

In particular, $\lim_{n\mapsto\infty} \langle (P(F) - P(F_n))(u), u\rangle = 0$ for all $u \in H$. But F is the disjoint union $F = F_n \cup (F - F_n)$, so $P(F) = P(F_n) + P(F - F_n)$. Since $P(F - F_n)$ is an orthogonal projection, for every $u \in H$, we have

$$\begin{aligned}
\|(P(F) - P(F_n))(u)\|^2 &= \|(P(F - F_n)(u)\|^2 \\
&= \langle P(F - F_n)(u), P(F - F_n)(u)\rangle \\
&= \langle P(F - F_n)(u), u\rangle = \langle (P(F) - P(F_n))(u), u\rangle,
\end{aligned}$$

so weak convergence, which means that $\lim_{n\mapsto\infty}\langle (P(F)-P(F_n))(u), u\rangle = 0$, implies strong convergence, namely $\lim_{n\mapsto\infty}\|(P(F)-P(F_n))(u)\| = 0$. □

Proposition 11.53. *If E and F are two disjoint Borel sets, then the ranges of $P(E)$ and $P(F)$ are orthogonal.*

Proof. Since $P(E)$ and $P(F)$ are orthogonal projection, for any $u, v \in H$, we have

$$\begin{aligned}\langle P(E)(u), P(F)(v)\rangle &= \langle P(F)P(E)(u), v\rangle \\ &= \langle P(E\cap F)(u), v\rangle \\ &= \langle P(\emptyset)(u), v\rangle = \langle 0, v\rangle = 0,\end{aligned}$$

as claimed. □

We can use families of projections on a Hilbert space satisfying the properties of Proposition 11.52 to define families of measures similar to the $\mu_{u,v}$ introduced earlier, and using these measures, to also define representations.

11.11 Projection-Valued Measures and Representations

Let $(\Omega, \mathcal{M})$ be a measure space, where $\mathcal{M}$ is a σ-algebra on the set Ω (since we use the notation $\mathcal{A}$ to denote an algebra, to avoid a clash of notation we denote a σ-algebra by $\mathcal{M}$, departing from our earlier notation).

Definition 11.20. Given a measure space $(\Omega, \mathcal{M})$ and a Hilbert space H, a *projection-valued measure* is a map $P\colon \mathcal{M}\to\mathcal{L}(H)$ assigning an orthogonal projection $P(E)$ in $\mathcal{L}(H)$ to every set $E\in\mathcal{M}$ such that the properties of Proposition 11.52 hold, namely:

(1) Each $P(E)$ is an orthogonal projection in $\mathcal{L}(H)$ (which means that $P(E)^2 = P(E) = P(E)^*$).

(2) $P(\emptyset) = 0$ and $P(\Omega) = I$.

(3) $P(E\cap F) = P(E)\circ P(F)$.

(4) For any family $(E_i)_{i\geq 1}$ of pairwise disjoint sets in $\mathcal{M}$, we have

$$P\left(\bigcup_{i\geq 1} E_i\right) = \sum_{i\geq 1} P(E_i),$$

which means that if we define F_n and F as $F_n = \bigcup_{i=1}^n E_i$ and $F = \bigcup_{i\geq 1} E_i$, then $\lim_{n\mapsto\infty}\|P(F)(u) - P(F_n)(u)\| = 0$ for all $u\in H$ (convergence in the strong operator topology).

We can now define analogs of the measures $\mu_{u,v}$.

Definition 11.21. Let P be a projection-valued measure of a measure space $(\Omega, \mathcal{M})$ in a Hilbert space H. For all $u, v \in H$, define $P_{u,v}(E)$ as

$$P_{u,v}(E) = \langle P(E)(u), v\rangle$$

for all $E \in \mathcal{M}$.

Properties (2) and (4) of Definition 11.20 imply that each $P_{u,v}$ is a complex measure on Ω. The map $(u, v) \mapsto P_{u,v}$ is obviously sesquilinear. Since $P(E)^* = P(E)$, we have

$$P_{v,u}(E) = \langle P(E)(v), u\rangle = \langle v, P(E)(u)\rangle = \overline{\langle P(E)(u), v\rangle} = \overline{P_{u,v}(E)},$$

so $P_{v,u} = \overline{P_{u,v}}$. Since $P(E)^* = P(E) = P(E)^2$, we also have

$$P_{u,u}(E) = \langle P(E)(u), u\rangle = \langle P(E)^2(u), u\rangle = \langle P(E)(u), P(E)^*(u)\rangle = \langle P(E)(u), P(E)(u)\rangle \geq 0$$

so each $P_{u,u}$ is a positive measure. Furthermore,

$$\|P_{u,u}\| = P_{u,u}(\Omega) = \langle P(\Omega)(u), u\rangle = \langle u, u\rangle = \|u\|^2 .$$

Finally, if $B(\Omega, \mathcal{M})$ denotes the space of bounded measurable functions on Ω, we will show that for all $u, v \in H$ and every $f \in B(\Omega, \mathcal{M})$, there is a unique continuous operator $U(f) \in \mathcal{L}(H)$ such that

$$\langle U(f)(u), v\rangle = \int_\Omega f\, dP_{u,v} \quad \text{for all } u, v \in H \text{ and for all } f \in B(\Omega, \mathcal{M}).$$

Without any other assumptions, $B(\Omega, \mathcal{M})$ is simply a commutative unital algebra (under pointwise multiplication). We can think of $U(f)$ as a generalized integral,

$$U(f) = \int f\, dP.$$

Since $\|P_{u,u}\| = \|u\|^2$, for any $f \in B(\Omega, \mathcal{M})$, we have

$$\left| \int f\, dP_{u,u} \right| \leq \|f\|_\infty \|P_{u,u}\| = \|f\|_\infty \|u\|^2 .$$

By linear algebra, we have the polarization identity

$$\begin{aligned}\langle P(E)(u), v\rangle = \langle P(E)(u), P(E)(v)\rangle = \frac{1}{4}(&\|P(E)(u+v)\|^2 - \|P(E)(u-v)\|^2 \\ &+ i(\|P(E)(u+iv)\|^2 - \|P(E)(u-iv)\|^2)),\end{aligned}$$

so

$$P_{u,v}(E) = \frac{1}{4}(P_{u+v,u+v}(E) - P_{u-v,u-v}(E) + i(P_{u+iv,u+iv}(E) - P_{u-iv,u-iv}(E)).$$

Note that all measures on the right-hand side are positive real measures. As a consequence,

$$\int f\, dP_{u,v} = \frac{1}{4}\left(\int f\, dP_{u+v,u+v} - \int f\, dP_{u-v,u-v} + i\left(\int f\, dP_{u+iv,u+iv} - \int f\, dP_{u-iv,u-iv}\right)\right),$$

and so, for all $u, v \in H$ such that $\|u\| = \|v\| = 1$, we have

$$\left|\int f\, dP_{u,v}\right| \le \frac{1}{4}\, \|f\|_\infty\, (\|u+v\|^2 + \|u-v\|^2 + \|u-iv\|^2 + \|u-iv\|^2) \le 4\, \|f\|_\infty\, .$$

Replacing u by $u/\, \|u\|$ and v by $v/\, \|v\|$ $(u, v \neq 0)$, we obtain

$$\left|\int f\, dP_{u,v}\right| \le 4\, \|f\|_\infty\, \|u\|\, \|v\|\, . \tag{$*_5$}$$

As a consequence, for f and u fixed we obtain a continuous semi-linear form on H, so by the Riesz representation theorem there is a unique $U(f)(u) \in H$ such that

$$\langle U(f)(u), v\rangle = \int f\, dP_{u,v} \quad \text{for all } u \in H,$$

but the map $u \mapsto U(f)(u)$ is also continuous, so we have a unique linear map $U(f) \in \mathcal{L}(H)$ such that

$$\langle U(f)(u), v\rangle = \int f\, dP_{u,v} \quad \text{for all } u, v \in H \text{ and all } f \in B(\Omega, \mathcal{M}). \tag{$*_6$}$$

It is customary to write

$$U(f) = \int f\, dP.$$

Remark: If f is a step function $f = \sum_{i=1}^n c_j \chi_{E_i}$ (with $c_j \in \mathbb{C}$), then

$$\int f\, dP_{u,v} = \sum_{i=1}^n c_i P_{u,v}(E_i) = \sum_{i=1}^n c_i \langle P(E_i)(u), v\rangle = \left\langle \sum_{i=1}^n c_i P(E_i)(u), v\right\rangle,$$

so in this case we have

$$\int f\, dP = \sum_{i=1}^n c_i P(E_i),$$

which is reassuring! Using the above fact and the definition of the integral using limits of step functions, the following result proven in Folland [33] (Theorem 1.43) shows that the map $f \mapsto U(f)$ as defined by $(*_6)$ is a representation of algebras. The preceding discussion and this last fact are combined in the following important theorem.

Theorem 11.54. *Let P be a projection-valued measure on a Hilbert space H. For every $f \in B(\Omega, \mathcal{M})$, there is a unique linear map $U(f) \in \mathcal{L}(H)$ such that*

$$\langle U(f)(u), v\rangle = \int f\, dP_{u,v} \quad \text{for all } u, v \in H \text{ and all } f \in B(\Omega, \mathcal{M}).$$

For short we write

$$U(f) = \int f\, dP.$$

The map $U\colon B(\Omega, \mathcal{M}) \to \mathcal{L}(H)$ is an isometric representation of algebras (it is linear and a multiplicative homomorphism). Moreover,

$$\|U(f)(u)\|^2 = \int |f|^2\, dP_{u,u} \quad \text{for all } u \in H \text{ and all } f \in B(\Omega, \mathcal{M}).$$

Remark: Rudin [80] defines the notion of *resolution of the identity*, which is different from the notion of projection-valued measure, but essentially equivalent. A resolution of the identity need not satisfy Condition (4) of Definition 11.20, but it is required that the $P_{u,v}$ as defined in Definition 11.21 are complex measures. Rudin also proves Theorem 11.54 in terms of resolutions of the identity; see Rudin [80], Theorem 12.21.

Let us now return to the case of a representation $U\colon \mathcal{C}_{\mathbb{C}}(K) \to \mathcal{L}(H)$ and its extension $\widetilde{U}\colon B(K) \to \mathcal{L}(H)$ where K is compact, and let $\Omega = K$ and let $\mathcal{M}$ be the σ-algebra generated by the open sets in K. Since in this case $P(E) = \widetilde{U}(\chi_E)$, we have

$$P_{u,v}(E) = \langle P(E)(u), v\rangle = \langle \widetilde{U}(\chi_E)(u), v\rangle = \int_K \chi_E\, d\mu_{u,v} = \mu_{u,v}(E),$$

for all E, so we deduce that

$$P_{u,v} = \mu_{u,v}.$$

In particular, these are Radon measures, so they are regular. What this shows is that any representation $\widetilde{U}\colon B(K) \to \mathcal{L}(H)$ where K is compact arises from some regular projection-valued measure on a Hilbert space H.

As an application, if $\mathcal{A}$ is a commutative unital C^*-subalgebra of $\mathcal{L}(H)$, we showed that the inverse Gelfand transform $\mathcal{G}^{-1}\colon \mathcal{C}(\mathsf{X}(\mathcal{A}); \mathbb{C}) \to \mathcal{A}$ is a representation of $\mathcal{C}(\mathsf{X}(\mathcal{A}); \mathbb{C})$ as continuous operators in $\mathcal{L}(H)$. Here we have $K = \mathsf{X}(\mathcal{A})$. Thus we obtain a version of the spectral theorem for a commutative unital C^*-subalgebra of $\mathcal{L}(H)$.

Theorem 11.55. *(Spectral Theorem, II) Let $\mathcal{A}$ be a commutative unital C^*-subalgebra of $\mathcal{L}(H)$. There is a regular projection-valued measure P on $\mathsf{X}(\mathcal{A})$ such that the following properties hold:*

$$\langle T(u), v\rangle = \int_{\mathsf{X}(\mathcal{A})} \mathcal{G}_T\, dP_{u,v} \quad \text{for all } u, v \in H \text{ and for all } T \in \mathcal{A},$$

and

$$\langle U(f)(u), v\rangle = \int_{\mathsf{X}(\mathcal{A})} f\, dP_{u,v} \quad \text{for all } u, v \in H \text{ and for all } f \in B(\mathsf{X}(\mathcal{A})),$$

where U is the extension of $\mathcal{G}^{-1}$ to $B(\mathsf{X}(\mathcal{A}))$. In short, we write

$$T = \int \mathcal{G}_T\, dP, \quad U(f) = \int f\, dP.$$

In fact, it can be shown that the projection-valued measure P as above is unique; see Folland [33] (Theorem 1.44). Theorem 11.55 is also proven by Rudin in terms of resolutions of the identity; see Rudin [80], Theorem 12.22.

As an application of Theorem 11.55 we obtain another version of the spectral theorem for normal continuous linear maps on a Hilbert space H.

Let T be a normal continuous linear map on a Hilbert space H and let $\mathcal{A}_T$ be the commutative unital C^*-algebra. of $\mathcal{L}(H)$ generated by T, T^* and I. Recall that Theorem 9.38 asserts that there is an isometric isomorphism $G\colon \mathcal{A}_T \to \mathcal{C}_{\mathbb{C}}(\sigma(T))$ such that

$$G(T) = \mathrm{id}_{\sigma(T)}$$

and that the inverse $G^{-1}\colon \mathcal{C}_{\mathbb{C}}(\sigma(T)) \to \mathcal{A}_T$ of G is a representation of $\mathcal{C}_{\mathbb{C}}(\sigma(T))$ in H. By applying Theorem 11.55 to the representation G^{-1} we obtain the following result.

Theorem 11.56. *(Spectral Theorem for Normal Bounded Operators, II) Let T be a normal continuous linear map on a Hilbert space H. There is a unique regular projection-valued measure P on $\sigma(T)$ such that*

$$\langle T(u), v\rangle = \int_{\sigma(T)} \mathrm{id}\, dP_{u,v} \quad \text{for all } u, v \in H.$$

In short, we write

$$T = \int \mathrm{id}\, dP.$$

Some authors (Rudin [80], Lax [64]) call the above result a *spectral decomposition* (or *resolution*) of T. Theorem 11.56 is proven by Rudin in terms of resolutions of the identity; see Rudin [80], Theorem 12.23.

Theorems 11.46 and 11.56 constitute two ways of generalizing the spectral theorem for normal linear maps on a finite-dimensional Hilbert space. The careful reader will notice that Theorem 11.56 holds even if the Hilbert space is not separable. Theorem 11.46 can be generalized to nonseparable Hilbert spaces at the expense of using uncountable Hilbert sums. Also observe that the projection-valued measure P in Theorem 11.56 is uniquely determined by T, whereas the measure space $(\Omega, \mathcal{M}, \mu)$ and the unitary map W of Theorem 11.46 are not.

The usefulness of projection-valued measures becomes more apparent when we generalize Theorem 11.55 to representations of arbitrary commutative unital involutive Banach algebras.

Theorem 11.57. *(Spectral Theorem, III) Let $\mathcal{A}$ be any commutative unital involutive Banach algebra. For any representation $U\colon \mathcal{A} \to \mathcal{L}(H)$ of $\mathcal{A}$ in a Hilbert space H, there is a regular projection-valued measure P on $\mathsf{X}(\mathcal{A})$ such that*

$$\langle U(a)(u), v\rangle = \int_{\mathsf{X}(\mathcal{A})} \mathcal{G}_a \, dP_{u,v} \quad \textit{for all } u, v \in H \textit{ and for all } a \in \mathcal{A},$$

which is abbreviated as

$$U(a) = \int \mathcal{G}_a \, dP, \quad a \in \mathcal{A}.$$

Sketch of proof. A more complete proof is given in Folland [33] (Section 1.5, Theorem 1.53). The key idea is to consider the closure $\mathcal{B}$ of $U(\mathcal{A})$ in $\mathcal{L}(H)$, because it is a commutative unital C^*-subalgebra of $\mathcal{L}(H)$, and so Theorem 11.55 applies to $\mathcal{B}$. Then there is a projection-valued measure P_0 on $\mathsf{X}(\mathcal{B})$ and we need to pull it back to $\mathsf{X}(\mathcal{A})$. Let us provide some details.

The map $U\colon \mathcal{A} \to \mathcal{B}$ induces a continuous map $U^*\colon \mathsf{X}(\mathcal{B}) \to \mathsf{X}(\mathcal{A})$ given by

$$U^*(h) = h \circ U, \quad \text{for all } h \in \mathsf{X}(\mathcal{B}).$$

Recall that for any $a \in \mathcal{A}$, we have $U(a) \in U(\mathcal{A}) \subseteq \mathcal{B} \subseteq \mathcal{L}(H)$ and $h\colon \mathcal{B} \to \mathbb{C}$, so we have $h \circ U\colon \mathcal{A} \to \mathbb{C}$. We claim that U^* is injective. Indeed, if $U^*(h_1) = U^*(h_2)$, then h_1 and h_2 (both in $\mathsf{X}(\mathcal{B})$) agree on $U(\mathcal{A}) \subseteq \mathcal{B}$, and since $\mathcal{B}$ is the closure of $U(\mathcal{A})$ in $\mathcal{L}(H)$, we must have $h_1 = h_2$. But $\mathsf{X}(\mathcal{B})$ is compact, so the injective continuous map U^* is a homeomorphism onto its image, which is compact in $\mathsf{X}(\mathcal{A})$.

As we said before, $\mathcal{B}$ is a commutative unital C^*-subalgebra of $\mathcal{L}(H)$, and so Theorem 11.55 applies to $\mathcal{B}$. By Theorem 11.55, there is a regular projection-valued measure P_0 on $\mathsf{X}(\mathcal{B})$ such that

$$T = \int \mathcal{G}_T \, dP_0 \quad \text{for all } T \in \mathcal{B}. \tag{$\dagger_1$}$$

We use U^* to define a regular projection-valued measure P on $\mathsf{X}(\mathcal{A})$ as follows: for every Borel set E on $\mathsf{X}(\mathcal{A})$, let

$$P(E) = P_0((U^*)^{-1}(E)),$$

where, $P_0((U^*)^{-1}(E))$ is really $P_0((U^*)^{-1}(E \cap U^*(\mathsf{X}(\mathcal{B}))))$. We leave it as an exercise to check that P is indeed a regular projection-valued measure on $\mathsf{X}(\mathcal{A})$.

Finally, observe that the Gelfand transforms of $\mathcal{B}$ and $\mathcal{A}$ are related as follows:

$$\mathcal{G}_{U(a)}(h) = \mathcal{G}_a(U^*(h)) \quad \text{for all } a \in \mathcal{A} \text{ and all } h \in \mathsf{X}(\mathcal{B}),$$

since

$$\mathcal{G}_{U(a)}(h) = h(U(a)) = U^*(h)(a) = \mathcal{G}_a(U^*(h)).$$

Then, since $U(a) \in \mathcal{B}$, by ($\dagger_1$), we have

$$U(a) = \int \mathcal{G}_{U(a)}(h) \, dP_0(h) = \int \mathcal{G}_a(U^*(h)) \, dP_0(h) = \int \mathcal{G}_a \, dP,$$

where the last equation is obtained by going back to the definitions of $\int \mathcal{G}_a(U^*(h))\, dP_0(h)$ and $\int \mathcal{G}_a\, dP$ in terms of the inner product on H and using the definition of P in terms of P_0. $\square$

The projection-valued mesure in Theorem 11.57 is unique; see Folland [33] (Section 1.5). Theorem 11.57 can be promoted to nonunital commutative involutive Banach algebras as long as the representation U is nondegenerate; see Folland [33] (Section 1.5, Theorem 1.54).

Theorem 11.58. *(Spectral Theorem, IV) Let $\mathcal{A}$ be any commutative involutive Banach algebra. For any nondegenerate representation $U\colon \mathcal{A} \to \mathcal{L}(H)$ of $\mathcal{A}$ in a Hilbert space H, there is a unique regular projection-valued measure P on $\mathsf{X}(\mathcal{A})$ such that*

$$\langle U(a)(u), v\rangle = \int_{\mathsf{X}(\mathcal{A})} \mathcal{G}_a\, dP_{u,v} \quad \textit{for all } u, v \in H \textit{ and for all } a \in \mathcal{A},$$

which is abbreviated as

$$U(a) = \int \mathcal{G}_a\, dP, \quad a \in \mathcal{A}.$$

The above theorem is crucial to the proof of Theorem 12.17 characterizing the unitary representations of an *abelian* locally compact group. Intuitively, the characters of G are glued by a suitable projection-valued measure. In turn Theorem 12.17 is a key result used in Mackey's theory for constructing induced representations; see Chapter 16, Proposition 16.1.

For any locally compact space X, $\mathcal{C}_0(X;\mathbb{C})$ is a nonunital C^*-algebra, and since by Proposition 9.22, $\mathsf{X}(\mathcal{C}_0(X;\mathbb{C}))$ is homeomorphism to X itself, we can view the isometric isomorphism from $\mathcal{C}_0(X;\mathbb{C})$ to $\mathcal{C}_0(\mathsf{X}(\mathcal{C}_0(X;\mathbb{C}));\mathbb{C})$ provided by the Gelfand transform (by Gelfand–Naimark) as the identity. In other words, we can view the Gelfand transform on $\mathcal{C}_0(X;\mathbb{C})$ as the identity. Then we have the following corollary of Theorem 11.58.

Theorem 11.59. *Let X be a locally compact space, and let $U\colon \mathcal{C}_0(X;\mathbb{C}) \to \mathcal{L}(H)$ be a nondegenerate representation of $\mathcal{C}_0(X;\mathbb{C})$ in a Hilbert space H. There is a unique regular projection-valued measure P on X such that*

$$\langle U(f)(u), v\rangle = \int_X f\, dP_{u,v} \quad \textit{for all } u, v \in H \textit{ and for all } f \in \mathcal{C}_0(X;\mathbb{C}),$$

which is abbreviated as

$$U(f) = \int f\, dP, \quad f \in \mathcal{C}_0(X;\mathbb{C}).$$

The above theorem is used in Section 16.2 to give an alternate definition of a system of imprimitivity; see Definition 16.4.

Chapter 12

Unitary Representations of Locally Compact Groups

In this chapter we discuss representations of locally compact groups. For simplicity, we begin with finite-dimensional representations, which are continuous group homomorphisms $U\colon G \to \mathbf{GL}(V)$, where V is a finite-dimensional complex vector space (see Section 12.1). Next we consider unitary representations, which are certain kinds of continuous homomorphisms $U\colon G \to \mathbf{U}(H)$, where H is a Hilbert space (typically separable), and $\mathbf{U}(H)$ is the group of unitary operators on H, that is, the continuous linear maps $f\colon H \to H$ that have a continuous inverse, and preserve the inner product; that is,

$$\langle f(x), f(y)\rangle = \langle x, y\rangle \qquad \text{for all } x, y \in H.$$

Then a unitary operator is a continuous linear map $f\colon H \to H$ such that $f^{-1} = f^*$, where f^* is the adjoint of f, the unique continuous linear map determined by the equation

$$\langle f^*(x), y\rangle = \langle x, f(y)\rangle \qquad \text{for all } x, y \in H.$$

The basic theory of unitary representations is discussed in Section 12.2.

There are three main results in this chapter.

The main first main result (first shown by Naimark) is that every unitary representation $U\colon G \to \mathbf{U}(H)$ of a locally compact group G defines a nondegenerate representation $U_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of the involutive Banach algebra $\mathrm{L}^1(G)$, and that conversely, for every nondegenerate representation $V\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of $\mathrm{L}^1(G)$, there is a unique unitary representation $U\colon G \to \mathbf{U}(H)$ of the group G such that $V = U_{\mathrm{ext}}$ (see Section 12.3, Theorem 12.14 and Theorem 12.15).

The bijection $U \mapsto U_{\mathrm{ext}}$ between unitary representations of a locally compact group G and nondegenerate representations of the algebra $\mathrm{L}^1(G)$ is a basic tool that allows the transfer of results about representations of algebras to representations of groups, and vice-versa. It will play a crucial role in the proof of the Peter–Weyl theorem.

The second main result (Theorem 12.17) is a characterization of the unitary representations $U\colon G \to \mathbf{U}(H)$ of a locally compact *abelian* group G in terms of projection-valued measures (as discussed in Section 11.11). This theorem plays a key role in the construction of induced representations using a method due to Mackey (the "Mackey machine"); see Chapter 16.

The third main result (Gelfand and Raikov, Godement) is that there is one-to-one correspondence between unitary cyclic representations of a locally compact group G and certain bounded continuous functions on G called functions of *positive type*.

Let $U\colon G \to \mathbf{U}(H)$ be a unitary representation of the locally compact group G in a Hilbert space H, let x_0 be any vector in H, and define the map $p = \psi_{U,x_0}$ by

$$p(s) = \psi_{U,x_0}(s) = \langle U(s)(x_0), x_0\rangle, \qquad s \in G.$$

It turns out that the function p is continuous and bounded and that it satisfies the following property:

$$\int (f^* * f)(s)p(s)\, d\lambda(s) \geq 0 \qquad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G),$$

where λ is a left Haar measure on G. Such functions are called functions of *positive type*. Remarkably, every continuous function p of positive type determines a unitary topologically cyclic representation U with a cyclic vector x_0, such that $p = \psi_{U,x_0}$ (see Theorem 12.19). The connection between cyclic unitary representations and functions of positive type is discussed in Section 12.5.

In Section 12.6 we present the Gelfand–Raikov theorem without proof (see Theorem 12.23). Informally, this theorem says that there is vast supply of irreducible unitary representations for any locally compact group. This is far from obvious a priori. For example, $\mathbf{SL}(2,\mathbb{R})$ does not have finite-dimensional unitary representations, and it is not that easy to find irreducible unitary representations.

Section 12.7 is devoted to measures of positive type, which generalize functions of positive type. A complex or σ-Radon measure μ is of positive type if

$$\int (f^* * f)(s)\, d\mu(s) = \int\int \overline{f(t)} f(ts)\, d\lambda(t)\, d\mu(s) \geq 0, \qquad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G).$$

The Dirac measure δ_e is a measure of positive type, and more generally, if ν is a complex measure, then the measure $\overline{\check{\nu}} * \nu$ is of positive type. The main point is that a measure μ of positive type defines a unitary representation U_μ of G in a separable Hilbert space H (see Theorem 12.27). This construction will be used in Section 17.8 to define the Plancherel transform.

Basically all the material of this chapter is presented in a more condensed form in Dixmier's classical book Dixmier [26] (see also the English translation published by the AMS).

12.1 Finite-Dimensional Group Representations

For simplicity, we begin with finite-dimensional representations.

Definition 12.1. Given a locally compact group G and a normed vector space V of dimension n, a *continuous linear representation of G in V of dimension* (or *degree*) n is a group homomorphism $U\colon G \to \mathbf{GL}(V)$, where $\mathbf{GL}(V)$ denotes the group of invertible linear maps from V to itself, such that the following condition holds:

(C) The map $g \mapsto U(g)(u)$ is continuous for every $u \in V$.

The space V, called the *representation space*, may be a real or a complex vector space. If V has a Hermitian (resp. Euclidean) inner product $\langle -, - \rangle$, we say that $U\colon G \to \mathbf{GL}(V)$ is a *continuous unitary representation* if

(U) Every linear map $U(g)$ is an *isometry*, that is,

$$\langle U(g)(u), U(g)(v) \rangle = \langle u, v \rangle, \qquad \text{for all } g \in G \text{ and all } u, v \in V.$$

Thus, a continuous linear representation of G is a map $U\colon G \to \mathbf{GL}(V)$ satisfying Condition (C) as well as the properties:

$$\begin{aligned} U(gh) &= U(g)U(h) \\ U(g^{-1}) &= U(g)^{-1} \\ U(1) &= \mathrm{id}_V \end{aligned}$$

for all $g, h \in G$. If U is a unitary representation, then we also have

$$(U(g))^{-1} = (U(g))^*.$$

If G is a finite group, the continuity requirement is omitted.

To avoid confusion when representations involving diffferent groups arise we denote the space of the representation U by V_U, and so we denote a representation as $U\colon G \to \mathbf{GL}(V_U)$.

Note that a major difference with the notion of representation of an algebra, is that for a group representation $U\colon G \to \mathbf{GL}(V)$, the linear map $U(g)$ *must be invertible* for every $g \in G$. For an algebra representation $U\colon A \to \mathcal{L}(H)$ (where H is a Hilbert space), the linear maps $U(s)$ are generally *not invertible*.

For simplicity of language, we usually abbreviate *continuous linear (or unitary) representation* as *(unitary) representation*. The representation space V is also called a *G-module*, since the representation $U\colon G \to \mathbf{GL}(V)$ is equivalent to the left action $\cdot\colon G \times V \to V$, with $g \cdot v = U(g)(v)$. The representation such that $U(g) = \mathrm{id}_V$ for all $g \in G$ is called the *trivial representation*.

It should be noted that because V is finite-dimensional, the condition that for every $u \in V$, the map $g \mapsto U(g)(u)$ is continuous, is actually equivalent to the fact that the map $g \mapsto U(g)$ from G to $\mathcal{L}(V)$ equipped with the operator norm induced by any norm on V, is continuous.

Indeed, for any basis of V, the fact that the map $g \mapsto U(g)(u)$ is continuous implies that the matrix $(U_{ij}(g))$ representing $U(g)$ consists of continuous functions on G.

Since the space V of a representation $U\colon G \to \mathbf{GL}(V)$ is finite-dimensional, say n, it is often convenient to pick a basis $(e_1, \ldots, e_n)$ of V, and then every invertible linear map $U(g) \in \mathbf{GL}(V)$ is represented by an $n \times n$ matrix that we denote $M_U(g) = (U_{ij}(g))$.[1] We obtain a continuous map $M_U\colon G \to \mathbf{GL}(n, \mathbb{C})$ assigning an invertible $n \times n$ complex matrix $M_U(g) = (U_{ij}(g))$ to $g \in G$ satisfying the properties

$$\begin{aligned} M_U(gh) &= M_U(g)M_U(h) \\ M_U(g^{-1}) &= (M_U(g))^{-1} \\ M_U(1) &= I_n \end{aligned}$$

for all $g, h \in G$. The continuity of M_U is equivalent to the fact that the n^2 functions $g \mapsto U_{ij}(g)$ are continuous. If U is a unitary representation, then we also have

$$(M_U(g))^{-1} = (M_U(g))^*.$$

If G is finite we drop the continuity requirement. Conversely we have the notion of representation in matrix form.

Definition 12.2. Given a locally compact group G a *continuous linear representation of G of dimension (or degree) n in matrix form* is a mapping $M_U\colon G \to \mathbf{GL}(n, \mathbb{C})$ assigning an invertible $n \times n$ complex matrix $M_U(g) = (U_{ij}(g))$ to $g \in G$ satisfying the properties

$$\begin{aligned} M_U(gh) &= M_U(g)M_U(h) \\ M_U(g^{-1}) &= (M_U(g))^{-1} \\ M_U(1) &= I_n \end{aligned}$$

for all $g, h \in G$, and such that the n^2 functions $g \mapsto U_{ij}(g)$ are continuous. If U is a unitary representation, then we also have

$$(M_U(g))^{-1} = (M_U(g))^*.$$

In this case M_U is a homomorphism $M_U\colon G \to \mathbf{U}(n)$. If G is finite we drop the continuity requirement.

A representation in matrix form $M_U\colon G \to \mathbf{GL}(n, \mathbb{C})$ (resp. $M_U\colon G \to \mathbf{U}(n)$) defines the representation $U\colon G \to \mathbf{GL}(\mathbb{C}^n)$ (resp. $U\colon G \to \mathbf{U}(\mathbb{C}^n)$) given by

$$(U(g))(z) = M_U(g)z, \quad z \in \mathbb{C}^n,\ g \in G.$$

[1] To be perfectly rigorous the matrix M_U should be indexed by the basis $\mathcal{E} = (e_1, \ldots, e_n)$, say as $M_U^{\mathcal{E}}$, but this is just too much decoration.

Since the notation $M_U(g)$ is quite heavy, we often write $M(g)$ instead of $M_U(g)$. This is an abuse of notation since $M(g)$ is a linear map and $M_U(g)$ is a matrix representing it in some basis, and thus depends on this basis. We also often identify a matrix representation with the representation associated with it. The same issue arises in linear algebra and we hope that the reader is already familiar with it and will not be confused.

Given any basis $(e_1, \ldots, e_n)$ of V, we may think of the scalar functions $g \mapsto U_{ij}(g)$ as *special functions* on G. As explained in Dieudonné [20] (see also Vilenkin [101]), essentially all special functions (Legendre polynomials, ultraspherical polynomials, Bessel functions *etc.*) arise in this way by choosing some suitable G and V.

Remark: In Chapter 15 we will consider the situation where G is a group not equipped with any topology, and V is a vector space, possibly infinite-dimensional, not equipped with any norm. Then a *linear representation of G in V* is simply a homomorphism $U\colon G \to \mathbf{GL}(V)$, which amounts to dropping Condition (C) from Definition 12.1. However, in this chapter and the next, all representations satisfy Condition (C).

Example 12.1. Consider the group $\mathfrak{S}_3$ of permutations on the set $\{1, 2, 3\}$. There are $3! = 6$ permutations

$$\sigma_1 = (1, 2, 3),\;\; \sigma_2 = (1, 3, 2),\;\; \sigma_3 = (2, 1, 3),\;\; \sigma_4 = (2, 3, 1),\;\; \sigma_5 = (3, 1, 2),\;\; \sigma_6 = (3, 2, 1).$$

The first permutation $\sigma_1 = (1, 2, 3)$ is the identity; the permutations

$$\sigma_2 = (1, 3, 2),\;\; \sigma_3 = (2, 1, 3),\;\; \sigma_6 = (3, 2, 1)$$

are transpositions and thus have negative signature, and the permutations

$$\sigma_4 = (2, 3, 1),\;\; \sigma_5 = (3, 1, 2)$$

are cyclic permutations and thus have positive signature. We obtain a representation $\rho_1\colon \mathfrak{S}_3 \to \mathbf{GL}(\mathbb{C}^3)$ as follows. If (e_1, e_2, e_3) is the canonical basis of $\mathbb{C}^3$, then $\rho_1(\sigma_i)$ is the linear map given by

$$\rho_1(\sigma_i)(e_j) = e_{\sigma_i(j)}, \quad 1 \leq i, j \leq 3.$$

In the basis (e_1, e_2, e_3), the linear maps $\rho_1(\sigma_i)$ are represented by the $3{\times}3$ matrices $M_1, \ldots, M_6$ given by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

This is an example of a permutation representation.

Here is another representation of the group $\mathfrak{S}_3$ in $\mathbb{C}^6$.

Example 12.2. This time we define the representation $\rho_{\mathbf{R}} \colon \mathfrak{S}_3 \to \mathbf{GL}(\mathbb{C}^6)$ as follows. Let $(e_{\sigma_1}, \ldots, e_{\sigma_6})$ be the canonical basis of $\mathbb{C}^6$ indexed by the permutations σ_i ($1 \leq i \leq 6$), and set

$$\rho_{\mathbf{R}}(\sigma_i)(e_{\sigma_j}) = e_{\sigma_i \circ \sigma_j}, \quad 1 \leq i, j \leq 6.$$

Note that the 6×6 matrix representing $\rho_{\mathbf{R}}(\sigma_i)$ in the basis $(e_{\sigma_1}, \ldots, e_{\sigma_6})$ consists of the permutation of the columns of the identity matrix I_6 whose indices are given by the ith row of the multiplication table of the group $\mathfrak{S}_3$. This multiplication table is given by

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 5 & 6 & 3 & 4 \\ 3 & 4 & 1 & 2 & 6 & 5 \\ 4 & 3 & 6 & 5 & 1 & 2 \\ 5 & 6 & 2 & 1 & 4 & 3 \\ 6 & 5 & 4 & 3 & 2 & 1, \end{pmatrix},$$

where we denote σ_i simply by i and where the (i, j) entry represents $\sigma_i \circ \sigma_j$. We obtain the following 6 matrices:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The permutation $\rho_{\mathbf{R}}$ is called the *regular representation* of $\mathfrak{S}_3$.

Example 12.3. For an example involving an infinite group, we describe a class of representations of $G = \mathbf{SL}(2, \mathbb{C})$, the group of complex matrices with determinant $+1$,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad ad - bc = 1.$$

Recall that $\mathcal{P}^{\mathbb{C}}_m(2)$ denotes the vector space of complex homogeneous polynomials of degree m in two variables (z_1, z_2). A complex homogeneous polynomials of degree m in two variables (z_1, z_2) is an expression of the form $P(z_1, z_2) = \sum_{i=0}^m c_i z_1^i z_2^{m-i}$, with $c_i \in \mathbb{C}$. For every matrix $A \in \mathbf{SL}(2, \mathbb{C})$, with

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

for every homogeneous polynomial $P \in \mathcal{P}_m^{\mathbb{C}}(2)$, we define $U_m(A)(P(z_1, z_2))$ by

$$U_m(A)(P(z_1, z_2)) = P(dz_1 - bz_2, -cz_1 + az_2).$$

If we think of the homogeneous polynomial $Q(z_1, z_2)$ as a function $P\binom{z_1}{z_2}$ of the vector $\binom{z_1}{z_2}$, then

$$U_m(A)\left(P\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) = PA^{-1}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = P\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

The expression above makes it clear that

$$U_m(AB) = U_m(A)U_m(B)$$

for any two matrices $A, B \in \mathbf{SL}(2, \mathbb{C})$, so U_m is indeed a representation of $\mathbf{SL}(2, \mathbb{C})$ into $\mathcal{P}_m^{\mathbb{C}}(2)$.

The representations U_m also yield representations of the subgroup $\mathbf{SU}(2)$ of $\mathbf{SL}(2, \mathbb{C})$. Recall that the group $\mathbf{SU}(2)$ consists of all 2×2 complex matrices

$$S = \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix}, \quad a\overline{a} + b\overline{b} = 1.$$

As above, the representation $U_m \colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_m^{\mathbb{C}}(2))$ is given by

$$U_m(S)(P(z_1, z_2)) = P(\overline{a}z_1 - bz_2, \overline{b}z_1 + az_2).$$

It can be shown that $\mathbf{SL}(2, \mathbb{C})$ has *no* nontrivial *unitary* finite-dimensional representations! This is because $\mathbf{SL}(2, \mathbb{C})$ is a connected simple noncompact Lie group with finite center; see Dieudonné [21] (Section 21.6, Problem 5).

There is a natural and useful notion of equivalence of representations.

Definition 12.3. Given any two representations $U_1 \colon G \to \mathbf{GL}(V_1)$ and $U_2 \colon G \to \mathbf{GL}(V_2)$, a *G-map* (or *morphism of representations*) $\varphi \colon U_1 \to U_2$ is a linear map $\varphi \colon V_1 \to V_2$ which is *equivariant*, which means that the following diagram commutes for every $g \in G$:

$$\begin{tikzcd} V_1 \arrow[r, "U_1(g)"] \arrow[d, "\varphi"'] & V_1 \arrow[d, "\varphi"] \\ V_2 \arrow[r, "U_2(g)"'] & V_2, \end{tikzcd}$$

i.e.

$$\varphi \circ U_1(g) = U_2(g) \circ \varphi, \qquad g \in G.$$

The space of all G-maps between two representations as above is denoted $\mathrm{Hom}_G(U_1, U_2)$. Two representations $U_1 \colon G \to \mathbf{GL}(V_1)$ and $U_2 \colon G \to \mathbf{GL}(V_2)$ are *equivalent* iff $\varphi \colon V_1 \to V_2$ is an invertible linear map (which implies that $\dim V_1 = \dim V_2$). In matrix form, the representations $U_1 \colon G \to \mathbf{GL}(n, \mathbb{C})$ and $U_2 \colon G \to \mathbf{GL}(n, \mathbb{C})$ are equivalent iff there is some invertible $n \times n$ matrix P so that

$$U_2(g) = PU_1(g)P^{-1}, \qquad g \in G.$$

If $W \subseteq V$ is a subspace of V, then in some cases, a representation $U\colon G \to \mathbf{GL}(V)$ yields a representation $U\colon G \to \mathbf{GL}(W)$. This is interesting because under certain conditions on G (*e.g.*, G compact) every representation may be decomposed into a "sum" of so-called irreducible representations (defined below), and thus the study of all representations of G boils down to the study of irreducible representations of G; for instance, see Knapp [57] (Chapter 4, Corollary 4.7), or Bröcker and tom Dieck [16] (Chapter 2, Proposition 1.9).

Definition 12.4. Let $U\colon G \to \mathbf{GL}(V)$ be a representation of G. If $W \subseteq V$ is a subspace of V, then we say that W is *invariant* (or *stable*) under U iff $U(g)(w) \in W$, for all $g \in G$ and all $w \in W$. If W is invariant under U, then we have a homomorphism, $U\colon G \to \mathbf{GL}(W)$, called a *subrepresentation* of G. A representation $U\colon G \to \mathbf{GL}(V)$ with $V \neq (0)$ is *irreducible* iff it only has the two subrepresentations $U\colon G \to \mathbf{GL}(W)$ corresponding to $W = (0)$ or $W = V$.

Example 12.4. The representation $\rho_1\colon \mathfrak{S}_3 \to \mathbf{GL}(\mathbb{C}^3)$ of Example 12.1 is reducible. Indeed, the one-dimensional subspace V_1 spanned by $e_1+e_2+e_3$ is invariant under ρ_1 since each $\rho(\sigma_i)$ permutes the indices $1, 2, 3$. The corresponding subrepresentation of $\mathfrak{S}_3$ in V_1 is equivalent to the irreducible trivial representation in $\mathbb{C}$, given by $\rho_{\mathrm{triv}}(\sigma_i) = 1$ $(1 \leq i \leq 6)$. The orthogonal complement V_2 of V_1 is the plane of equation

$$x_1 + x_2 + x_3 = 0,$$

which has $(e_1 - e_2, e_2 - e_3)$ as a basis. It is easy to see that the subspace V_2 is also invariant under ρ_1. It is instructive to find an equivalent representation of ρ_1 in the basis (v_1, v_2, v_3) given by

$$\begin{aligned}
v_1 &= (1/3)(e_1 + e_2 + e_3)\\
v_2 &= (1/3)(e_1 - e_2)\\
v_3 &= (1/3)(e_2 - e_3).
\end{aligned}$$

The change of basis matrix P from the basis (e_1, e_2, e_3) to the basis (v_1, v_2, v_3) is

$$P = \begin{pmatrix} 1/3 & 1/3 & 0\\ 1/3 & -1/3 & 1/3\\ 1/3 & 0 & -1/3 \end{pmatrix},$$

whose inverse is

$$P^{-1} = \begin{pmatrix} 1 & 1 & 1\\ 2 & -1 & -1\\ 1 & 1 & -2 \end{pmatrix}.$$

Using the linear map φ from $\mathbb{C}^3$ to itself given by P^{-1} (which transforms the coordinates of a vector in $\mathbb{C}^3$ over the basis (e_1, e_2, e_3) to the coordinates of this vector over the basis (v_1, v_2, v_3)), we obtain the equivalent representation ρ_1' given by

$$\rho_1'(\sigma_i) = \varphi\rho_1(\sigma_i)\varphi^{-1},$$

and over the basis (v_1, v_2, v_3), the matrices representing the linear maps $\rho_1'(\sigma_i)$ are the matrices $P^{-1}M_iP$ shown below:

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&1&0\\0&1&-1\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&-1&1\\0&0&1\end{pmatrix},$$
$$\begin{pmatrix}1&0&0\\0&0&-1\\0&1&-1\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&-1&1\\0&-1&0\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&0&-1\\0&-1&0\end{pmatrix}.$$

Some of the above matrices are not unitary. We can fix this by choosing an orthonormal basis (w_1, w_2, w_3) with $w_1 = (1/\sqrt{3})v_1$, a basis of V_1, and (w_2, w_3), a basis of V_2. For example we can pick

$$\begin{aligned} w_1 &= (1/\sqrt{3})(e_1 + e_2 + e_3) \\ w_2 &= (1/\sqrt{2})(e_1 - e_2) \\ w_3 &= (1/\sqrt{6})(e_1 + e_2 - 2e_3). \end{aligned}$$

The change of basis matrix Q from the basis (e_1, e_2, e_3) to the basis (w_1, w_2, w_3) is

$$Q = \begin{pmatrix}1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6}\\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6}\\ 1/\sqrt{3} & 0 & -2/\sqrt{6}\end{pmatrix}$$

and $Q^{-1} = Q^\top$. We obtain an equivalent representation $\rho_1''(\sigma_i)$ and over the basis (w_1, w_2, w_3), the unitary matrices representing the linear maps $\rho_1''(\sigma_i)$ are the matrices $Q^{-1}M_iQ$ shown below:

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&1/2&\sqrt{3}/2\\0&\sqrt{3}/2&-1/2\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&-1&0\\0&0&1\end{pmatrix},$$
$$\begin{pmatrix}1&0&0\\0&-1/2&-\sqrt{3}/2\\0&\sqrt{3}/2&-1/2\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&-1/2&\sqrt{3}/2\\0&-\sqrt{3}/2&-1/2\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&1/2&-\sqrt{3}/2\\0&-\sqrt{3}/2&-1/2\end{pmatrix}.$$

It is now clear that the subspace V_1 spanned by w_1 and the subspace V_2 spanned by w_2 and w_3 are invariant. It is not hard to show that the subrepresentation of ρ_1'' in V_2 is irreducible. This representation is usually called the *standard representation* of $\mathfrak{S}_3$; see Fulton and Harris [36], Section 1.3, Thus we have two irreducible representations of $\mathfrak{S}_3$, the second one being two-dimensional. It turns out that $\mathfrak{S}_3$ only has one more irreducible representation. How do we find it? The answer is, as a subrepresentation of the regular representation.

Recall the regular representation $\rho_{\mathbf{R}} \colon \mathfrak{S}_3 \to \mathbf{GL}(\mathbb{C}^6)$ of $\mathfrak{S}_3$ from Example 12.2. The notion of regular representation can be defined for any finite group. Let G be a finite group

with $g = |G|$ elements. We define the representation $\rho_{\mathbf{R}}\colon G \to \mathbf{GL}(\mathbb{C}^g)$ as follows. Let $(e_{s_1}, \ldots, e_{s_g})$ be the canonical basis of $\mathbb{C}^g$ indexed by the g elements of G and set

$$\rho_{\mathbf{R}}(s_i)(e_{s_j}) = e_{s_i s_j}, \quad 1 \leq i, j \leq g.$$

It can be shown that every irreducible finite-dimensional representation $\rho_i\colon G \to \mathbf{GL}(\mathbb{C}^{n_i})$ of the finite group G is equivalent to a subrepresentation of the regular representation $\rho_{\mathbf{R}}\colon G \to \mathbf{GL}(\mathbb{C}^g)$ of G in $\mathbb{C}^g$ (where $g = |G|$), and each irreducible representation $\rho_i\colon G \to \mathbf{GL}(\mathbb{C}^{n_i})$ occurs n_i times in the regular representation; see Proposition 13.17. Furthermore, if there are h irreducible representations $\rho_i\colon G \to \mathbf{GL}(\mathbb{C}^{n_i})$ (up to equivalence), then

$$n_1^2 + \cdots + n_h^2 = g;$$

see Section 13.2, Example 13.2. Moreover the number h of irreducible representations of G (up to equivalence) is equal to the number of conjugacy classes of G; see Section 13.2, Example 13.2. The proof of these standard facts of representation theory can be found in Serre [90], Fulton and Harris [36], Simon [93], Hall [43], or any book on representation theory. We also proved these facts in Section 13.2 (Example 13.2) and in Section 13.3 (Proposition 13.17) as a special case of results applying to compact groups.

If G is finite of order $g = |G|$, if we write $G = \{s_1, \ldots, s_g\}$ and denote the canonical basis vectors of $\mathbb{C}^g$ as $(e_{s_1}, \ldots, e_{s_g})$, then there is an isomorphism between $\mathbb{C}^g$ and the vector space $\mathbb{C}^G$ of functions from G to $\mathbb{C}$ defined such that to every vector $x = z_{s_1} e_{s_1} + \cdots + z_{s_g} e_{s_g}$ in $\mathbb{C}^g$ we assign the function $f_x\colon G \to \mathbb{C}$ given by

$$f_x(s_i) = z_{s_i}, \quad 1 \leq i \leq g.$$

Now by the definition of the regular representation $\rho_{\mathbf{R}}$ of G, we have

$$\rho_{\mathbf{R}}(s_i)(x) = \rho_{\mathbf{R}}(s_i)\left(\sum_{j=1}^{g} z_{s_j} e_{s_j}\right) = \sum_{j=1}^{g} z_{s_j} e_{s_i s_j}, \quad 1 \leq i \leq g. \tag{$*_1$}$$

If we let $s_i s_j = s_k$, then $s_j = s_i^{-1} s_k$, $z_{s_j} e_{s_i s_j} = z_{s_i^{-1} s_k} e_{s_k}$, and the vector $y = \rho_{\mathbf{R}}(s_i)(x) = \sum_{k=1}^{g} z_{s_i^{-1} s_k} e_{s_k}$ corresponds the function f_y given by

$$f_y(s_k) = z_{s_i^{-1} s_k}, \quad 1 \leq k \leq g.$$

Therefore, $(*_1)$ induces the representation $\mathbf{R}\colon G \to \mathbf{GL}(\mathbb{C}^G)$ given by

$$(\mathbf{R}_{s_i}(f))(s_k) = f(s_i^{-1} s_k), \quad f \in \mathbb{C}^G, \ 1 \leq i, k \leq g. \tag{$*_2$}$$

The representation $\mathbf{R}$ is also called the *regular representation* of G in $\mathbb{C}^G$. It is a special case of the notion of regular representation defined in Definition 12.11 for locally compact groups. To be very precise it is the *left regular representation* of G because it acts on the left on functions in $\mathbb{C}^G$. At first glance the term $s_i^{-1} s_k$ may seem wrong, but it is necessary to

use s_i^{-1} instead of s_i to insure that $\mathbf{R}$ is a left action on functions in $\mathbb{C}^G$. We already noticed this fact in Section 8.2, Definition 8.7. There is also a right regular representation defined by

$$(\mathbf{R}^r_{s_i}(f))(s_k) = f(s_k s_i), \quad f \in \mathbb{C}^G,\ 1 \leq i, k \leq g. \tag{$*_3$}$$

Representations as given by $(*_2)$ are said to be representations by *left shifts*, and representations as given by $(*_3)$ are said to be representations by *right shifts*.

Obviously the notion of left regular representation (and right regular representation) makes sense for any group G, finite or infinite, and any subspace $\mathcal{F}$ of the vector space all functions in $\mathbb{C}^G$, namely it is the representation $\mathbf{R}\colon G \to \mathbf{GL}(\mathcal{F})$ given by

$$(\mathbf{R}_s(f))(t) = f(s^{-1}t), \quad f \in \mathcal{F},\ s, t \in G. \tag{$*_4$}$$

If G is an infinite locally compact groups, in this it is necessary to replace the vector space $\mathbb{C}^G$ of the representation by a space of functions defined on G, namely $\mathrm{L}^2_\lambda(G; \mathbb{C})$ (where λ is a left Haar measure on G).

If V has a hermitian inner product, then we can prove that any irreducible linear representation $U\colon G \to \mathbf{GL}(V)$ of a group G, finite or infinite, where U is not assumed to satisfy Condition (C), is equivalent to some (irreducible) subrepresentation $\rho\colon G \to \mathbf{GL}(\mathcal{F})$ of the left regular representation $\mathbf{R}\colon G \to \mathbf{GL}(\mathbb{C}^G)$. The key to the construction is the mapping $\varphi\colon V \to \mathbb{C}^G$ converting a vector $u \in V$ to a function $f_u \in \mathbb{C}^G$ defined as follows: pick any nonzero vector $a \in V$ and define $\varphi(u) = f_u$ by

$$f_u(s) = \langle U(s^{-1})(u), a\rangle, \quad u \in V,\ s \in G. \tag{$*_5$}$$

The reason for using $U(s^{-1})$ is that we want the left regular representation. If we use $U(s)$, then we obtain the right regular representation, as in Vilenkin [101] (Chapter I, Section 2.4). Since $U(s^{-1})$ is a linear map and the inner product is linear in its first argument, the function φ is linear. The trick is to see what is $f_{U(s)(u)}(t)$ $(s, t \in G)$. By definition,

$$\begin{aligned} f_{U(s)(u)}(t) &= \langle U(t^{-1})(U(s)(u)), a\rangle \\ &= \langle U(t^{-1}s)(u), a\rangle \\ &= \langle U((s^{-1}t)^{-1})(u), a\rangle \\ &= f_u(s^{-1}t), \end{aligned}$$

which we record as the equation

$$f_{U(s)(u)}(t) = f_u(s^{-1}t). \tag{$*_6$}$$

Also observe that

$$f_a(e) = \langle U(e^{-1})(a), a\rangle = \langle U(e)(a), a\rangle = \langle a, a\rangle,$$

so $f_a(e) \neq 0$ since $a \neq 0$. Using the above considerations we can prove the following result.

Proposition 12.1. *If V has a hermitian inner product, then any irreducible linear representation $U\colon G \to \mathbf{GL}(V)$ of a finite or infinite group G (where U is not assumed to satisfy Condition (C)) is equivalent to some (irreducible) subrepresentation $\rho\colon G \to \mathbf{GL}(\mathcal{F})$ of the left regular representation $\mathbf{R}\colon G \to \mathbf{GL}(\mathbb{C}^G)$. The linear map $\varphi\colon V \to \mathbb{C}^G$ defined above is injective, $\mathcal{F} = \varphi(V)$, and $\varphi\colon V \to \mathcal{F}$ provides the equivalence between U and ρ.*

Proof. Since $\rho\colon G \to \mathbf{GL}(\mathcal{F})$ is a subrepresentation of the regular representation of G it is given by
$$(\rho(s)(f))(t) = f(s^{-1}t), \quad f \in \mathcal{F},\ s, t \in G.$$
Let us first verify that the diagram
$$\begin{tikzcd} V \arrow[r, "U(s)"] \arrow[d, "\varphi"'] & V \arrow[d, "\varphi"] \\ \mathcal{F} \arrow[r, "\rho(s)"'] & \mathcal{F}, \end{tikzcd}$$
commutes, that is,
$$\varphi(U(s)(u)) = \rho(s)(\varphi(u)), \qquad s \in G, u \in V,$$
which means that
$$\varphi(U(s)(u))(t) = \rho(s)(\varphi(u))(t), \qquad s, t \in G, u \in V.$$
By definition of φ and $(*_6)$,
$$\varphi(U(s)(u))(t) = f_{U(s)(u)}(t) = f_u(s^{-1}t),$$
and
$$\rho(s)(\varphi(u))(t) = (\rho(s)(f_u))(t) = f_u(s^{-1}t),$$
which verifies the commutativity of the diagram. Consequently, as φ is surjective on $\mathcal{F}$ by definition, it suffices to prove that φ is injective to conclude that φ is an equivalence between U and ρ. Let $\mathcal{K}$ be the kernel of φ. We prove that $\mathcal{K}$ is invariant under U. For any $u \in V$ we have $u \in \mathcal{K}$ iff $f_u = 0$ iff $f_u(t) = 0$ for all $t \in G$ iff $f_u(s^{-1}t) = 0$ for all $t \in G$ and all $s \in G$ (since for fixed s, the map $t \mapsto s^{-1}t$ is a bijection of G), which by $(*_6)$ is equivalent to $f_{U(s)(u)}(t) = 0$ for all $t \in G$ iff $U(s)(u) \in \mathcal{K}$ for all $s \in G$. So $\mathcal{K}$ is indeed invariant. Since U is irreducible, either $\mathcal{K} = (0)$ of $\mathcal{K} = V$. But we observed earlier that $f_a(e) \neq 0$ so $\mathcal{K} = V$ is impossible since it means that $f_u(t) = 0$ for all $u \in V$ and all $t \in G$. Therefore, $\mathcal{K} = (0)$ and the map φ in injective. $\square$

If V is a Hilbert space, U is a unitary representation, all functions f_u belong to $\mathrm{L}^2(G)$, and the map $\varphi\colon V \to \mathrm{L}^2(G)$ is well-behaved, then ρ is a unitary subrepresentation of the regular representation of G in $\mathrm{L}^2(G)$. This is the case for compact groups; see Proposition 13.17. Proposition 12.1 implies that if G is finite, since $\mathbb{C}^G$ is isomorphic to $\mathbb{C}^{|G|}$, then the dimension of the vector space V involved in an irreducible representation of G is at most the cardinality of G.

We now return to the regular representation of Example 12.2.

Example 12.5. It is easy to see that the symmetric group has three conjugacy classes, $\{\sigma_1\}$, $\{\sigma_2, \sigma_3, \sigma_6\}$ and $\{\sigma_4, \sigma_5\}$, so it has three irreducible representations. Going back to the regular representation $\rho_{\mathbf{R}} \colon \mathfrak{S}_3 \to \mathbf{GL}(\mathbb{C}^6)$, we see that the one-dimensional subspace V_1 spanned by $e_1 + e_2 + e_3 + e_4 + e_5 + e_6$ is invariant so the representation $\rho_{\mathbf{R}}$ is reducible. The subrepresentation of $\rho_{\mathbf{R}}$ in V_1 is equivalent to the trivial representation, which is irreducible. Although this is not obvious, there is another one-dimensional irreducible representation, which is the representation induced by the signature function ϵ on permutations. Recall that for any permutation σ, its signature $\epsilon(\sigma)$ is $+1$ if σ is the composition of an even number of transpositions, -1 if it is the composition of an odd number of transpositions. The map $\epsilon \colon \mathfrak{S}_n \to \mathbb{C}$ is a homomorphism and it yields the irreducible representation $\rho_\epsilon \colon \mathfrak{S}_n \to \mathbf{U}(1)$ given by

$$(\rho_\epsilon(\sigma))(z) = \epsilon(\sigma)z, \quad z \in \mathbb{C}.$$

Then we see that the subspace V_2 spanned by the vector $e_1 - e_2 - e_3 + e_4 + e_5 - e_6$ (which corresponds to the signatures $+1, -1, -1, +1, +1, -1$ of the permutations $\sigma_1, \ldots, \sigma_6$) is invariant under $\rho_{\mathbf{R}}$, and the subrepresentation of $\rho_{\mathbf{R}}$ to V_2 is equivalent to the irreducible representation ρ_ϵ. The orthogonal complement V_3 of $V_1 \oplus V_2$ is the intersection of the two hyperplanes in $\mathbb{C}^6$ given by the equations

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 &= 0 \\ x_1 - x_2 - x_3 + x_4 + x_5 - x_6 &= 0, \end{aligned}$$

a subspace of dimensional 4. By adding and subtracting these equations we see that the subspace V_3 is also defined by the equations

$$\begin{aligned} x_1 + x_4 + x_5 &= 0 \\ x_2 + x_3 + x_6 &= 0. \end{aligned}$$

We can prove directly that V_3 is invariant under $\rho_{\mathbf{R}}$, but since the representation $\rho_{\mathbf{R}}$ is actually unitary, we prefer using results form the next section.

An easy but crucial lemma about irreducible representations is "Schur's Lemma."

Lemma 12.2. *(Schur's Lemma) Let $U_1 \colon G \to \mathbf{GL}(V)$ and $U_2 \colon G \to \mathbf{GL}(W)$ be any two real or complex representations of a group G. If U_1 and U_2 are irreducible, then the following properties hold:*

(i) Every G-map $\varphi \colon U_1 \to U_2$ is either the zero map or an isomorphism.

(ii) If U_1 is a complex representation, then every G-map $\varphi \colon U_1 \to U_1$ is of the form $\varphi = \lambda \mathrm{id}$, for some $\lambda \in \mathbb{C}$.

Proof. (i) Observe that the kernel $\mathrm{Ker}\ \varphi \subseteq V$ of φ is invariant under U_1. Indeed, for every $v \in \mathrm{Ker}\ \varphi$ and every $g \in G$, we have

$$\varphi(U_1(g)(v)) = U_2(g)(\varphi(v)) = U_2(g)(0) = 0,$$

so $U_1(g)(v) \in \mathrm{Ker}\ \varphi$. Thus, $U_1 \colon G \to \mathbf{GL}(\mathrm{Ker}\ \varphi)$ is a subrepresentation of U_1, and as U_1 is irreducible, either $\mathrm{Ker}\ \varphi = (0)$ or $\mathrm{Ker}\ \varphi = V$. In the second case, $\varphi = 0$. If $\mathrm{Ker}\ \varphi = (0)$, then φ is injective. However, $\varphi(V) \subseteq W$ is invariant under U_2, since for every $v \in V$ and every $g \in G$,

$$U_2(g)(\varphi(v)) = \varphi(U_1(g)(v)) \in \varphi(V),$$

and as $\varphi(V) \neq (0)$ (as $V \neq (0)$ since U_1 is irreducible) and U_2 is irreducible, we must have $\varphi(V) = W$; that is, φ is an isomorphism.

(ii) Since V is a complex vector space, the linear map φ has some eigenvalue $\lambda \in \mathbb{C}$. Let $E_\lambda \subseteq V$ be the eigenspace associated with λ. The subspace E_λ is invariant under U_1, since for every $u \in E_\lambda$ and every $g \in G$, we have

$$\varphi(U_1(g)(u)) = U_1(g)(\varphi(u)) = U_1(g)(\lambda u) = \lambda U_1(g)(u),$$

so $U_1 \colon G \to \mathbf{GL}(E_\lambda)$ is a subrepresentation of U_1, and as U_1 is irreducible and $E_\lambda \neq (0)$, we must have $E_\lambda = V$. ☐

An interesting corollary of Schur's Lemma is the following fact:

Proposition 12.3. *Every complex irreducible representation $U \colon G \to \mathbf{GL}(V)$ of a commutative group G is one-dimensional.*

Proof. Since G is abelian, we claim that for every $g \in G$, the map $\tau_g \colon V \to V$ given by $\tau_g(v) = U(g)(v)$ for all $v \in V$ is a G-map. This amounts to checking that the following diagram commutes

$$\begin{tikzcd} V \arrow[r, "U(g_1)"] \arrow[d, "\tau_g"'] & V \arrow[d, "\tau_g"] \\ V \arrow[r, "U(g_1)"'] & V \end{tikzcd}$$

for all $g, g_1 \in G$. This is equivalent to checking that

$$\tau_g(U(g_1)(v)) = U(g)(U(g_1)(v)) = U(gg_1)(v) = U(g_1)(\tau_g(v)) = U(g_1)(U(g)(v)) = U(g_1 g)(v)$$

for all $v \in V$, that is, $U(gg_1)(v) = U(g_1 g)(v)$, which holds since G is commutative (so $gg_1 = g_1 g$).

By Schur's Lemma (Lemma 12.2 (ii)), $\tau_g = \lambda_g \mathrm{id}$ for some $\lambda_g \in \mathbb{C}$. It follows that any subspace of V is invariant. If the representation is irreducible, we must have $\dim(V) = 1$ since otherwise V would contain a one-dimensional invariant subspace, contradicting the assumption that U is irreducible. ☐

12.2 Unitary Group Representations

We now generalize representations to allow the representing space to be a *complex Hilbert space* (typically separable).

Definition 12.5. Given a locally compact group G and a complex Hilbert space H, a *unitary representation of* G *in* H is a group homomorphism $U\colon G \to \mathbf{U}(H)$, where $\mathbf{U}(H)$ is the group of unitary operators on H, such that:

(C) The map $g \mapsto U(g)(u)$ is continuous for every $u \in H$.

(U) Every linear map $U(g)$ is an isometry; that is,

$$\langle U(g)(u), U(g)(v)\rangle = \langle u, v\rangle, \qquad \text{for all } g \in G \text{ and all } u, v \in H.$$

In particular $U(g)$ is continuous and

$$(U(g))^{-1} = (U(g))^* \qquad \text{for all } g \in G.$$

As in Definition 12.1, to avoid confusion when representations involving diffferent groups arise we denote the space of the representation U by H_U, and so we denote a representation as $U\colon G \to \mathbf{U}(H_U)$.

Remark: Sometimes, a unitary representation as in Definition 12.5 is called a *continuous* unitary representation. Note that if H is infinite-dimensional, the map $g \mapsto U(g)$ is *not necessarily continuous*. For a counter-example involving the regular representation of an infinite compact group G in $\mathrm{L}^2(G)$, see Dieudonné [21] (Chapter XXI, Section 1, Problem 3). However, the left action $U^a\colon G \times H \to H$ associated with U given by

$$U^a(s, x) = U(s)(x), \qquad \text{for all } s \in G \text{ and all } x \in H$$

is continuous. Indeed, since $U(s)$ is a unitary map, we have $\|U(s)(w)\| = \|w\|$ for all $w \in H$, so for all $s, t \in G$ and all $x, y \in H$, we have

$$\begin{aligned}
\|U^a(s,x) - U^a(t,y)\| &\leq \|U(s)(x) - U(s)(y)\| + \|U(s)(y) - U(t)(y)\| \\
&= \|U(s)(x-y)\| + \|U(s)(y) - U(t)(y)\| \\
&= \|x - y\| + \|U(s)(y) - U(t)(y)\|,
\end{aligned}$$

and since by hypothesis, for any fixed $y \in H$, the map $s \mapsto U(s)(y)$ is continuous, we see that the action U^a is continuous. Conversely, if the action $U^a\colon G \times H \to H$ is continuous, then obviously the map $s \mapsto U(s)(y)$ is continuous, so U is a unitary representation.

The notion of morphism of unitary representations and of equivalence is adapted as follows.

Definition 12.6. Given any two unitary representations $U_1 \colon G \to \mathbf{U}(H_1)$ and $U_2 \colon G \to \mathbf{U}(H_2)$, a *G-map* (or *morphism of representations*) $\varphi \colon U_1 \to U_2$ is a continuous linear map which is *equivariant*, which means that the following diagram commutes for every $g \in G$:

$$\begin{tikzcd} H_1 \arrow[r, "U_1(g)"] \arrow[d, "\varphi"'] & H_1 \arrow[d, "\varphi"] \\ H_2 \arrow[r, "U_2(g)"'] & H_2, \end{tikzcd}$$

i.e.

$$\varphi \circ U_1(g) = U_2(g) \circ \varphi, \qquad g \in G.$$

The space of all G-maps between two representations as above is denoted $\mathrm{Hom}_G(U_1, U_2)$. A G-map is also called an *intertwining operator*. Two unitary representations $U_1 \colon G \to \mathbf{U}(H_1)$ and $U_2 \colon G \to \mathbf{U}(H_2)$ are *equivalent* iff $\varphi \colon H_1 \to H_2$ is an invertible linear isometry whose inverse is also continuous; thus $U_2(g) = \varphi \circ U_1(g) \circ \varphi^{-1}$, for all $g \in G$.

When $U_1 = U_2$, the space of G-maps $\mathrm{Hom}_G(U, U)$ is a unital subalgebra of $\mathcal{L}(H)$ denoted by $\mathcal{C}(U)$ and is called the *commutant* or *centralizer* of U. Observe that

$$\mathcal{C}(U) = \{\varphi \in \mathcal{L}(H) \mid \varphi \circ U(g) = U(g) \circ \varphi \quad \text{for all } g \in G\}.$$

It is easy to show that the unital subalgebra $\mathcal{C}(U)$ of $\mathcal{L}(H)$ is actually a C^*-algebra and that it is closed in $\mathcal{L}(H)$ under weak limits (see Definition 11.19(3)). By the *von Neumann density theorem*, it is also closed in $\mathcal{L}(H)$ under strong limits (Definition 11.19(2)); see Folland [33], Section 1.6. Such a C^*-algebra is a *von Neumann algebra*.

Given a unitary representation $U \colon G \to \mathbf{U}(H)$, the definition of an invariant subspace $W \subseteq H$ is the same as in Definition 12.4. If $W \subseteq H$ is invariant under U, we say that the subrepresentation $U \colon G \to \mathbf{U}(W)$ is *closed* if W is closed in H. As in the case of unitary representations of algebras, the notion of subrepresentation is only well defined for closed invariant subspaces of H. However, by Proposition 12.5, since the closure $\overline{W}$ of an invariant subspace W is closed, the notion of subrepresentation of G in $\overline{W}$ is well defined.

In the definition of an *irreducible* unitary representation $U \colon G \to \mathbf{U}(H)$ $(H \neq (0))$, we require that the only *closed* subrepresentations $U \colon G \to \mathbf{U}(W)$ of the representation $U \colon G \to \mathbf{U}(H)$ correspond to $W = (0)$ or $W = H$.

If $U \colon G \to \mathbf{U}(H)$ is an irreducible unitary representation, then the subspace spanned by the set $\{U(s)(u) \mid s \in G,\, u \in H\}$ is dense in H. As for representations of algebras, we can define topologically cyclic representations and cyclic vectors.

Definition 12.7. Let $U \colon G \to \mathbf{U}(H)$ be a unitary representation of G in H. A vector $x_0 \in H$ is called a *totalizer*, or *totalizing vector*, or *cyclic vector* for the representation U if the subspace of H spanned by the set $\{U(s)(x_0) \mid s \in G\}$ is dense in H. Equivalently if $\mathcal{M}_{x_0}$ denotes the closure of the set $\{U(s)(x_0) \mid s \in G\}$, called the *cyclic subspace* generated by x_0, which is invariant under U, then x_0 is a totalizer (a cyclic vector) if $\mathcal{M}_{x_0} = H$. A representation which admits a totalizer is said to be *topologically cyclic*.

The importance of totalizers stems from the following result which is the analog of Proposition 11.9 for group representations. In fact, the proof is essentially the same.

Proposition 12.4. *Let $U\colon G \to \mathbf{U}(H)$ be a unitary representation of G in H. Then H is the Hilbert sum of a sequence $(H_\alpha)_{\alpha\in\Lambda}$ of closed subspaces $H_\alpha \neq (0)$ of H invariant under U, and such that the restriction of U to each H_α is topologically cyclic. If H is separable, the family Λ is countable (possibly finite).*

Proposition 12.4 is proven in the separable case in Dieudonné [24] (Chapter XV, Section 5), and in general, using Zorn's lemma; see Folland [33] (Chapter 3, Proposition 3.3).

Hilbert sums of unitary representations of a locally compact group are defined just as in the case of an algebra; see Definition 11.6. We also have the following version of Proposition 11.4 for group representations.

Proposition 12.5. *Let $U\colon G \to \mathbf{U}(H)$ be a unitary representation of G in H.*

(1) If the subspace E of H is invariant under U, then its closure $\overline{E}$ is also invariant under U.

(2) Let E be a closed subspace of H invariant under U. If $E^\perp$ is the orthogonal complement of E in H, then $E^\perp$ is invariant under U. If $U_1(s)$ and $U_2(s)$ are the restrictions of $U(s)$ to E and $E^\perp$, then $H = E \oplus E^\perp$ (the algebraic direct sum), and the representation U is the Hilbert sum of the representations U_1 and U_2.

Proof. Part (1) is easy to prove and follows from the continuity of $U(s)$; see Dieudonné [25], (Chapter III, Section 11). For Part (2), let $x \in E$ and $y \in E^\perp$. For any $s \in G$ we have

$$\langle x, U(s)(y)\rangle = \langle (U(s))^*(x), y\rangle = \langle (U(s))^{-1}(x), y\rangle = \langle U(s^{-1})(x), y\rangle = 0,$$

since E is invariant under U, so $U(s^{-1})(x) \in E$, and since $E^\perp$ is the orthogonal complement of E and $y \in E^\perp$. Then $U(s)(y)$ is orthogonal to all $x \in E$, which means that $U(s)(y) \in E^\perp$, so $E^\perp$ is invariant under U. The last property is obvious. $\square$

One should realize that Property (2) of Proposition 12.5 fails for nonunitary representations. For example, the map

$$U\colon x \mapsto \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

is a representation of $\mathbb{R}$ in $\mathbb{C}^2$, but the only nontrivial invariant subspace is the subspace spanned by $(1, 0)$, which is one-dimensional. The problem is that because $\mathbb{R}$ is not compact, there is no way to define an inner product on $\mathbb{C}^2$ invariant under U.

However, using the Haar measure, Theorem 8.36 shows that if H is a finite-dimensional hermitian space, then there is an inner-product on H for which the linear maps $U(s)$ are unitary.

Theorem 12.6. *(Complete Reducibility) Let $U : G \to \mathbf{U}(H)$ be a linear representation of a compact group G in a finite-dimensional hermitian space H of dimension $n \geq 1$. There is a hermitian inner product $\langle -, - \rangle$ on H such that $U : G \to \mathbf{U}(H)$ is a unitary representation of G in the hermitian space $(H, \langle -, - \rangle)$. The representation U is the direct sum of a finite number of irreducible unitary representations.*

Proof. As we noted in the discussion following Definition 12.1 the representation $U : G \to \mathbf{U}(H)$ is a continuous linear map $g \mapsto U(g)$ from G to $\mathcal{L}(H)$ equipped with any norm. Since G is compact and H is finite-dimensional, Theorem 8.36 yields an inner product on H which is invariant under U.

We proceed by complete induction on the dimension $n \geq 1$ of H. When $n = 1$, the representation is automatically irreducible. If $n > 1$ and the representation is not irrreducible, then it has some invariant subspace H_1 of dimension n_1 with $1 \leq n_1 < n$. By Proposition 12.5, the orthogonal complement $H_2 = H_1^{\perp}$ of H_1 is also invariant under U, and its dimension n_2 satisfies $n_2 \geq 1$ and $n_1 + n_2 = n$, with $n > 1$ and $1 \leq n_1 < n$, so we also have $1 \leq n_2 < n$. We can apply the induction hypothesis to the subrepresentations $U : G \to \mathbf{U}(H_1)$ and $U : G \to \mathbf{U}(H_2)$, with $H = H_1 \oplus H_2$, and we obtain a collection of irreducible representations of G whose direct sum is U. $\square$

Theorem 12.6 is very significant because it shows that the study of *arbitrary finite-dimensional* representations of a compact group G reduces to the study of the *irreducible unitary (finite-dimensional) representations* of G.

Example 12.6. The regular representation $\rho_{\mathbf{R}} : \mathfrak{S}_3 \to \mathbf{GL}(\mathbb{C}^6)$ of $\mathfrak{S}_3$ from Example 12.2 is obviously unitary. Theorem 12.6 tells us that $\rho_{\mathbf{R}}$ is the direct sum of irreducible representations, and in Example 12.5 we already found two irreducible representations which are one-dimensional. The discussion before Example 12.5 also shows that the standard representation (see Example 12.4) must occur in the representation $\rho_{\mathbf{R}}$, and if there are h irreducible representations, the equation $n_1^2 + \cdots + n_h^2 = g = 6$ implies that $1 + 1 + 2^2 + \cdots + n_h^2 = 6$, so $h = 3$ and the standard representation occurs twice. Therefore the orthogonal complement V_3 of the direct sum $V_1 \oplus V_2$ given by the equations

$$\begin{aligned} x_1 + x_4 + x_5 &= 0 \\ x_2 + x_3 + x_6 &= 0 \end{aligned}$$

must be the direct sum of 2 two-dimensional invariant subspaces. With a little help from `Matlab` we find that the subspace V_1^3 spanned by the vectors

$$e_1 + e_2 - e_3 - e_4, \;\; e_3 + e_4 - e_5 - e_6$$

is invariant under $\rho_{\mathbf{R}}$, the subspace V_2^3 spanned by the vectors

$$e_1 - e_3 - e_4 + e_6, \;\; e_2 + e_4 - e_5 - e_6,$$

is also invariant under $\rho_{\mathbf{R}}$, both V_1^3 and V_2^3 are orthogonal to $V_1 \oplus V_2$, and

$$\mathbb{C}^6 = V_1 \oplus V_2 \oplus V_1^3 \oplus V_2^3.$$

To show that V_1^3 is invariant we observe that V_1^3 is also spanned by

$$e_1 + e_2 - e_3 - e_4,\ \ e_3 + e_4 - e_5 - e_6,\ \ e_1 + e_2 - e_5 - e_6,$$

and the action of $\rho_{\mathbf{R}}(\sigma_i)$ is to permute these vectors, possibly flipping signs, and similarly V_2^3 is also spanned by

$$e_1 - e_3 - e_4 + e_6,\ \ e_2 + e_4 - e_5 - e_6,\ \ e_1 + e_2 - e_3 - e_5,$$

and the action of $\rho_{\mathbf{R}}(\sigma_i)$ is also to permute these vectors, possibly flipping signs. According to our previous discussion these two sub-representations of $\mathfrak{S}_3$ in V_1^3 and V_2^3 are equivalent to the standard representation given in Example 12.4. Thus we identified explicitly the three irreducible representations of $\mathfrak{S}_3$ as subrepresentations of the regular representation.

The analog of Proposition 11.6 holds for unitary group representations.

Proposition 12.7. *Let $U\colon G \to \mathbf{U}(H)$ be a unitary representation of G in H. A closed subspace E of H is invariant under U iff $P_E U(g) = U(g) P_E$ for all $g \in G$, in other words, $P_E \in \mathcal{C}(U) = \mathrm{Hom}_G(U, U)$, where $P_E\colon H \to E$ is the orthogonal projection of H onto E.*

Proof. Assume that $P_E \in \mathrm{Hom}_G(U, U)$, so that the following diagram commutes:

$$\begin{tikzcd} H \arrow[r, "U(g)"] \arrow[d, "P_E"'] & H \arrow[d, "P_E"] \\ H \arrow[r, "U(g)"'] & H. \end{tikzcd}$$

For any $x \in E$, since P_E is the orthogonal projection of H onto E, we have $P_E(x) = x$, so

$$P_E(U(g)(x)) = U(g)(P_E(x)) = U(g)(x),$$

which shows that $U(g)(x) \in E$, and thus that E is invariant under U.

Conversely, assume that E is invariant under U. Since E is closed, by a well-known result of Hilbert space theory, we have $H = E \oplus E^{\perp}$, an algebraic direct sum. For any $x \in E$, since E is invariant under U, we have $U(g)(x) \in E$ for all $g \in G$, and since P_E is a projection onto E, we have

$$U(g)(P_E(x)) = U(g)(x) = P_E(U(g)(x)) \qquad \text{for all } x \in E.$$

By Proposition 12.5, the subspace $E^{\perp}$ is also invariant under U. For any $x \in E^{\perp}$, we have $U(g)(x) \in E^{\perp}$, so $P_E(x) = P_E(U(g)(x)) = 0$, and we have

$$U(g)(P_E(x)) = 0 = P_E(U(g)(x)) \qquad \text{for all } x \in E^{\perp}.$$

Since $U = E \oplus E^{\perp}$, we have

$$U(g)(P_E(x)) = P_E(U(g)(x)) \qquad \text{for all } x \in H,$$

namely, $P_E \in \mathrm{Hom}_G(U, U)$. $\square$

Proposition 12.7 yields a method for proving that a unitary representation $U\colon G \to \mathbf{U}(H)$ is irreducible. Indeed, if U is reducible, then there is some nonzero G-map $\varphi \in \mathrm{Hom}_G(U, U)$ which is *not invertible*. Thus, *if every nonzero G-map in* $\mathrm{Hom}_G(U, U)$ *is invertible, then U must be irreducible*. This technique is illustrated in the next example.

Example 12.7. Recall the representations $U_m\colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_m^{\mathbb{C}}(2))$ from Example 12.3, where $\mathcal{P}_m^{\mathbb{C}}(2)$ denotes the vector space of complex homogeneous polynomials $P(z_1, z_2) = \sum_{k=0}^{m} c_k z_1^k z_2^{m-k}$ of degree m ($c_i \in \mathbb{C}$). The $m+1$ monomials $P_k = z_1^k z_2^{m-k}$ ($0 \leq k \leq m$) form a basis of $\mathcal{P}_m^{\mathbb{C}}(2)$. In the physics literature, it is customary to index homogeneous polynomials in terms of $\ell = m/2$, which is an integer when m is even but a half integer when m is odd. In this context, the number $\ell = m/2$ is the *spin* of a particle. In terms of $\ell = m/2$, a homogeneous polynomial is written as

$$P(z_1, z_2) = \sum_{k=-\ell}^{\ell} c_k z_1^{\ell-k} z_2^{\ell+k},$$

where it is assumed that $\ell + k = j$ where j takes the *integral* values $j = 0, 1, \ldots, 2\ell = m$, so that $\ell - k = 2\ell - (\ell + k) = 2\ell - j$ takes the values $2\ell, 2\ell - 1, \ldots, 0$. Note that $k = j - \ell = j - m/2$ with $j = 0, 1, \ldots, 2\ell = m$, so k is an integer only if m is even. The physics notation makes it easier to make the connection between the matrix expression of the representations U_m (renamed as U_ℓ) and the special functions expressed in terms of Jacobi polynomials; see Vilenkin [101] (Chapter III, Sections 2 and 3).

For every matrix $S \in \mathbf{SU}(2)$, with

$$S = \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix}, \quad a\overline{a} + b\overline{b} = 1,$$

for every homogeneous polynomial $P \in \mathcal{P}_m^{\mathbb{C}}(2)$, $U_m(S)(P(z_1, z_2))$ is defined by

$$U_m(S)(P(z_1, z_2)) = P(\overline{a}z_1 - bz_2, \overline{b}z_1 + az_2). \tag{U_m}$$

As defined, the representations U_m are not unitary, but since $\mathbf{SU}(2)$ is compact, we can apply Theorem 12.6 to find an invariant inner product on $\mathcal{P}_m^{\mathbb{C}}(2)$. This can actually be done quite explicitly; we will come back to this point later.

Proposition 12.8. *The representations* $U_m\colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_m^{\mathbb{C}}(2))$ *are irreducible.*

Proof. To prove that the representations U_m are irreducible, it suffices to prove that every nonzero equivariant map A in $\mathrm{Hom}_{\mathbf{SU}(2)}(U_m, U_m)$ is invertible. Actually, we will prove that $A = \lambda\mathrm{id}$, with $\lambda \in \mathbb{C}, \lambda \neq 0$. A nice and rather short proof is given in Bröcker and tom Dieck [16], Chapter 2, Proposition 5.1. The trick is to consider the matrices

$$r_x(\varphi) = \begin{pmatrix} e^{i\varphi} & 0 \\ 0 & e^{-i\varphi} \end{pmatrix}, \quad 0 < \varphi < \pi.$$

Plugging the matrix $r_x(\varphi)$ and $P = P_k = z_1^k z_2^{m-k}$ in Equation (U_m) yields

$$U_m(r_x(\varphi))(P_k) = (e^{-i\varphi}z_1)^k(e^{i\varphi}z_2)^{m-k} = e^{i(m-2k)\varphi}z_1^k z_2^{m-k} = e^{i(m-2k)\varphi}P_k.$$

Therefore, $(P_0, \ldots, P_m)$ is a basis (in fact, orthogonal) of eigenvectors of $U_m(r_x(\varphi))$ for the eigenvalues $(e^{im\varphi}, e^{i(m-2)\varphi}, \ldots, e^{-im\varphi})$. We can pick φ such that these eigenvalues are all distinct, for example $\varphi = 2\pi/m$. Now if $A \in \mathrm{Hom}_{\mathbf{SU}(2)}(U_m, U_m)$ is equivariant, then $U_m(r_x(\varphi))A = AU_m(r_x(\varphi))$, so for $k = 0, \ldots, m$ we have

$$U_m(r_x(\varphi))AP_k = AU_m(r_x(\varphi))P_k = Ae^{i(m-2k)\varphi}P_k = e^{i(m-2k)\varphi}AP_k.$$

The above implies that either $AP_k = 0$ or AP_k is an eigenvector of $U_m(r_x(\varphi))$ for the eigenvalue $e^{i(m-2k)\varphi}$. Since φ was chosen so that the eigenvalues $(e^{im\varphi}, \ldots, e^{i(m-2)\varphi}, \ldots, e^{-im\varphi})$ are all distinct, each eigenspace is one-dimensional, so $AP_k = c_kP_k$ for some $c_k \in \mathbb{C}$, $c_k \neq 0$. In either case,

$$AP_k = c_kP_k$$

for some $c_k \in \mathbb{C}$. We will now prove that $c_0 = c_1 = \cdots = c_m$. This shows that $A = c_0\mathrm{id}_{m+1}$, and since A is not the zero map, $c_0 \neq 0$, so A is invertible, as desired.

To prove that the c_k have the same value, we use the matrices

$$r_y(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}, \quad t \in \mathbb{R}.$$

Since A is equivariant, $AU_m(r_y(t)) = U_m(r_y(t))A$, so we need to compute $AU_m(r_y(t))P_m$ and $U_m(r_y(t))AP_m$. Since $P_m = z_1^m$ and $AP_k = c_kP_k$, using Equation (U_m) we have

$$\begin{aligned}
AU_m(r_y(t))P_m &= A(z_1\cos t + z_2\sin t)^m \\
&= A\sum_{k=1}^{m}\binom{m}{k}(\cos t)^k(\sin t)^{m-k}z_1^k z_2^{m-k} \\
&= \sum_{k=1}^{m}\binom{m}{k}(\cos t)^k(\sin t)^{m-k}AP_k \\
&= \sum_{k=1}^{m}\binom{m}{k}(\cos t)^k(\sin t)^{m-k}c_kP_k.
\end{aligned}$$

We also have

$$\begin{aligned} U_m(r_y(t))AP_m &= U_m(r_y(t))c_mP_m = c_mU_m(r_y(t))P_m = c_m(z_1\cos t + z_2 \sin t)^m \\ &= \sum_{k=1}^{m} \binom{m}{k} (\cos t)^k (\sin t)^{m-k} c_m P_k. \end{aligned}$$

Since $AU_m(r_y(t))P_m = U_m(r_y(t))AP_m$, comparing coefficients (since these equations hold for all $t \in \mathbb{R}$) we obtain

$$c_k = c_m, \quad 0 \leq k \leq m.$$

Therefore, on the basis $(P_0, \ldots, P_m)$ we have $AP_k = c_0P_k$, which means that $A = c_0\mathrm{id}_{m+1}$, as claimed. $\square$

Therefore, the representations $U_m \colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_m^{\mathbb{C}}(2))$ are irreducible unitary representations of $\mathbf{SU}(2)$. In fact, they constitute all of them up to equivalence, but this is harder to prove. A good strategy is to use properties of the characters of compact groups; see Section 13.2.

The groups $\mathbf{SU}(2)$ and $\mathbf{SO}(3)$ are intimately related by the adjoint representation that we review next. Details can be found in Gallier and Quaintance [39] (Chapter 15) and Gallier [37] (Chapter 9). The group $\mathbf{SU}(2)$ turns out to be the group of unit quaternions but all we need here is Theorem 12.9.

The group $\mathbf{SU}(2)$ is the group of 2×2 complex matrices q of the form

$$q = \begin{pmatrix} a + ib & c + id \\ -(c - id) & a - ib \end{pmatrix}, \quad a, b, c, d \in \mathbb{R},\ a^2 + b^2 + c^2 + d^2 = 1.$$

If we get rid of the condition $a^2 + b^2 + c^2 + d^2 = 1$, the set of *all* matrices X of the form

$$X = \begin{pmatrix} a + ib & c + id \\ -(c - id) & a - ib \end{pmatrix}, \quad a, b, c, d \in \mathbb{R}$$

is a real vector space which turns out to be closed under multiplication and in which every nonzero element has a multiplicative inverse. It is a the skew-field of *quaternions*, denoted $\mathbb{H}$.

The group $\mathbf{SU}(2)$ is a Lie group whose Lie algebra $\mathfrak{su}(2)$ is defined as follows.

Definition 12.8. The (real) vector space $\mathfrak{su}(2)$ of 2×2 *skew Hermitian matrices with zero trace* is given by

$$\mathfrak{su}(2) = \left\{ \begin{pmatrix} ix & y + iz \\ -y + iz & -ix \end{pmatrix} \;\middle|\; (x, y, z) \in \mathbb{R}^3 \right\}.$$

Observe that for every matrix $A \in \mathfrak{su}(2)$, we have $A^* = -A$, that is, A is skew Hermitian, and that $\mathrm{tr}(A) = 0$. Also note that $\mathfrak{su}(2) \subseteq \mathbb{H}$. The quaternions in $\mathfrak{su}(2)$ are also called *pure quaternions* (they have no "real part" a).

Definition 12.9. The *adjoint representation* of the group $\mathbf{SU}(2)$ is the group homomorphism $\mathrm{Ad}\colon \mathbf{SU}(2) \to \mathbf{GL}(\mathfrak{su}(2))$ defined such that for every $q \in \mathbf{SU}(2)$, with

$$q = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \in \mathbf{SU}(2),$$

we have

$$\mathrm{Ad}_q(A) = qAq^*, \quad A \in \mathfrak{su}(2),$$

where q^* is the inverse of q (since $\mathbf{SU}(2)$ is a unitary group) and is given by

$$q^* = \begin{pmatrix} \overline{\alpha} & -\beta \\ \overline{\beta} & \alpha \end{pmatrix}.$$

One needs to verify that the map Ad_q is an invertible linear map from $\mathfrak{su}(2)$ to itself, and that Ad is a group homomorphism, which is easy to do.

In order to associate a rotation ρ_q (in $\mathbf{SO}(3)$) to q, we need to embed $\mathbb{R}^3$ into $\mathbf{su}(2) \subseteq \mathbb{H}$ as the pure quaternions, by

$$\psi(x, y, z) = \begin{pmatrix} ix & y + iz \\ -y + iz & -ix \end{pmatrix}, \quad (x, y, z) \in \mathbb{R}^3.$$

Then q defines the map ρ_q (on $\mathbb{R}^3$) given by

$$\rho_q(x, y, z) = \psi^{-1}(q\psi(x, y, z)q^*).$$

Therefore, modulo the isomorphism ψ, the linear map ρ_q is the linear isomorphism Ad_q. Now the reason why this is interesting is summarized in the following result proven in Gallier [37] (Chapter 9).

Theorem 12.9. *Let $\rho\colon \mathbf{SU}(2) \to \mathbf{SO}(3)$ be the map given by*

$$\rho_q(x, y, z) = \psi^{-1}(q\psi(x, y, z)q^*), \quad q \in \mathbf{SU}(2),\ (x, y, z) \in \mathbb{R}^3.$$

The map $\rho\colon \mathbf{SU}(2) \to \mathbf{SO}(3)$ is a surjective homomorphism whose kernel is $\{I, -I\}$. We have $\rho_q = I_3$ iff $u = (b, c, d) = 0$ iff $|a| = 1$. If $u \neq 0$, then either $a = 0$ and ρ_q is a rotation by π around the axis of rotation determined by the vector $u = (b, c, d)$, or $0 < |a| < 1$ and ρ_q is the rotation around the axis of rotation determined by the vector $u = (b, c, d)$ and the angle of rotation $\theta \neq \pi$ with $0 < \theta < 2\pi$, is given by

$$\tan(\theta/2) = \frac{\|u\|}{a}.$$

Here we are assuming that a basis (w_1, w_2) has been chosen in the plane orthogonal to $u = (b, c, d)$ such that (w_1, w_2, u) is positively oriented, that is, $\det(w_1, w_2, u) > 0$ (where w_1, w_2, u are expressed over the canonical basis (e_1, e_2, e_3), which is chosen to define positive orientation).

Remark: Under the orientation defined above, we have

$$\cos(\theta/2) = a, \quad 0 < \theta < 2\pi.$$

Note that the condition $0 < \theta < 2\pi$ implies that θ is uniquely determined by the above equation. This is not the case if we choose π such that $-\pi < \theta < \pi$ since both θ and $-\theta$ satisfy the equation, and this shows why the condition $0 < \theta < 2\pi$ is preferable. If $0 < a < 1$, then $0 < \theta < \pi$, and if $-1 < a < 0$, then $\pi < \theta < 2\pi$. In the second case, ρ_q is also the rotation of axis $-u$ and of angle $-(2\pi - \theta) = \theta - 2\pi$ with $0 < 2\pi - \theta < \pi$, but this time the orientation of the plane orthogonal to $-u = (b, c, d)$ is the opposite orientation from before. This orientation is given by (w_2, w_1), so that $(w_2, w_1, -u)$ has positive orientation. Since the quaternions q and $-q$ define the same rotation, we may assume that $a > 0$, in which case $0 < \theta < \pi$, but we have to remember that if $a < 0$ and if we pick $-q$ instead of q, the vector defining the axis of rotation becomes $-u$, which amounts to flipping the orientation of the plane orthogonal to the axis of rotation.

Because there is a surjective homomorphism $\rho\colon \mathbf{SU}(2) \to \mathbf{SO}(3)$ whose kernel is $\{-I, I\}$, the irreducible representations of $\mathbf{SO}(3)$ can also be determined (up to equivalence).

Example 12.8. If $U\colon \mathbf{SO}(3) \to \mathbf{U}(H)$ is an irreducible unitary representation of $\mathbf{SO}(3)$, then $V = \rho \circ U$ is a unitary representation $V\colon \mathbf{SU}(2) \to \mathbf{U}(H)$ of $\mathbf{SU}(2)$ which must be irreducible, and $V(-I)$ is the identity. Conversely, an irreducible unitary representation $V\colon \mathbf{SU}(2) \to \mathbf{U}(H)$ of $\mathbf{SU}(2)$ descends to an irreducible unitary representation $U\colon \mathbf{SO}(3) \to \mathbf{U}(H)$ iff $V(-I) = \mathrm{id}$. Now by definition of U_m,

$$U_m(-I)(P_k) = (-z_1)^k(-z_2)^{m-k} = (-1)^m z_1^k z_2^{m-k} = (-1)^m P_k.$$

Therefore, $U_m(-I) = \mathrm{id}_{m+1}$ iff $(-1)^m = 1$ iff m is even. In summary we obtained the folowing result.

Proposition 12.10. *The unitary representations $W_\ell\colon \mathbf{SO}(3) \to \mathbf{GL}(\mathcal{P}^{\mathbb{C}}_{2\ell}(2))$ given by*

$$W_\ell(\rho_q) = U_{2\ell}(q) \quad q \in \mathbf{SU}(2), \quad \ell \geq 0$$

are irreducible. Observe that $\mathcal{P}^{\mathbb{C}}_{2\ell}(2)$ has odd dimension $2\ell + 1$.

We will prove later that every irreducible unitary representation of $\mathbf{SU}(2)$ is equivalent to some representation U_m, and that every irreducible unitary representation of $\mathbf{SO}(3)$ is equivalent to some representation W_ℓ; see Proposition 14.1. We will also present a more pleasant description of the irreducible unitary representation of $\mathbf{SO}(3)$ in terms of spaces of harmonic polynomials.

Remark: The representations $U_m\colon \mathbf{SL}(2, \mathbb{C}) \to \mathbf{GL}(\mathcal{P}^{\mathbb{C}}_m(2))$ are not unitary, but they are irreducible. If some nontrivial proper subspace F of $\mathcal{P}^{\mathbb{C}}_m(2)$ was invariant under U_s for all $s \in \mathbf{SL}(2, \mathbb{C})$, then F would also be invariant under U_s for all $s \in \mathbf{SU}(2)$, contradicting the

irreducibility of $U_m\colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_m^{\mathbb{C}}(2))$. The group $\mathbf{SL}(2, \mathbb{C})$ is the complexification of the group $\mathbf{SU}(2)$.

There is a generalization of Schur's lemma to (complex) unitary representations, which says that if a unitary representation $U\colon G \to \mathbf{U}(H)$ is irreducible, then every G-map in $\mathrm{Hom}_G(U, U)$ is of the form $\alpha\, \mathrm{id}_H$, for some $\alpha \in \mathbb{C}$.

The proof is much more difficult because a linear operator on an infinite-dimensional vector space may not have eigenvectors! It uses some results from the spectral theory of algebras, in particular, Theorem 11.51.

We state the Schur's lemma for unitary representations without proof and refer the interested reader to Folland [33] for a proof (Chapter 3, Proposition 3.5).

Theorem 12.11. *(Schur's lemma for unitary representations) The following properties hold.*

(1) A (complex) unitary representation $U\colon G \to \mathbf{U}(H)$ is irreducible iff every G-map in $\mathcal{C}(U) = \mathrm{Hom}_G(U, U)$ is of the form $\alpha\, \mathrm{id}_H$, for some $\alpha \in \mathbb{C}$.

(2) Let $U_1\colon G \to \mathbf{U}(H_1)$ and $U_2\colon G \to \mathbf{U}(H_2)$ be two complex unitary representations. If U_1 and U_2 are equivalent, then $\mathrm{Hom}_G(U_1, U_2)$ is one-dimensional; otherwise we have $\mathrm{Hom}_G(U_1, U_2) = (0)$.

As in the case of representations in finite dimensional vector spaces, an important corollary of Theorem 12.11 is the following result.

Proposition 12.12. *Every complex irreducible unitary representation $U\colon G \to \mathbf{U}(H)$ of a locally compact abelian group G in a Hilbert space H is one-dimensional.*

Proof. If G is abelian, then $U(s) \circ U(t) = U(t) \circ U(s)$ for all $s, t \in G$, which implies that $U(s) \in \mathcal{C}(U)$ for all $s \in G$. If U is irreducible, then by Part (1) of Schur's lemma, we have $U(s) = \alpha_s \mathrm{id}$ for some $\alpha_s \in \mathbb{C}$. It follows that every one-dimensional subspace of H is invariant, so H itself is one-dimensional. ☐

12.3 Unitary Representations of G and $\mathrm{L}^1(G)$

In this section we discuss the crucial fact that every unitary representation $U\colon G \to \mathbf{U}(H)$ of a locally compact group G defines a nondegenerate representation $U_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of the involutive Banach algebra $\mathrm{L}^1(G)$, and that conversely, for every nondegenerate representation $V\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of $\mathrm{L}^1(G)$, there is a unique unitary representation $U\colon G \to \mathbf{U}(H)$ of the group G such that $V = U_{\mathrm{ext}}$. These results hold for any Hilbert space H, but when dealing with Hilbert sums H is assumed to be separable.

Dieudonné [21] (Chapter XXI, Section 1) proves the above results under the simplifying assumption that G is metrizable, separable, and unimodular (and of course locally compact).

One of the reasons is that Dieudonné only shows the existence of the Haar measure for a metrizable, separable, locally compact group. We prove it for metrizable locally compact groups.

The bijection holds for any locally compact group, not necessarily unimodular, and is proven in Folland [33] (Chapter 3). Since the technical details are not particularly illuminating, we will give an outline of the constructions and proofs, using the simplifying assumption that G is metrizable. This includes the case of Lie groups. The involution $f \mapsto f^*$ in $\mathrm{L}^1(G)$ is given by $f^*(s) = \Delta(s^{-1})\overline{f(s^{-1})}$, but not much simplification is afforded if we assume that G is unimodular.

First we show that representations of $\mathrm{L}^1(G)$ are continuous.

Proposition 12.13. *Let G be a locally compact group and let $V\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ be a representation. We have*

$$\|V(f)\| \leq \|f\|_1 \qquad \textit{for all } f \in \mathcal{L}^1(G), \tag{$*$}$$

and thus V is continuous.

Proof. If G is discrete, this follows by Proposition 11.1. Otherwise, we can extend V to a representation of the unital subalgebra $\mathrm{L}^1(G) \oplus \mathbb{C}\delta_e$ of $\mathcal{M}^1(G)$ by setting $V(f\, d\lambda + \alpha\delta_e) = V(f) + \alpha\, \mathrm{id}_H$, and then we apply Proposition 11.1 to this representation. $\square$

Our goal is to construct a nondegenerate representation of the algebra $\mathrm{L}^1(G)$ in H from a continuous unitary representation $U\colon G \to \mathbf{U}(H)$ of G. Technically it is more advantageous to construct a nondegenerate representation of the algebra $\mathcal{M}^1(G)$ of complex regular Borel measures (see Definition 7.22) from a continuous unitary representation $U\colon G \to \mathbf{U}(H)$ of G but to motivate the construction let us stick with $\mathrm{L}^1(G)$. We need to define a map $\widetilde{U}\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ which is an algebra homomorphism. For every function $f \in \mathrm{L}^1(G)$, an obvious candidate $\widetilde{U}(f)$ for a continuous linear map from H to itself is

$$\widetilde{U}(f)(x) = \int f(s)U(s)(x)\, d\lambda(s), \qquad f \in \mathrm{L}^1(G),\ x \in H, \tag{1}$$

where λ is a left Haar measure on G. However the right-hand side is an integral over the vector-valued function $s \mapsto f(s)U(s)(x)$ from G to H (in general, an infinite-dimensional vector space) so the theory of integration that we have presented does not apply. We will see how to circumvent this difficulty using weak integrals a little later, but since this method works if G is a *finite* group, let us assume temporarily that G is a finite group.

Let G be a finite group of order $|G|$. In this case, the algebras $\mathrm{L}^1(G)$ and $\mathrm{L}^2(G)$ are the same and equal to the space $[G \to \mathbb{C}]$ of functions from G to $\mathbb{C}$. Convolution of two functions $f, h\colon G \to \mathbb{C}$ is given by

$$(f * h)(s) = \frac{1}{|G|} \sum_{s_1 s_2 = s} f(s_1)h(s_2) = \frac{1}{|G|} \sum_{t \in G} f(t)h(t^{-1}s). \tag{2}$$

Recall that for every $s \in G$, the function $\delta_s \colon G \to \mathbb{C}$ is given by

$$\delta_s(t) = \begin{cases} 1 & \text{if } t = s \\ 0 & \text{if } t \neq s. \end{cases}$$

We define an involution $f \mapsto f^*$ on $\mathrm{L}^1(G)$ by $f^*(s) = \overline{f(s^{-1})}$. Then $\mathrm{L}^1(G) = [G \to \mathbb{C}]$ is a unital involutive algebra under convolution with unit δ_e (where e is the identity element of G).

Using the discrete analog of (1) where the integral is replaced by a sum, given a unitary representation $U \colon G \to \mathbf{U}(H)$ of G where H is finite-dimensional, define $\widetilde{U}(f)(x)$ by

$$\widetilde{U}(f)(x) = \frac{1}{|G|} \sum_{s \in G} f(s) U(s)(x), \quad x \in H,\ f \in \mathrm{L}^1(G). \tag{3}$$

It is not hard to prove that $\widetilde{U} \colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ is an algebra representation; for details, see Simon [93] (Chapter II, Section 3). For instance, it is instructive to verify that

$$\widetilde{U}(f * h) = \widetilde{U}(f) \circ \widetilde{U}(h).$$

This result for finite groups is generalized to locally compact metrizable groups in Theorem 12.14.

Conversely, let $V \colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ be an algebra representation. Then we can construct a unitary group representation $U \colon G \to \mathbf{U}(H)$ from V such that $\widetilde{U} = V$.

If we define $U \colon G \to \mathbf{U}(H)$ by

$$U(s) = V(\delta_s), \quad s \in G, \tag{4}$$

we can verify that U is a unitary representation such that $\widetilde{U} = V$; for details, see Simon [93] (Chapter II, Section 3). This result for finite groups is generalized to locally compact metrizable groups in Theorem 12.15.

As an application of the first construction going from a representation of G to a representation of $\mathrm{L}^1(G)$ consider the left regular representation of G. The space $\mathrm{L}^2(G) = \mathrm{L}^1(G)$ has the hermitian inner product given by

$$\langle f, g \rangle = \frac{1}{|G|} \sum_{s \in G} f(s) \overline{g(s)}, \quad f, g \in \mathrm{L}^2(G). \tag{5}$$

For any $s \in G$, define $\mathbf{R}_s \colon \mathrm{L}^2(G) \to \mathrm{L}^2(G)$ by

$$(\mathbf{R}_s(f))(t) = f(s^{-1}t), \quad s, t \in G,\ \mathrm{L}^2(G). \tag{6}$$

It is easily verified that the map $s \mapsto \mathbf{R}_s$ is a linear representation of G in $\mathrm{L}^2(G)$, and since the inner product is left and right invariant under G, each $\mathbf{R}_s$ is unitary, so the map

$\mathbf{R}\colon G \to \mathbf{U}(\mathrm{L}^2(G))$ is a unitary representation of G called the *left regular representation of* G. The left regular representation is generalized to locally compact metrizable groups in Definition 12.11.

We leave it as an exercise to prove that if we apply (3) to define $\widetilde{\mathbf{R}}(f)$ we find that

$$(\widetilde{\mathbf{R}}(f))(g) = f * g, \quad f, g \in \mathrm{L}^2(G). \tag{7}$$

The algebra representation $\widetilde{\mathbf{R}}\colon \mathrm{L}^1(G) \to \mathcal{L}(\mathrm{L}^2(G))$ is generalized to locally compact metrizable groups in Definition 12.12. If G is a finite group, then $\mathrm{L}^1(G) = \mathrm{L}^2(G)$, but for infinite groups this is generally false.

We now return to the situation when G is a locally compact metrizable group and H is Hilbert space (for simplicity we may assume that H is separable). Recall that if λ is a left Haar measure on G, then we have an embedding of $\mathrm{L}^1(G)$ into $\mathcal{M}^1(G)$ given by $f \mapsto f\,d\lambda$.

As we stated ealier, given a unitary representation $U\colon G \to \mathbf{U}(H)$ we need to construct an algebra representation of $\mathrm{L}^1(G)$ but technically it is preferable to construct an algebra representation (an algebra homomorphism) $\widetilde{U}\colon \mathcal{M}^1(G) \to \mathcal{L}(H)$ of the measure algebra $\mathcal{M}^1(G)$.

Pick any complex regular Borel measure $\mu \in \mathcal{M}^1(G)$. We need to define $\widetilde{U}(\mu)$ as a continuous linear map from H to itself. An obvious candidate is

$$\widetilde{U}(\mu)(x) = \int U(s)(x)\, d\mu(s), \qquad x \in H$$

but the right-hand side is an integral over the vector-valued function $s \mapsto U(s)(x)$ from G to H, and μ is generally not a positive measure, so the theory of integration that we have presented does not apply. The theory does extend to complex measures (see Schwartz [86]), but we do not know whether this type of integral has the properties needed to obtain the desired results, so instead we will resort to a so-called weak integral. The idea is to use the duality between the Hilbert space H and its dual H', the space of continuous linear forms. So, for $x \in H$ fixed, we define the semilinear form $\Phi_{\mu,x}\colon H \to \mathbb{C}$ given by

$$\Phi_{\mu,x}(y) = \int \langle U(s)(x), y\rangle\, d\mu(s), \qquad y \in H;$$

this form is semilinear because $\Phi_{\mu,x}(y_1 + y_2) = \Phi_{\mu,x}(y_1) + \Phi_{\mu,x}(y_2)$, but $\Phi_{\mu,x}(\lambda y) = \overline{\lambda}\Phi_{\mu,x}(y)$. The function $s \mapsto \langle U(s)(x), y\rangle$ is continuous and bounded because $\|U(s)(x)\| = \|x\|$ since $U(s)$ is unitary, so it is μ-integrable (recall that $|\mu|(X)$ is finite). Using the Cauchy-Schwarz inequality we also have

$$|\Phi_{\mu,x}(y)| = \left|\int \langle U(s)(x), y\rangle\, d\mu(s)\right| \leq \|\mu\|\,\|x\|\,\|y\|\,,$$

so the semilinear form $\Phi_{\mu,x}$ is continuous. By the Riesz representation theorem for Hilbert spaces (Theorem D.9), there is a unique vector $\widetilde{U}(\mu)(x) \in H$ such that

$$\langle \widetilde{U}(\mu)(x), y\rangle = \Phi_{\mu,x}(y) \qquad \text{for all } y \in H.$$

If we let $y = \widetilde{U}(\mu)(x)$ in the inequality

$$|\langle \widetilde{U}(\mu)(x), y\rangle| \leq \|\mu\| \, \|x\| \, \|y\| \,,$$

we get

$$\left\| \widetilde{U}(\mu)(x) \right\| \leq \|\mu\| \, \|x\| \,,$$

and so

$$\left\| \widetilde{U}(\mu) \right\| \leq \|\mu\| \,. \tag{C}$$

This shows that $\widetilde{U}(\mu)$ is a continuous linear map (in $\mathcal{L}(H)$, we use the operator norm induced by the Hermitian norm on H).

Definition 12.10. Given any complex regular Borel measure $\mu \in \mathcal{M}^1(G)$, for every $x \in H$, let $\Phi_{\mu,x} \colon H \to \mathbb{C}$ be the continuous linear form given by

$$\Phi_{\mu,x}(y) = \int \langle U(s)(x), y\rangle \, d\mu(s), \qquad y \in H.$$

The unique vector $\widetilde{U}(\mu)(x) \in H$ such that

$$\langle \widetilde{U}(\mu)(x), y\rangle = \Phi_{\mu,x}(y) = \int \langle U(s)(x), y\rangle \, d\mu(s) \qquad \text{for all } y \in H$$

is called the *weak integral* of the function $s \mapsto U(s)(x)$ from G to H with respect to μ, and is denoted by

$$\int U(s)(x) \, d\mu(s) = \widetilde{U}(\mu)(x).$$

Observe that

$$\widetilde{U}(\delta_s) = U(s) \qquad \text{for all } s \in G, \tag{$\widetilde{U}(\delta_s)$}$$

where δ_s is the Dirac measure at s. Also, when $\mu = f\, d\lambda$ with $f \in \mathrm{L}^1(G)$, we have

$$\langle \widetilde{U}(f\, d\lambda)(x), y\rangle = \int f(s) \langle U(s)(x), y\rangle \, d\lambda(s) \qquad \text{for all } y \in H.$$

For simplicity of notation, we also write $\widetilde{U}(f)$ instead of $\widetilde{U}(f\, d\lambda)$ and we write

$$\widetilde{U}(f)(x) = \int f(s) U(s)(x) \, d\lambda(s).$$

The next step is to show that the map $\mu \mapsto \widetilde{U}(\mu)$ is a representation of the unital involutive Banach algebra $\mathcal{M}^1(G)$.

Theorem 12.14. *Let G be a metrizable locally compact group, and let $U\colon G \to \mathbf{U}(H)$ be a unitary representation of G in H. The map $\widetilde{U}\colon \mathcal{M}^1(G) \to \mathcal{L}(H)$ defined above is a representation of the unital involutive Banach algebra $\mathcal{M}^1(G)$. The restriction $U_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of $\widetilde{U}$ to the involutive Banach algebra $\mathrm{L}^1(G)$ is nondegenerate. The theorem also holds for any arbitrary locally compact group G.*

Proof. Theorem 12.14 is proven in Dieudonné [21] (Chapter XXI, Section 1, Theorem 21.1.6). We leave the verification that $\widetilde{U}(\mu)$ is linear as exercise. Let us verify that $\widetilde{U}(\mu * \nu) = \widetilde{U}(\mu) \circ \widetilde{U}(\nu)$. Recall the definition of the convolution of measures, Definition 8.21. For all $x, y \in H$ and all $s \in G$, we have

$$\begin{aligned}
\langle \widetilde{U}(\mu * \nu)(x), y\rangle &= \int \langle U(s)(x), y\rangle\, d(\mu * \nu) = \int\!\!\int \langle U(st)(x), y\rangle\, d\mu(s)\, d\nu(t) \\
&= \int\!\!\int \langle U(t)(x), U(s)^*(y)\rangle\, d\nu(t)\, d\mu(s) = \int \langle \widetilde{U}(\nu)(x), U(s)^*(y)\rangle\, d\mu(s) \\
&= \int \langle U(s)(\widetilde{U}(\nu)(x)), y\rangle\, d\mu(s) \\
&= \langle \widetilde{U}(\mu)(\widetilde{U}(\nu)(x)), y\rangle,
\end{aligned}$$

which proves that $\widetilde{U}(\mu * \nu) = \widetilde{U}(\mu) \circ \widetilde{U}(\nu)$.

Next recall that

$$\int \varphi(s) d\overline{\mu}(s) = \overline{\int \overline{\varphi(s)}\, d\mu(s)}$$

and

$$\int \varphi(s) d\check{\mu}(s) = \int \varphi(s^{-1})\, d\mu(s),$$

see Proposition 7.24 and Proposition 8.45. Then using the fact that since U is a unitary representation we have $(U(s))^* = U(s^{-1})$, we have

$$\begin{aligned}
\langle (\widetilde{U}(\mu))^*(x), y\rangle &= \langle x, \widetilde{U}(\mu)(y)\rangle = \overline{\langle \widetilde{U}(\mu)(y), x\rangle} \\
&= \overline{\int \langle U(s)(y), x\rangle\, d\mu(s)} \\
&= \int \overline{\langle U(s)(y), x\rangle}\, d\overline{\mu}(s) = \int \langle x, U(s)(y)\rangle\, d\overline{\mu}(s) \\
&= \int \langle (U(s))^*(x), y\rangle\, d\overline{\mu}(s) = \int \langle U(s^{-1})(x), y\rangle\, d\overline{\mu}(s) \\
&= \int \langle U(s)(x), y\rangle\, d\check{\overline{\mu}}(s) = \langle \widetilde{U}(\check{\overline{\mu}})(x), y\rangle,
\end{aligned}$$

which proves that $(\widetilde{U}(\mu))^* = \widetilde{U}(\check{\overline{\mu}})$.

Recall from Definition 11.8 that the algebra representation $\widetilde{U}$ is nondegenerate iff $\widetilde{U}(f\,d\lambda)(x) = 0$ for all $f \in \mathrm{L}^1(G)$ implies that $x = 0$. To prove that the restriction of $\widetilde{U}$ to $\mathrm{L}^1(G)$ is nondegenerate, since G is metrizable, we can find a neighborhood base of e (the identity element of G) consisting of a sequence (V_n) of open neighborhoods of e such that $V_{n+1} \subset V_n$ for all n.[2] Fix $s \in G$. Using Proposition A.39, for every $n \geq 1$ we can define a positive function $u_n \in \mathcal{K}_{\mathbb{R}}(G)$ of compact support contained in sV_n, such that $\int u_n\,d\lambda = 1$. Since for any fixed $x \in H$ the map $s \mapsto U(s)(x)$ is continuous (Condition (2) of Definition 12.5), for every $x \in H$ and every $\epsilon > 0$, there is some $n > 0$ such that

$$\|U(t)(x) - U(s)(x)\| \leq \epsilon, \qquad \text{for all } t \in sV_n.$$

For all $y \in H$, since $\int u_n\,d\lambda = 1$, we have

$$\langle \widetilde{U}(u_n\,d\lambda)(x) - U(s)(x), y\rangle = \int \langle U(t)(x) - U(s)(x), y\rangle u_n(t)\,d\lambda(t),$$

so if we choose $y = \widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)$, we get

$$\left\|\widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\right\|^2 = \int \langle U(t)(x) - U(s)(x), \widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\rangle u_n(t)\,d\lambda(t).$$

Using the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\left\|\widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\right\|^2 &= \int \langle U(t)(x) - U(s)(x), \widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\rangle u_n(t)\,d\lambda(t) \leq \\
&\int \left|\langle U(t)(x) - U(s)(x), \widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\rangle u_n(t)\right|\,d\lambda(t) \\
&\leq \int \|U(t)(x) - U(s)(x)\|\,\|\widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\|u_n(t)\,d\lambda(t) \\
&\leq \|\widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\|\epsilon \int u_n(t)\,d\lambda(t) = \|\widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\|\epsilon,
\end{aligned}$$

so we deduce that

$$\left\|\widetilde{U}(u_n\,d\lambda)(x) - U(s)(x)\right\| \leq \epsilon.$$

If there was some $x \neq 0$ such that $\widetilde{U}(f\,d\lambda)(x) = 0$ for all $f \in \mathrm{L}^1(G)$, then for $f = u_n$ we would have $U(s)(x) = 0$ for all $s \in G$, which is absurd for $s = e$ (since $U(e) = \mathrm{id}$). Therefore the restriction of $\widetilde{U}$ to $\mathrm{L}^1(G)$ is nondegenerate.

If G is not metrizable, we have to use a more general neighborhood base and a filter argument $\square$

For simplicity of notation, we write $U_{\mathrm{ext}}(f)$ instead of $U_{\mathrm{ext}}(f\,d\lambda)$. The following converse holds.

[2] This is where we assumption that G is metrizable is used. Otherwise, we may have to use an uncountable family.

Theorem 12.15. *Let G be a metrizable and locally compact group. For every nondegenerate representation $V\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of $\mathrm{L}^1(G)$, there is a unique unitary representation $U\colon G \to \mathbf{U}(H)$ of the group G such that $V = U_{\mathrm{ext}}$. Consequently, the map $U \mapsto U_{\mathrm{ext}}$ is a bijection between the set of unitary representations of the group G and the set of nondegenerate representations of the involutive Banach algebra $\mathrm{L}^1(G)$. Furthermore, a closed subspace E of H is invariant under the linear map $U(s)$ for every $s \in G$ if and only if it is invariant under the linear map $V(f)$ for every $f \in \mathrm{L}^1(G)$ (in fact, since $\mathcal{K}_{\mathbb{R}}(G)$ is dense in $\mathrm{L}^1(G)$, for every $f \in \mathcal{K}_{\mathbb{R}}(G)$). Consequently, the map $U \mapsto U_{\mathrm{ext}}$ is a bijection between the set of irreducible unitary representations of G and the set of nondegenerate topologically irreducible representations of $\mathrm{L}^1(G)$. If $H = \bigoplus_n H_n$ is a Hilbert sum, then there is a bijection between the Hilbert sum $U = \bigoplus_n U_n$ of the unitary representations $U_n\colon G \to \mathbf{U}(H_n)$ and the Hilbert sum $U_{\mathrm{ext}} = \bigoplus_n (U_n)_{\mathrm{ext}}$ of the unitary nondegenerate representations $(U_n)_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(H_n)$. The theorem also holds for any arbitrary locally compact group G.*

Proof. Theorem 12.15 is proven in Dieudonné [21] (Chapter XXI, Section 1, Theorem 21.1.7). Folland [33] (Chapter, Theorem 3.11) gives a different proof that applies to any locally compact group.

The proof of Theorem 12.14 shows that $U(s)(x)$ is the limit of a sequence $\widetilde{U}(u_n\, d\lambda)(x)$, with $u_n \in \mathcal{K}_{\mathbb{R}}(G)$, which shows that the map $U \mapsto U_{\mathrm{ext}}$ is injective. It also shows that if the closed subspace E is invariant under the $\widetilde{U}(f)$ with $f \in \mathcal{K}_{\mathbb{R}}(G)$, then it is invariant under the maps $U(s)$ for all $s \in G$. Conversely, by definition of $\widetilde{U}(\mu)$, it is immediate that if the closed subspace E is invariant under the maps $U(s)$ for all $s \in G$, then it is invariant under the $\widetilde{U}(f)$ with $f \in \mathcal{K}_{\mathbb{R}}(G)$,

Given a nondegenerate representation $V\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of $\mathrm{L}^1(G)$, we need to construct a unitary representation $U\colon G \to \mathbf{U}(H)$ of the group G such that $V = U_{\mathrm{ext}}$. If G is a finite group, then we can use (4) to define U by $U(s) = V(\delta_s)$. Unfortunately, if G is infinite then $\delta_s \notin \mathrm{L}^1(G)$, so we have to proceed differently.

The idea is that $U(s)(y)$ is the limit of a sequence $V(u_n)(y)$ for a sequence (u_n) of functions that tends to the Dirac delta function at s. To make this rigorous, we proceed as follows.

Consider the subspace E of H spanned by the set

$$\{V(f)(x) \mid f \in \mathcal{L}^1(G),\, x \in H\}.$$

Since V is nondegenerate, by Proposition 11.8, E is dense in H. Pick $s \in G$ and define a neighborhood base of e consisting of a sequence (V_n) of open neighborhoods of e such that $V_{n+1} \subset V_n$ for all n and a sequence (u_n) of functions $u_n \in \mathcal{K}_{\mathbb{R}}(G)$ of compact support contained in sV_n, as in the proof of Theorem 12.14, so that $\int u_n\, d\lambda = 1$. Since the Haar measure is left-invariant, $\int (\delta_{s^{-1}} * u_n)\, d\lambda = \int u_n\, d\lambda$, and since the function u_n has support contained in sV_n, the function $\delta_{s^{-1}} * u_n$ has support contained in V_n. For any open subset W containing e, since the V_n a form a neighborhood base of e, we have $V_n \subseteq W$ for n

large enough, so $G - W \subseteq G - V_n$ and as a result, since $\delta_{s^{-1}} * u_n$ has support contained in V_n, $\int_{G-W}(\delta_{s^{-1}} * u_n)\, d\lambda = 0$ as n tends to infinity. Then Proposition 8.50 shows that $\lim_{n\mapsto\infty} \|\delta_{s^{-1}} * u_n * f - f\|_1 = 0$, and since

$$\|u_n * f - \delta_s * f\|_1 = \|\delta_s * \delta_{s^{-1}} * u_n * f - \delta_s * f\|_1 \leq \|\delta_s\| \, \|\delta_{s^{-1}} * u_n * f - f\|_1 ,$$

we deduce that $\lim_{n\mapsto\infty} \|u_n * f - \delta_s * f\|_1 = 0$. Alternatively, we can prove this by going back to the proof of Proposition 8.50.

Remark: In [33] (Theorem 3.11), Folland defines functions ψ_{V_n} with compact support contained in V_n so that $\lim_{n\mapsto\infty} \|\psi_{V_n} * f - f\|_1 = 0$. The connection with our u_n is that $u_n = \delta_s * \psi_{V_n}$ and $\lim_{n\mapsto\infty} \|(\delta_s * \psi_{V_n}) * f - \delta_s * f\|_1 = 0$.

By Proposition 12.13, we have

$$\|V(f)\| \leq \|f\|_1 \qquad \text{for all } f \in \mathcal{L}^1(G). \tag{$*_1$}$$

Applying V to $u_n * f - \delta_s * f$, using the above inequality, we get

$$\lim_{n\mapsto\infty} \|V(u_n) \circ V(f) - V(\delta_s * f)\| = 0.$$

The above proves that for every linear combination $y = \sum_k V(f_k)(x_k) \in E$, with $f_k \in \mathcal{L}^1(G)$ and $x_k \in H$, the sequence $(V(u_n)(y))$ has a limit in H equal to $\sum_k V(\delta_s * f_k)(x_k)$, because

$$\begin{aligned} \left\| V(u_n)\left(\sum_k V(f_k)(x_k)\right) - \sum_k V(\delta_s * f_k)(x_k) \right\| &\leq \sum_k \|V(u_n)(V(f_k)(x_k)) - V(\delta_s * f_k)(x_k)\| \\ &\leq \sum_k \|V(u_n) \circ V(f_k) - V(\delta_s * f_k)\| \, \|x_k\| . \end{aligned}$$

and since

$$\lim_{n\mapsto\infty} \|V(u_n) \circ V(f_k) - V(\delta_s * f_k)\| = 0,$$

we also have

$$\lim_{n\mapsto\infty} \left\| V(u_n)\left(\sum_k V(f_k)(x_k)\right) - \sum_k V(\delta_s * f_k)(x_k) \right\| = 0. \tag{$*_2$}$$

Therefore, for $y = \sum_k V(f_k)(x_k)$, we define $U(s)(y)$ by

$$U(s)(y) = U(s)\left(\sum_k V(f_k)(x_k)\right) = \sum_k V(\delta_s * f_k)(x_k). \tag{$*_3$}$$

We obtain a linear map $U(s)$ from E to H such that

$$U(s) \circ V(f) = V(\delta_s * f), \qquad \text{for all } f \in \mathcal{L}^1(G), \tag{$\dagger$}$$

which shows that $U(s)$ maps E into itself. Note that $(*_2)$ says that

$$\lim_{n\mapsto\infty} \|V(u_n)(y) - U(s)(y)\| = 0, \quad \text{for all } y \in E, \tag{$*_4$}$$

which means that $V(u_n)$ converges strongly to $U(s)$ on E. By $(*_1)$, we have $\|V(u_n)\| \leq \|u_n\|_1 = 1$, so by $(*_4)$

$$\|U(s)(y)\| \leq \|y\| \qquad \text{for all } y \in E \text{ and all } s \in G,$$

and $U(s)$ extends uniquely to a continuous map on H, also denoted $U(s)$. What we just did also shows that

$$\|U(s)\| \leq 1 \qquad \text{for all } s \in G. \tag{$\dagger\dagger$}$$

It remains to prove that the map $s \mapsto U(s)$ is a unitary representation of G and that $V = U_{\text{ext}}$.

By definition, for any $y = \sum_k V(f_k)(x_k) \in E$,

$$U(s)(y) = \sum_k V(\delta_s * f_k)(x_k),$$

and the above expression is continuous in s.

For all $s, t \in G$ and all $f \in \mathcal{L}^1(G)$, using $(\dagger)$ we have

$$\begin{aligned}
U(st) \circ V(f) &= V(\delta_{st} * f) \\
&= V(\delta_s * (\delta_t * f)) \\
&= U(s) \circ V(\delta_t * f) \\
&= U(s) \circ U(t) \circ V(f),
\end{aligned}$$

which implies that $U(st)(y) = U(s)(U(t)(y))$ for all $y \in E$, and then by continuity $U(st) = U(s) \circ U(t)$ in $\mathcal{L}(H)$. By $(\dagger)$, we also have $U(e) = \mathrm{id}_H$.

By $(\dagger\dagger)$, we have $\|U(s)(x)\| \leq \|x\|$ for all $s \in G$ and all $x \in H$, so $\|U(s^{-1})(x)\| \leq \|x\|$, and then $\|x\| = \|U(s^{-1}s)(x)\| = \|U(s^{-1})(U(s)(x))\| \leq \|U(s)(x)\|$, so $\|U(s)(x)\| = \|x\|$, and since $U(s)$ is linear, by the polarization identity for a hermitian inner product, $U(s)$ is a continuous isometry. Therefore, U is a unitary representation of G in H.

To prove that $V = U_{\text{ext}}$, we use the fact that by Theorem 5.51, the dual $\mathrm{L}^1(G)'$ of $\mathrm{L}^1(G)$ is isomorphic to $\mathrm{L}^\infty(G)$. This means that for every continuous form $\Phi \in \mathrm{L}^1(G)'$, there is a unique function $h \in \mathrm{L}^\infty(G)$ such that

$$\Phi(f) = \int f(s)h(s)\, d\lambda(s) \qquad \text{for all } f \in \mathrm{L}^1(G).$$

With some abuse of notation, we write $\Phi(f) = (h, f) = \int f(s)h(s)\, d\lambda(s)$.

We use the following trick (see Dieudonné [21] (Chapter XXI, Section 1, Theorem 21.1.7). For all $f, g \in \mathcal{L}^1(G)$ and all $h \in \mathcal{L}^\infty(G)$,

$$(h, f * g) = \int f(s)(h, \delta_s * g)\, d\lambda(s). \tag{$**$}$$

Indeed, using the fact that $(\delta_s * g)(t) = g(s^{-1}t)$ and Fubini's theorem, we have

$$\begin{aligned}
(h, f * g) &= \int (f * g)(t) h(t)\, d\lambda(t) \\
&= \int \left(\int f(s) g(s^{-1}t)\, d\lambda(s) \right) h(t)\, d\lambda(t) \\
&= \int \left(\int f(s)(\delta_s * g)(t)\, d\lambda(s) \right) h(t)\, d\lambda(t) \\
&= \int f(s) \left(\int (\delta_s * g)(t) h(t)\, d\lambda(t) \right) d\lambda(s) \\
&= \int f(s)(h, \delta_s * g)\, d\lambda(s).
\end{aligned}$$

For any fixed pair $x, y \in H$, the map $f \mapsto \langle V(f)(x), y\rangle$ is a continuous linear form on $\mathcal{L}^1(G)$, so there is a unique $h \in \mathcal{L}^\infty(G)$ such that $(h, f) = \langle V(f)(x), y\rangle$, for all $f \in \mathrm{L}^1(G)$, and we get

$$\begin{aligned}
\langle V(f)(V(g)(x)), y\rangle &= \langle V(f * g)(x), y\rangle && V \text{ is an algebra homomorphism} \\
&= (h, f * g) = \int f(s)(h, \delta_s * g)\, d\lambda(s) && \text{by } (**) \\
&= \int f(s) \langle V(\delta_s * g)(s), y\rangle\, d\lambda(s) \\
&= \int \langle U(s)(V(g)(x)), y\rangle f(s)\, d\lambda(s) && \text{by } (\dagger) \\
&= \langle \widetilde{U}(f)(V(g)(x)), y\rangle && \text{by definition of } \widetilde{U}(f).
\end{aligned}$$

This proves that $\langle \widetilde{U}(f)(z), y\rangle = \langle V(f)(z), y\rangle$ for all $z \in E$ and all $y \in H$, and since E is dense in H, since U_{ext} is the restriction of $\widetilde{U}$ to $\mathrm{L}^1(G)$, we conclude that $U_{\mathrm{ext}}(f) = V(f)$.

If G is not metrizable, we have to use a more general neighborhood base and a filter argument. $\square$

Since the preceding proof involves many technical details, a summary of the proof focusing on the main points should be helpful.

First pick a neighborhood base of e (the identity element of G) consisting of a sequence (V_n) of open neighborhoods of e such that $V_{n+1} \subset V_n$ for all n. Second, for every $s \in G$, for every $n \geq 1$ define a positive function $u_n \in \mathcal{K}_{\mathbb{R}}(G)$ of support contained in sV_n, such

that $\int u_n \, d\lambda = 1$. Then we proved that $V(u_n)(V(f)(x))$ converges to $V(\delta_s * f)(x)$, for any $f \in \mathrm{L}^1(G)$ and any $x \in H$, and since by definition $U(s)(V(f)(x)) = V(\delta_s * f)(x)$, actually $V(u_n)(V(f)(x))$ converges to $U(s)(V(f)(x))$. But the set of linear combinations of terms of the form $V(f)(x)$ is dense in H, so we proved that $V(u_n)(y)$ converges to $U(s)(y)$ for all $y \in H$, which is strong convergence of $V(u_n)$ to $U(s)$.

As an application of Theorem 12.14, we obtain an injective representation of $\mathrm{L}^1(G)$ into $\mathrm{L}^2(G)$ which will be needed in the proof of the Peter–Weyl theorem. It is shown in Dieudonné [24] (Chapter XIV, Section 9, Theorem 14.9.2) that for every $s \in G$, for any $f \in \mathrm{L}^2(G)$, we have $\delta_s * f = \lambda_s(f) \in \mathrm{L}^2(G)$. By left-invariance of the (left) Haar measure, we have $\|\delta_s * f\|_2 = \|f\|_2$. Consequently, the map $f \mapsto \delta_s * f = \lambda_s(f)$, denoted $\mathbf{R}(s)$, is a unitary operator on $\mathrm{L}^2(G)$. Furthermore, if G is unimodular, by Theorem 14.10.6.3 of Dieudonné [24] (Chapter XIV, Section 10), the map $s \mapsto \mathbf{R}(s)$ is continuous and so it is a unitary representation of G in $\mathrm{L}^2(G)$. If G is not unimodular, then the continuity follows from the argument in Proposition 2.41 of Folland [33].

Definition 12.11. The representation $\mathbf{R}\colon G \to \mathbf{U}(\mathrm{L}^2(G))$ given by

$$(\mathbf{R}(s)(f))(t) = \lambda_s(f)(t) = f(s^{-1}t), \qquad f \in \mathrm{L}^2(G),\ s, t \in G,$$

is called the *left regular representation* of G in $\mathrm{L}^2(G)$.

By Theorem 12.14, we obtain a representation $\mathbf{R}_{\mathrm{ext}}$ of $\mathrm{L}^1(G)$ in $\mathrm{L}^2(G)$ (a homomorphism from $\mathrm{L}^1(G)$ to $\mathcal{L}(\mathrm{L}^2(G))$). Going back to Definition 12.10 of a weak integral,

$$\langle \widetilde{U}(f)(x), y\rangle = \int f(s)\langle U(s)(x), y\rangle \, d\lambda(s) \qquad \text{for all } y \in H,$$

it is not hard to prove that

$$\mathbf{R}_{\mathrm{ext}}(f)(g) = f * g,$$

with $f \in \mathrm{L}^1(G)$ and $g \in \mathrm{L}^2(G)$ (in the equation defining $\widetilde{U}(f) = \mathbf{R}_{\mathrm{ext}}(f)$, x is the function g and y is a function h). Using Proposition 8.50, it can be shown that $\mathbf{R}_{\mathrm{ext}}$ is injective, because if $f * g$ is zero almost everywhere for all $g \in \mathrm{L}^2(G)$, then $f = 0$ almost everywhere.

Definition 12.12. The representation $\mathbf{R}_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(\mathrm{L}^2(G))$ given by

$$(\mathbf{R}_{\mathrm{ext}}(f))(g) = f * g, \qquad f \in \mathrm{L}^1(G),\, g \in \mathrm{L}^2(G),$$

is called the *left regular representation* of $\mathrm{L}^1(G)$ in $\mathrm{L}^2(G)$.

12.4 Unitary Representations of LCA Groups

We know from Proposition 12.12 that every irreducible unitary representation $U\colon G \to \mathbf{U}(H)$ of a locally compact abelian group G is one-dimensional. But then every irreducible

representation of G is uniquely defined by a *character*, as introduced in Definition 10.1. This is because if $U\colon G \to \mathbf{U}(1)$ is a unitary representation of G, then $U(s)$ is a unitary map of $\mathbb{C}$ for every $s \in G$, which means that there is a complex number of unit length, say $\chi(s) \in \mathbb{T}$, such that

$$U(s)(z) = \chi(s)z, \quad \text{for all } z \in \mathbb{C},$$

and for all $s_1, s_2 \in G$ we have

$$\chi(s_1s_2)z = U(s_1s_2)(z) = U(s_1)(U(s_2)(z)) = \chi(s_1)\chi(s_2)z \quad \text{for all } z \in \mathbb{C},$$

which implies that

$$\chi(s_1s_2) = \chi(s_1)\chi(s_2).$$

But then $\chi\colon G \to \mathbb{T}$ is a character of G, and so every unitary representation $U\colon G \to \mathbf{U}(1)$ of G is of the form

$$U(s)(z) = \chi(s)z, \quad \text{for all } s \in G \text{ and all } z \in \mathbb{C}$$

for a unique character $\chi \in \widehat{G}$.

It is remarkable that *any* unitary representation $U\colon G \to \mathbf{U}(H)$ of a locally compact abelian group G can be expressed in terms of a projection-valued measure, as discussed in Section 11.11. Intuitively, the projection-valued measure glues the characters in the dual group $\widehat{G}$.

In order to state our theorem we need to recall the fundamental fact that for a locally compact abelian group G, the dual group $\widehat{G}$ and the space $\mathsf{X}(\mathrm{L}^1(G))$ of algebra characters of $\mathrm{L}^1(G)$ are homeomorphic; see Theorem 10.6. More precisely, the map $j\colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$ given by

$$j(\chi)(f) = \int_G \chi(a)f(a)\, d\lambda(a), \qquad \chi \in \widehat{G},\ f \in \mathrm{L}^1(G),$$

is a homeomorphism (where λ is a Haar measure on G). As a matter of notation, we denote the group characters in $\widehat{G}$ by χ and the algebra characters in $\mathsf{X}(\mathrm{L}^1(G))$ by ζ. Then our map is also expressed by $\chi \mapsto \zeta_\chi = j(\chi)$, with

$$\zeta_\chi(f) = \int_G \chi(a)f(a)\, d\lambda(a).$$

Also recall that the Gelfand map from $\mathrm{L}^1(G)$ to $\mathsf{X}(\mathrm{L}^1(G))$ is given by $\mathcal{G}_f(\zeta) = \zeta(f)$ and that

$$\zeta_\chi(f) = \mathcal{G}_f(\zeta_\chi) = \overline{\mathcal{F}}(f)(\chi),$$

where $\overline{\mathcal{F}}(f)$ is the Fourier co-transform of f.

The other fact that we need to recall is that every unitary representation $U\colon G \to \mathbf{U}(H)$ of G induces a nondegenerate representation $U_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of the involutive Banach algebra $\mathrm{L}^1(G)$; see Theorem 12.14. Before stating our next theorem we need to address a

notational issue, which is to make sense of the "integrals" $\int_{\widehat{G}} \chi(s)\, dP(\chi)$ and $\int_{\widehat{G}} \zeta_\chi(f)\, dP(\chi)$, for $s \in G$ and $f \in \mathrm{L}^1(G)$.

Recall that $\int_{\widehat{G}} \chi(s)\, dP(\chi)$ is the unique continuous linear map T in $\mathcal{L}(H)$ such that

$$\langle T(u), v\rangle = \int_{\widehat{G}} \chi(s)\, dP_{u,v}(\chi) \quad \text{for all } u, v \in H,$$

where for any Borel set E on $\widehat{G}$, the finite Radon measure $P_{u,v}$ is defined by

$$P_{u,v}(E) = \langle P(E)(u), v\rangle.$$

We are actually a bit sloppy because the integrand should be the function $\chi \mapsto \chi(s)$, evaluation at s. It would be more rigorous to introduce for every $s \in G$ the evaluation map $\mathrm{eval}_s^{\widehat{G}} \colon \widehat{G} \to \mathbb{C}$ given by

$$\mathrm{eval}_s^{\widehat{G}}(\chi) = \chi(s), \quad \chi \in \widehat{G}.$$

Then the integral $\int_{\widehat{G}} \chi(s)\, dP_{u,v}(\chi)$ is really

$$\int_{\widehat{G}} \mathrm{eval}_s^{\widehat{G}}\, dP_{u,v}.$$

Similarly $\int_{\widehat{G}} \zeta_\chi(f)\, dP(\chi)$ is the unique continuous linear map S in $\mathcal{L}(H)$ such that

$$\langle S(u), v\rangle = \int_{\widehat{G}} \zeta_\chi(f)\, dP_{u,v}(\chi) \quad \text{for all } u, v \in H.$$

This time, for every $f \in \mathsf{X}(\mathrm{L}^1(G))$, we have the evaluation map $\mathrm{eval}_f^{\mathsf{X}(\mathrm{L}^1(G))} \colon \mathsf{X}(\mathrm{L}^1(G)) \to \mathbb{C}$ given by

$$\mathrm{eval}_f^{\mathsf{X}(\mathrm{L}^1(G))}(\zeta) = \zeta(f), \quad f \in \mathsf{X}(\mathrm{L}^1(G)).$$

But note that $\mathrm{eval}_f^{\mathsf{X}(\mathrm{L}^1(G))} = \mathcal{G}_f$, where $\mathcal{G}$ is the Gelfand map from $\mathrm{L}^1(G)$ to $\mathsf{X}(\mathrm{L}^1(G))$! Then

$$\zeta_\chi(f) = \mathcal{G}_f(\zeta_\chi) = \mathcal{G}_f(j(\chi)) = (\mathcal{G}_f \circ j)(\chi),$$

so the second integral $\int_{\widehat{G}} \zeta_\chi(f)\, dP_{u,v}(\chi)$ is really

$$\int_{\widehat{G}} (\mathcal{G}_f \circ j)\, dP_{u,v}.$$

Another technical point comes up in the proof of Theorem 12.17, which is the fact that we use the notion of direct image of a measure.

Definition 12.13. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measure spaces, and let $\varphi \colon X \to Y$ be a map such that $\varphi^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$ (φ is a measurable map; see Definition 5.1). If μ is a (positive) measure on $(X, \mathcal{A})$, we define the *direct image* $\varphi_*\mu$ of μ as the measure on $(Y, \mathcal{B})$ given by

$$\varphi_*\mu(B) = \mu(\varphi^{-1}(B)), \quad B \in \mathcal{B}.$$

We leave it as an exercise to prove that $\varphi_*\mu$ is a measure. Then we have the following result.

Proposition 12.16. *With the notations of Definition 12.13, if $g \in \mathcal{L}^1_{\varphi_*\mu}(Y, \mathcal{B}, \mathbb{C})$, then $g \circ \varphi \in \mathcal{L}^1_{\mu}(X, \mathcal{A}, \mathbb{C})$ and*

$$\int_X (g \circ \varphi)\, d\mu = \int_Y g\, d(\varphi_*\mu).$$

The proof is not difficult but if you get stuck, see Folland [33] (Proposition 10.1) or Lang [62] (Chapter VI, Exercise 8). Proposition 12.16 can be extended to complex Radon measures on locally compact spaces where $\mathcal{A}$ and $\mathcal{B}$ are the Borel σ-algebras on X and Y, respectively.

Now that we have given a precise meaning to our generalized integrals we can state the following important result.

Theorem 12.17. *Let G be a locally compact abelian group with Haar measure λ. For every unitary representation $U \colon G \to \mathbf{U}(H)$ of G there is a unique regular projection-valued measure P on the dual group $\widehat{G}$ such that*

$$\begin{aligned} U(s) &= \int_{\widehat{G}} \chi(s)\, dP(\chi), \quad s \in G \\ U_{\mathrm{ext}}(f) &= \int_{\widehat{G}} \zeta_\chi(f)\, dP(\chi), \quad f \in \mathrm{L}^1(G). \end{aligned}$$

According to the preceding remarks, a more rigorous statement of the above equations is

$$\begin{aligned} U(s) &= \int_{\widehat{G}} \mathrm{eval}_s^{\widehat{G}}\, dP, \quad s \in G \\ U_{\mathrm{ext}}(f) &= \int_{\widehat{G}} (\mathcal{G}_f \circ j)\, dP, \quad f \in \mathrm{L}^1(G), \end{aligned}$$

where j is the homeomorphism $j \colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$. Moreover, a continuous linear map $T \in \mathcal{L}(H)$ belongs to $\mathcal{C}(U)$ iff T commutes with $P(E)$ for every Borel set $E \subseteq \widehat{G}$.

Proof sketch. The statement about commuting operators is proven in Folland [33]; see Theorem 1.54 and Theorem 3.12(b). The second equation in Theorem 12.17 follows from Theorem 11.58 (Spectral Theorem IV). Indeed, this theorem says that there is a unique regular projection-valued measure P^L on $\mathsf{X}(\mathrm{L}^1(G))$ such that

$$U_{\mathrm{ext}}(f) = \int_{\mathsf{X}(\mathrm{L}^1(G))} \mathcal{G}_f\, dP^L = \int_{\mathsf{X}(\mathrm{L}^1(G))} \mathcal{G}_f(\zeta)\, dP^L(\zeta), \quad f \in \mathrm{L}^1(G).$$

using the homeomorphism $j \colon \widehat{G} \to \mathsf{X}(\mathrm{L}^1(G))$ we define the projection-valued measure P on $\widehat{G}$ given by

$$P(E) = P^L(j(E))$$

for every Borel set E on $\widehat{G}$. Because j is a homeomorphism, P is the direct image of the measure P^L by j^{-1} and P^L is the direct image of the measure P by j, so by Proposition 12.16 we have

$$\int_{\widehat{G}} (\mathcal{G}_f \circ j)\, dP = \int_{\mathsf{X}(\mathrm{L}^1(G))} \mathcal{G}_f\, dP^L.$$

Thus we proved that

$$U_{\mathrm{ext}}(f) = \int_{\widehat{G}} (\mathcal{G}_f \circ j)\, dP = \int_{\widehat{G}} \zeta_\chi(f)\, dP(\chi), \quad f \in \mathrm{L}^1(G),$$

as claimed.

The first equation follows from the second equation but the proof is more involved. The argument uses the technique from the proof of Theorem 12.15. To simplify notation, write $V = U_{\mathrm{ext}}$. We need to recover U from V. The idea is that $U(s)(y)$ $(s \in G, y \in H)$ is the limit of the sequence $V(u_n)(y)$ for a sequence (u_n) of functions that tends to the Dirac delta function at s. If G is metrizable we can use the proof method of Theorem 12.15. In this case, we introduced a neighborhood base of e consisting of a sequence (V_n) of open neighborhoods of e such that $V_{n+1} \subset V_n$ for all n and a sequence of positive functions $u_n \in \mathcal{K}_{\mathbb{R}}(G)$ of support contained in sV_n, such that $\int u_n\, d\lambda = 1$. We proved that $V(u_n)(V(f)(x))$ converges to $V(\delta_s * f)(x)$, for any $f \in \mathrm{L}^1(G)$ and any $x \in H$, and since by definition $U(s)(V(f)(x)) = V(\delta_s * f)(x)$, actually $V(u_n)(V(f)(x))$ converges to $U(s)(V(f)(x))$. But the set of linear combinations of terms of the form $V(f)(x)$ is dense in H, so we proved that $V(u_n)(y)$ converges to $U(s)(y)$ for all $y \in H$, which is strong convergence of $V(u_n)$ to $U(s)$. Let us take a closer look at

$$V(u_n) = \int_{\widehat{G}} \zeta_\chi(u_n)\, dP(\chi).$$

The u_n have support in sV_n, where the V_n are neighborhoods of e, so the functions $\delta_{s^{-1}} * u_n$ have support in V_n, and $\int (\delta_{s^{-1}} * u_n)\, d\lambda = 1$. But since ζ is an algebra homomorphism with respect to convolution, and since by Proposition 10.19,

$$\zeta_\chi(\delta_s) = \overline{\mathcal{F}}(\delta_s)(\chi) = \chi(s),$$

we obtain

$$\begin{aligned} V(u_n) &= \int_{\widehat{G}} \zeta_\chi(u_n)\, dP(\chi) \\ &= \int_{\widehat{G}} \zeta_\chi(\delta_s * (\delta_{s^{-1}} * u_n))\, dP(\chi) \\ &= \int_{\widehat{G}} \zeta_\chi(\delta_s)\zeta_\chi(\delta_{s^{-1}} * u_n)\, dP(\chi) \\ &= \int_{\widehat{G}} \chi(s)\zeta_\chi(\delta_{s^{-1}} * u_n)\, dP(\chi). \end{aligned}$$

For every $\epsilon > 0$, for every compact subset K of $\widehat{G}$, consider the set

$$W_{K,\epsilon} = \{a \in G \mid |\chi(a) - 1| < \epsilon, \text{ for all } \chi \in K\}.$$

It is easily checked that $W_{K,\epsilon}$ is a neighborhood of e. For $V_n \subseteq W_{K,\epsilon}$, for all $\chi \in K$, since $\int (\delta_{s^{-1}} * u_n)\, d\lambda = 1$ we have

$$|\zeta_\chi(\delta_{s^{-1}} * u_n) - 1| = \left| \int_{W_{K,\epsilon}} (\chi(a) - 1)(\delta_{s^{-1}} * u_n)\, d\lambda \right| < \epsilon. \tag{$*_1$}$$

For every $\epsilon > 0$, since $P_{u,v}$ is a finite Radon measure, there is a compact $K \subseteq \widehat{G}$ such that $|\mu_{u,v}|(\widehat{G} - K) < \epsilon$. Since

$$V(u_n) = \int_{\widehat{G}} \chi(s) \zeta_\chi(\delta_{s^{-1}} * u_n)\, dP(\chi),$$

we also have

$$\left\langle \left(V(u_n) - \int_{\widehat{G}} \chi(s)\, dP(\chi) \right)(u), v \right\rangle = \int_{\widehat{G}} \chi(s) (\zeta_\chi(\delta_{s^{-1}} * u_n) - 1)\, dP_{u,v}.$$

The integral on the right can be written as

$$\int_K \chi(s) (\zeta_\chi(\delta_{s^{-1}} * u_n) - 1)\, dP_{u,v} + \int_{\widehat{G} - K} \chi(s) (\zeta_\chi(\delta_{s^{-1}} * u_n) - 1)\, dP_{u,v}.$$

For all n such that $V_n \subseteq W_{K,\epsilon}$, by $(*_1)$ the first integral is bounded by ϵ . Since $|\chi(a)| = 1$, we have $|\chi(a) - 1| \leq 2$, so for all $\chi \in \widehat{G} - K$, since $\int |\delta_{s^{-1}} * u_n|\, d\lambda = 1$, we have

$$|\zeta_\chi(\delta_{s^{-1}} * u_n) - 1| = \left| \int_{\widehat{G} - K} (\chi(a) - 1)(\delta_{s^{-1}} * u_n)\, d\lambda \right| \leq 2, \tag{$*_2$}$$

and since $|\mu_{u,v}|(\widehat{G} - K) < \epsilon$ and $|\chi(a)| = 1$, the second integral

$$\int_{\widehat{G} - K} \chi(s) (\zeta_\chi(\delta_{s^{-1}} * u_n) - 1)\, dP_{u,v}$$

is bounded by 2ϵ. Finally the above argument shows that

$$\langle U(s)(u), v \rangle = \lim_{n \mapsto \infty} \langle V(u_n)(u), v \rangle = \int_{\widehat{G}} \chi(s)\, dP_{u,v}(\chi),$$

as claimed.

Otherwise we need to use a more general neighborhod base and a filter (or net) argument. Technically this is achieved by Theorem 3.11 of Folland [33], which relies on Proposition 2.42. Then another limit argument very similar to the one we gave above shows that the equation

$$U(s) = \int_{\widehat{G}} \chi(s)\, dP(\chi), \quad s \in G$$

follows from the equation

$$U_{\mathrm{ext}}(f) = \int_{\widehat{G}} \zeta_\chi(f)\, dP(\chi), \quad f \in \mathrm{L}^1(G).$$

The details of this proof are worked out in Folland [33] after Lemma 4.46. □

Theorem 12.17 plays a crucial role in Mackey's theory for constructing induced representations; see Chapter 16, Proposition 16.1.

As a corollary of Theorem 12.17, since by Corollary 10.11 the characters of $\mathbb{R}^n$ are the homomorphisms

$$x \mapsto e^{iy\cdot x}, \qquad x, y \in \mathbb{R}^n,$$

where $y \cdot x$ is the Euclidean product in $\mathbb{R}^n$, we obtain the following result due to Stone.

Theorem 12.18. *(Stone) For every unitary representation $U\colon \mathbb{R}^n \to \mathbf{U}(H)$ of $\mathbb{R}^n$, there is a unique projection measure P on $\mathbb{R}^n$ such that*

$$U(x) = \int_{\mathbb{R}^n} e^{iy\cdot x}\, dP(y), \quad x \in \mathbb{R}^n.$$

12.5 Functions of Positive Type and Unitary Representations

There is deep and fruitful connection between topologically cyclic unitary representations $U\colon G \to \mathbf{U}(H)$ and certain kinds of continuous functions $p \in \mathcal{C}(G;\mathbb{C})$ called functions of positive type.

Let $U\colon G \to \mathbf{U}(H)$ be a unitary representation of the locally compact group G in a Hilbert space H, let x_0 be any vector in H, and define the map $p = \psi_{U,x_0}$ by

$$p(s) = \psi_{U,x_0}(s) = \langle U(s)(x_0), x_0\rangle, \qquad s \in G.$$

Note that this definition is analogous to Definition 11.11 which involves representations of algebras, but here we are dealing with a group representation. By its very definition the function ψ_{U,x_0} is continuous, but it is also bounded, because $U(s)$ is unitary for every $s \in G$, so $\|U(s)(x_0)\| = \|x_0\|$, which implies by Cauchy–Schwarz that

$$|\psi_{U,x_0}(s)| = |\langle U(s)(x_0), x_0\rangle| \leq \|U(s)(x_0)\|\,\|x_0\| = \|x_0\|^2 = \psi_{U,x_0}(e),$$

for all $s \in G$. Consequently,

$$\|p\|_\infty = p(e)$$

and $p = \psi_{U,x_0} \in \mathcal{L}^\infty(G;\mathbb{C})$, so we obtain a continuous linear form $\omega\colon \mathcal{L}^1(G;\mathbb{C}) \to \mathbb{C}$ given by

$$\omega(f) = \int f(s)p(s)\, d\lambda(s) = \int f(s)\langle U(s)(x_0), x_0\rangle\, d\lambda(s), \qquad \text{for all } f \in \mathcal{L}^1(G;\mathbb{C}).$$

We recognize above the weak integral $U_{\mathrm{ext}}(f)$, so we have

$$\omega(f) = \int f(s)\langle U(s)(x_0), x_0\rangle \, d\lambda(s) = \langle U_{\mathrm{ext}}(f)(x_0), x_0\rangle.$$

The term on the right-hand side is exactly the term of Definition 11.11, so by Proposition 11.10, ω is a positive linear form, which means that $\omega(f^* * f) \geq 0$ for all $f \in \mathcal{L}^1(G; \mathbb{C})$, that is,

$$\int (f^* * f)(s)p(s)\, d\lambda(s) \geq 0 \qquad \text{for all } f \in \mathcal{L}^1(G; \mathbb{C}).$$

But $f^*(s) = \Delta(s^{-1})\overline{f(s^{-1})}$, so by changing t to t^{-1}, by Fubini, the left invariance of the Haar measure, and Proposition 8.27,

$$\begin{aligned}
\int (f^* * f)(s)p(s)\, d\lambda(s) &= \int\int \Delta(t^{-1})\overline{f(t^{-1})} f(t^{-1}s)p(s)\, d\lambda(t)\, d\lambda(s) \\
&= \int\int \overline{f(t)} f(ts)p(s)\, d\lambda(t)\, d\lambda(s) \\
&= \int\int \overline{f(t)} f(ts)p(s)\, d\lambda(s)\, d\lambda(t) \\
&= \int\int p(t^{-1}s)\overline{f(t)} f(s)\, d\lambda(s)\, d\lambda(t).
\end{aligned}$$

Observe that we also have

$$\psi_{U,x_0}(s^{-1}) = \overline{\psi_{U,x_0}(s)},$$

because

$$\psi_{U,x_0}(s^{-1}) = \langle U(s^{-1})(x_0), x_0\rangle = \langle (U(s))^*(x_0), x_0\rangle = \langle x_0, U(s)(x_0)\rangle = \overline{\psi_{U,x_0}(s)}.$$

Definition 12.14. If G is a locally compact group, then a continuous function $p \in \mathcal{C}(G; \mathbb{C})$ is of *positive type* if

$$\int (f^* * f)(s)p(s)\, d\lambda(s) \geq 0 \qquad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G),$$

or equivalently if

$$\int\int p(t^{-1}s)\overline{f(t)} f(s)\, d\lambda(s)\, d\lambda(t) \geq 0 \qquad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G).$$

The set of functions of positive type is denoted by $\mathcal{P}$.

We have $\mathcal{P} \subseteq \mathcal{C}(G; \mathbb{C}) \cap \mathcal{L}^1(G; \mathbb{C})$ and $\|p\|_\infty = p(e)$ for all $p \in \mathcal{P}$.

Remark: If p is of positive type, then $\int (f^* * f)(s)p(s)\, d\lambda(s) \geq 0$ for all $f \in \mathrm{L}^1(G)$. Indeed, $\mathcal{K}_{\mathbb{C}}(G)$ is dense in $\mathrm{L}^1(G)$, and for any sequence (f_n) with $f_n \in \mathcal{K}_{\mathbb{C}}(G)$ converging to f in

$\mathrm{L}^1(G)$, the sequence $f_n^* * f_n$ converges to $f^* * f$ in $\mathrm{L}^1(G)$, and this implies that the sequence $\int (f_n^* * f_n)(s)p(s)\, d\lambda(s) \geq 0$ converges to $\int (f^* * f)(s)p(s)\, d\lambda(s) \geq 0$.

Every constant function with a nonnegative value is of positive type. We see this using the fact that the Haar measure is a positive measure and by applying Proposition 7.24 to the integral $\int\int \overline{f(t)} f(s)\, d\lambda(s)\, d\lambda(t)$. For every $f \in \mathcal{L}^2(G;\mathbb{C})$, we have the left regular representation of G in $\mathrm{L}^2(G)$ with $(\mathbf{R}(s))(f) = \lambda_s f$ (see Definition 12.11), and we have

$$\langle (\mathbf{R}(s))(\overline{f}), \overline{f}\rangle = \int \overline{f(s^{-1}t)} f(t)\, d\lambda(t) = \int \check{\overline{f}}(t^{-1}s) f(t)\, d\lambda(t) = (f * \check{\overline{f}})(s),$$

so as a special case of a function p of the form ψ_{U,x_0}, $f * \check{\overline{f}} = \psi_{\mathbf{R},\overline{f}}$ is of positive type.

We showed that the functions of the form ψ_{U,x_0} are of positive type. Remarkably, every continuous function p of positive type determines a unitary topologically cyclic representation U with a cyclic vector x_0, such that $p = \psi_{U,x_0}$. Before stating our next theorem we need to recall that by Theorem 8.34, if G is a metrizable, separable, locally compact group, then $\mathrm{L}^1(G)$ is separable.

Theorem 12.19. *Let G be a metrizable, separable, locally compact group. For any continuous function $p \in \mathcal{C}(G;\mathbb{C})$, the following properties are equivalent:*

(a) There is a unitary representation $U\colon G \to \mathbf{U}(H)$ of G in a separable Hilbert space H and a vector $x_0 \in H$ such that $p = \psi_{U,x_0}$.

(b) The function p is of positive type, that is,

$$\int\int p(t^{-1}s)\overline{f(t)} f(s)\, d\lambda(s)\, d\lambda(t) \geq 0 \qquad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G).$$

(c) The function p is bounded by $p(e) \geq 0$, $\check{\overline{p}} = p$, and for every complex measure $\mu \in \mathcal{M}^1(G)$, we have

$$\int p(s)\, d(\check{\overline{\mu}} * \mu)(s) = \int\int p(t^{-1}s)\, d\overline{\mu}(t)\, d\mu(s) \geq 0.$$

If p satisfies the above conditions, then there exists a topologically cyclic unitary representation V_1 of G in a separable Hilbert space H_1 and a cyclic vector x_1 such that $p = \psi_{V_1,x_1}$. The topologically cyclic representation is unique up to equivalence, in the sense that if V_2 is another topologically cyclic representation in a separable Hilbert H_2 and if x_2 is a cyclic vector for V_2 such that $p = \psi_{V_2,x_2}$, then there is an isomorphism $T\colon H_1 \to H_2$ such that $T(x_1) = x_2$ and $V_2 = TV_1T^{-1}$.

Proof. We follow Dieudonné [22] (Chapter XXII, Section 1, Theorem 22.1.3). We already proved that (a) implies (b). Let us prove that (b) implies (c).

By Proposition 8.45, we have $\int \varphi(t) d\check{\mu}(t) = \int \varphi(t^{-1})\, d\mu(t)$. By the definition of the convolution of measures, we obtain

$$\int p(z)\, d(\check{\overline{\mu}} * \mu)(z) = \int\!\!\int p(ts)\, d\check{\overline{\mu}}(t)\, d\mu(s) = \int\!\!\int p(t^{-1}s)\, d\overline{\mu}(t)\, d\mu(s). \qquad (*_1)$$

For every complex measure μ, the union of all the open sets A of measure zero (that is, $\mu(A) = 0$) has measure zero, so there is a largest open set of measure zero.

Definition 12.15. The *support* $\mathrm{supp}(\mu)$ of the measure μ is the complement of the largest open set of measure zero.

The support of the measure μ has the property that for every $x \in \mathrm{supp}(\mu)$, for every neighborhood V of x, there is a continuous function f with compact support contained in V such that $\int f(s)\, d\mu(s) \neq 0$; see Dieudonné [24] (Chapter XIII, Section 19).

Let us first assume that μ has compact support. In this case, for any $f \in \mathcal{K}_{\mathbb{C}}(G)$, we have $\mu * f \in \mathcal{K}_{\mathbb{C}}(G)$ (see Dieudonné [24] (Chapter XIV, Section 14, 14.5.4 and 14.9.2), and it follows that

$$0 \leq \int\!\!\int p(t^{-1}s)\overline{(\mu * f)(t)}(\mu * f)(s)\, d\lambda(s)\ d\lambda(t),$$

and since by Definition 8.24,

$$(\mu * f)(t) = \int f(x^{-1}t)\, d\mu(x),$$

using Proposition 7.24 and Proposition 8.45, we have

$$\begin{aligned}
&\int\!\!\int p(t^{-1}s)\overline{(\mu * f)(t)}(\mu * f)(s)\, d\lambda(s)\ d\lambda(t) \\
&\quad = \int\!\!\int p(t^{-1}s)\left(\overline{\int f(x^{-1}t)\, d\mu(x)}\right)\left(\int f(y^{-1}s)\, d\mu(y)\right) d\lambda(s)\ d\lambda(t) \\
&\quad = \int\!\!\int p(t^{-1}s)\left(\int \overline{f(x^{-1}t)}\, d\overline{\mu}(x)\right)\left(\int f(y^{-1}s)\, d\mu(y)\right) d\lambda(s)\ d\lambda(t) \\
&\quad = \int\!\!\int p(t^{-1}s)\left(\int \overline{f(xt)}\, d\check{\overline{\mu}}(x)\right)\left(\int f(ys)\, d\check{\mu}(y)\right) d\lambda(s)\ d\lambda(t) \\
&\quad = \int\!\!\int \left(\int\!\!\int p(t^{-1}xy^{-1}s)\overline{f(t)}\, d\lambda(t) f(s)\ d\lambda(s)\right) d\check{\overline{\mu}}(x)\, d\check{\mu}(y) \\
&\quad = \int\!\!\int \left(\int\!\!\int \overline{f}(t) f(s) p(t^{-1}x(s^{-1}y)^{-1})\, d\lambda(t)\, d\lambda(s)\right) d\check{\overline{\mu}}(x)\, d\check{\mu}(y).
\end{aligned}$$

Define the group $G \times G$ as the Cartesian product $G \times G$ with the multiplication

$$(s_1, t_1)(s_2, t_2) = (s_1 s_2, t_1 t_2).$$

Then the convolution of the functions $(t,s) \mapsto F(t,s) = \overline{f}(t)f(s)$ and $(x,y) \mapsto \Pi(x,y) = p(xy^{-1})$ is given by

$$\begin{aligned}(x,y) \mapsto \iint F(t,s)\Pi((t^{-1},s^{-1})(x,y))\, d\lambda(t)\, d\lambda(s) &= \iint F(t,s)\Pi(t^{-1}x, s^{-1}y)\, d\lambda(t)\, d\lambda(s) \\ &= \iint \overline{f}(t)f(s)p(t^{-1}xy^{-1}s)\, d\lambda(t)\, d\lambda(s).\end{aligned}$$

This suggests using the regularization method (Proposition 8.50). Let (V_n) be a fundamental system of compact neighborhoods of e such that $V_{n+1} \subseteq V_n$ for all n, and let f_n be a continuous function $f_n \geq 0$ with compact support contained in V_n and such that $\int f_n\, d\lambda = 1$. Since f_n is real, $\overline{f_n} = f_n$. Then if we let $F_n(t,s) = \overline{f}_n(t)f_n(s) = f_n(t)f_n(s)$, we have $\iint F_n(t,s)\, d\lambda(t)\, d\lambda(s) = \iint f_n(t)f_n(s)\, d\lambda(t)\, d\lambda(s) = 1$, and by Proposition 8.50, the sequence of functions

$$(F_n * \Pi)(x,y)$$

converges uniformly to the function $(x,y) \mapsto p(xy^{-1})$ on every compact subset. By passing to the limit, using Proposition 13.19.3 of Dieudonné [24] (Chapter XIII, Section 19) which says that on a compact subset we can interchange the integral and the limit and $(*_1)$, we obtain the inequality

$$\iint p(xy^{-1})\, d\check{\overline{\mu}}(x)\, d\check{\mu}(y) = \iint p(xy)\, d\check{\overline{\mu}}(x)\, d\mu(y) = \int p(s) d(\check{\overline{\mu}} * \mu)(s) \geq 0.$$

Now we show that p is bounded by $p(e)$. For any finite subset $\{s_1, \ldots, s_n\}$ of G and for any complex numbers $\xi_1, \ldots, \xi_n$, the linear functional α on $\mathcal{K}_{\mathbb{C}}(G)$ given by

$$\alpha(f) = \sum_{j=1}^{n} \xi_j f(s_j)$$

is continuous, so by Radon–Riesz III (Theorem 7.30), there is a unique complex measure μ corresponding to α, called an *atomic measure*. For the measure μ, the inequality in (c) becomes

$$\sum_{j,k} p(s_j^{-1}s_k)\overline{\xi}_j\xi_k \geq 0.$$

This means that the sesquilinear form Φ defined by $\Phi(x,y) = \sum_{i,j=1}^{n} p(s_i^{-1}s_j)x_i\overline{y}_j$ must satisfy the property $\Phi(x,x) \geq 0$ for all $x \in \mathbb{C}^n$. Since

$$\Phi(x+y, x+y) = \Phi(x,x) + \Phi(x,y) + \Phi(y,x) + \Phi(y,y),$$

we see that $\Phi(x,y)+\Phi(y,x)$ must be real. By replacing x by ix, we see that $i\Phi(x,y)-i\Phi(y,x)$ must be real, so we must have

$$\Phi(y,x) = \overline{\Phi(x,y)}.$$

Therefore, the matrix $(p(s_j^{-1}s_k))$ is hermitian positive semidefinite. In particular, when $n = 2$ and with the set $\{e, s\}$, the matrix

$$\begin{pmatrix} p(e) & p(s) \\ p(s^{-1}) & p(e) \end{pmatrix}$$

must be hermitian positive semidefinite, which implies that $p(e) \geq 0$,

$$p(s^{-1}) = \overline{p(s)},$$

so $\overline{\check{p}} = p$, and

$$p(e)^2 - p(s)p(s^{-1}) = p(e)^2 - p(s)\overline{p(s)} \geq 0,$$

and thus

$$|p(s)| \leq p(e), \qquad \text{for all } s \in G,$$

namely, p is bounded (by $(p(e))$).

Let us now consider an arbitrary complex measure μ. By Proposition A.49, since G is locally compact and metrizable, there is a sequence (K_n) of compact subsets of G such that $K_n \subseteq K_{n+1}$ and $G = \bigcup_n K_n$. Then it can be shown that $\lim_{n\mapsto\infty} |\mu|(G - K_n) = 0$ (see Dieudonné [24], Chapter XIII, Section 8, Proposition 13.8.7). By Radon–Riesz III, the continuous linear functional $f \mapsto \int \chi_{K_n} f\, d\mu$ (with $f \in \mathcal{K}_{\mathbb{C}}(G)$) corresponds to a measure μ_n of compact support K_n. Then it can be shown that $\lim_{n\mapsto\infty} \|\mu - \mu_n\| = 0$, and also $\lim_{n\mapsto\infty} \left\|\check{\overline{\mu}} * \mu - \check{\overline{\mu}}_n * \mu_n\right\| = 0$ (see Dieudonné [24], Chapter XIV, Section 6, Proposition 14.6.2). Since μ_n has compact support, by our previous result $\int p(s) d(\check{\overline{\mu}}_n * \mu_n)(s) \geq 0$, but p is bounded and by the dominated convergence theorem, $\int p(s) d(\check{\overline{\mu}}_n * \mu_n)(s)$ tends to $\int p(s) d(\check{\overline{\mu}} * \mu)(s)$, and thus $\int p(s) d(\check{\overline{\mu}} * \mu)(s) \geq 0$.

Finally, we prove that (c) implies (a). Consider the linear form φ_p defined on the unital involutive Banach algebra $\mathcal{M}^1(G)$ given by

$$\varphi_p(\mu) = \int p(s)\, d\mu(s), \qquad \mu \in \mathcal{M}^1(G),$$

where p is a bounded continuous function satisfying (c), and with the involution of $\mathcal{M}^1(G)$ being given by $\mu^* = \check{\overline{\mu}}$. Observe that $\varphi_p(\check{\overline{\mu}} * \mu) = \int p(s)\, d(\check{\overline{\mu}} * \mu)(s) \geq 0$, by (c). Therefore, φ_p is a positive linear form on $\mathcal{M}^1(G)$, according to Definition 11.10. Recall that $\mathrm{L}^1(G) \oplus \mathbb{C}\delta_e$ is also a unital involutive Banach algebra, and the restriction of φ_p to $\mathrm{L}^1(G) \oplus \mathbb{C}\delta_e$ is also a positive linear form. By Proposition 11.38(1), the linear form φ_p is continuous.

Recall from Proposition 11.11 that φ_p induces a positive Hilbert form γ given by $\gamma(\mu, \nu) = \varphi_p(\nu^* * \mu)$. Then we are almost in the position of applying Proposition 11.37 to obtain a representation of the algebra $\mathrm{L}^1(G) \oplus \mathbb{C}\delta_e$, but it is not clear that Condition (U) is satisfied so we proceed directly.

Proposition 11.35 applies to the positive Hilbert form γ. To simplify notation, write $A = \mathrm{L}^1(G) \oplus \mathbb{C}\delta_e$. If

$$\mathfrak{n} = \{\mu \in A \mid \gamma(\mu, \mu) = \varphi_p(\mu^* * \mu) = 0\},$$

then $\mathfrak{n}$ is a left ideal in A and $A/\mathfrak{n} = H_0$ is a hermitian space with the inner product given by

$$\langle \pi(\mu), \pi(\nu) \rangle = \gamma(\mu, \nu) = \varphi_p(\nu^* * \mu),$$

where $\pi\colon A \to A/\mathfrak{n} = H_0$ is the quotient map. Observe that by Proposition 11.38, φ_p is a continuous linear form such that $\|\varphi_p\| = \varphi_p(e)$, so we have

$$\|\pi(\mu)\|^2 = \gamma(\mu, \mu) = \varphi_p(\mu^* * \mu) \leq \varphi_p(e) \|\mu^* * \mu\| \leq \varphi_p(e) \|\mu\|^2,$$

so π is continuous. Since A is separable, so is H_0. The completion H of H_0 is a separable Hilbert space. As in the proof of Proposition 11.37, the endomorphism $V(\mu)$ given by

$$V(\mu)(\pi(\nu)) = \pi(\mu * \nu)$$

extends to a continuous map $V(\mu)\colon H \to H$ which is a representation of A (left multiplication). Since A has a unit element δ_e, we see that

$$V(\mu)(\pi(\delta_e)) = \pi(\mu * \delta_e) = \pi(\mu),$$

so $x_0 = \pi(\delta_e)$ is a cyclic vector for V. Since

$$V(\mu)(x_0) = \pi(\mu),$$

we have

$$\langle V(\mu)(x_0), x_0 \rangle = \langle \pi(\mu), \pi(\delta_e) \rangle = \gamma(\mu, \delta_e) = \varphi_p(\delta_e^* * \mu) = \varphi_p(\mu).$$

We also claim that the representation V is nondegenerate. It suffices to prove that the set of elements of the form $f * g$ with $f, g \in \mathcal{L}^1(G)$ is dense in $\mathcal{L}^1(G)$ (Property (N)). But this follows immediately by regularization (Proposition 8.50).

We can now apply Theorem 12.15, to obtain a unitary representation $U\colon G \to \mathbf{U}(H)$, topologically cyclic, and such that $U_{\text{ext}} = V$. We know from (†) in the proof of Theorem 12.15 that U is given by

$$U(s) \circ V(\nu) = V(\delta_s * \nu),$$

and since $V(\mu)(x_0) = \pi(\mu)$, this means that

$$U(s)(\pi(\nu)) = \pi(\delta_s * \nu).$$

In particular, $U(s)(x_0) = \pi(\delta_s)$. Since $H_U = \overline{\{U(s)(x_0) \mid s \in G\}}$ is invariant under $U(s)$ for every $s \in G$, by Theorem 12.15, the closed subset H_U is also invariant under $V(\mu)$ for all $\mu \in A$, but $H_0 = \{V(\mu)(x_0) \mid \mu \in A\}$ and $x_0 \in H_U$, so we must have $H_U = H_0$, and $\{U(s)(x_0) \mid s \in G\}$ is dense in H. Therefore, x_0 is a cyclic vector for U, which means that the set $\{\pi(\delta_s) \mid s \in G\}$ is dense in H.

We have

$$\langle U(s)(x_0), x_0 \rangle = \langle \pi(\delta_s), \pi(\delta_e) \rangle = \gamma(\delta_s, \delta_e) = \varphi_p(\delta_e^* * \delta_s) = \varphi_p(\delta_s) = \int p(t)\, d\delta_s(t) = p(s),$$

so $p = \psi_{U,x_0}$, as desired. The uniqueness of U up to equivalence follows from Proposition 11.36. $\square$

In the next section we present the Gelfand–Raikov theorem.

12.6 The Gelfand–Raikov Theorem

We will not prove the Gelfand–Raikov theorem but we will prove several technical propositions needed for its proof that are of independent interest.

Proposition 12.20. *Let p be a function of positive type on G. For all $s, t \in G$, we have*

$$|p(s) - p(t)|^2 \leq 2p(e)(p(e) - \Re(p(s^{-1}t))).$$

Proof. By Theorem 12.19, we may asssume that there is cyclic unitary representation U and a cyclic vector x_0 such that

$$p(s) = \langle U(s)(x_0), x_0\rangle.$$

This immediately implies that $p(e) = \langle U(e)(x_0), x_0\rangle = \langle x_0, x_0\rangle = \|x_0\|^2$. By Cauchy–Schwarz and the fact that $p(s^{-1}) = \overline{p(s)}$, we have

$$\begin{aligned}
|p(s) - p(t)|^2 &= |\langle (U(s) - U(t))(x_0), x_0\rangle|^2 \\
&\leq \|x_0\|^2 \|(U(s) - U(t))(x_0)\|^2 \\
&= p(e)(\|U(s)(x_0)\|^2 + \|U(t)(x_0)\|^2 - 2\Re(\langle U(s)(x_0), U(t)(x_0)\rangle)) \\
&= p(e)(2\|x_0\|^2 - 2\Re(\langle U(t^{-1}s)(x_0), x_0\rangle)) \\
&= 2p(e)(p(e) - \Re(p(t^{-1}s))) = 2p(e)(p(e) - \Re(\overline{p(s^{-1}t)})) \\
&= 2p(e)(p(e) - \Re(p(s^{-1}t))),
\end{aligned}$$

as claimed. $\square$

Let $p = \psi_{U,x_0}$ be a function of positive type given by a cyclic unitary representation $U\colon G \to \mathbf{U}(H)$ with cyclic vector x_0. If (a_n) is a Hilbert basis of H (recall that the Hilbert space H is separable), then we can write

$$U(s)(x_0) = \sum_n p_n(s)a_n,$$

where each function p_n is continuous, and we have

$$\sum_n |p_n(s)|^2 = \|x_0\|^2, \qquad \text{for all } s \in G.$$

We deduce that

$$p(s^{-1}t) = \langle U(t)(x_0), U(s)(x_0)\rangle = \sum_n \overline{p_n(s)}p_n(t), \tag{$*_2$}$$

with $\sum_n |p_n(s)p_n(t)| \leq \|x_0\|^2$.

Proposition 12.21. *The product pq of two functions p and q of positive type on G is a function of positive type.*

Proof. Using a Hilbert basis as above, $(*_2)$, and a corollary of the dominated convergence theorem (Proposition 5.37), for every $f \in \mathcal{K}_{\mathbb{C}}(G)$, we have

$$\int\int p(s^{-1}t)q(s^{-1}t)\overline{f(s)}f(t)\,d\lambda(s)\,d\lambda(t) = \sum_n \int\int q(s^{-1}t)\overline{p_n(s)f(s)}p_n(t)f(t)\,d\lambda(s)\,d\lambda(t),$$

but since q is also of positive type, we have

$$\int\int q(s^{-1}t)\overline{p_n(s)f(s)}p_n(t)f(t)\,d\lambda(s)\,d\lambda(t) \geq 0, \qquad \text{for all } n,$$

so

$$\int\int p(s^{-1}t)q(s^{-1}t)\overline{f(s)}f(t)\,d\lambda(s)\,d\lambda(t) \geq 0,$$

that is, pq is of positive type. $\square$

Two subsets of the set $\mathcal{P}$ of continuous functions of positive type on G come up in the proof of the Gelfand–Raikov theorem and are of particular interest:

$$\begin{aligned} \mathcal{P}_1 &= \{f \in \mathcal{P} \mid f(e) = 1\} = \{f \in \mathcal{P} \mid \|f\|_\infty = 1\} \\ \mathcal{P}_0 &= \{f \in \mathcal{P} \mid 0 \leq f(e) \leq 1\} = \{f \in \mathcal{P} \mid \|f\|_\infty \leq 1\}. \end{aligned}$$

Since $\mathcal{P} \subseteq \mathcal{C}(G;\mathbb{C}) \cap \mathcal{L}^1(G;\mathbb{C})$, the space $\mathcal{P}$ can be given several topologies. The subspace $\mathcal{P}_1$ is particularly important because of its role in the proof of the Gelfand–Raikov theorem and remarkably, three natural topologies on $\mathcal{P}_1$ coincide.

Remark: The above notation is from Folland [33] (Chapter 3). Unfortunately, Dieudonné denotes $\mathcal{P}_1$ as $\mathcal{P}_0$.

The sets $\mathcal{P}_0$ and $\mathcal{P}_1$ are convex and bounded ($\mathcal{P}$ itself is a convex cone). Recall the definition of an *extreme point*. Given a nonempty convex set S, a point a of the boundary of S is *extreme* (or *extremal*) if $S - \{a\}$ is still convex. Equivalently, there does not exist two distinct points $x, y \in S$ such that $a = (1-\lambda)x + \lambda y$, with $0 < \lambda < 1$.

Let $\mathcal{E}(\mathcal{P}_0)$ (resp. $\mathcal{E}(\mathcal{P}_1)$) be the set of extreme points of $\mathcal{P}_0$ (resp. $\mathcal{P}_1$). The following results are shown in Folland [33] (Chapter 3, Theorem 3.25 and Lemma 3.26).

Theorem 12.22. *If $p \in \mathcal{P}_1$, then the cyclic unitary representation U associated with p given by Theorem 12.19 is irreducible iff $p \in \mathcal{E}(\mathcal{P}_1)$. We have $\mathcal{E}(\mathcal{P}_0) = \mathcal{E}(\mathcal{P}_1) \cup \{0\}$.*

In order to state the next results we need to define the weak *-topology on $\mathrm{L}^\infty(G)$. Recall from Theorem 5.51 that $\mathrm{L}^\infty(G)$ is isomorphic to the dual $(\mathrm{L}^1(G))'$ of $\mathrm{L}^1(G)$ under the pairing $(-,-)\colon \mathrm{L}^1(G) \times \mathrm{L}^\infty(G) \to \mathbb{C}$ given by

$$(f, g) = \int f(s)g(s)\,d\lambda(s).$$

Every function $g \in \mathrm{L}^\infty(G)$ defines the continuous linear form in $(\mathrm{L}^1(G))'$ given by $f \mapsto (f, g)$, for every $f \in \mathrm{L}^1(G)$, and every linear form in $(\mathrm{L}^1(G))'$ arises in this fashion for a unique function $g \in \mathrm{L}^\infty(G)$.

Definition 12.16. The weak *-topology on $\mathrm{L}^\infty(G)$ is the topology of pointwise convergence on $(\mathrm{L}^1(G))'$. This topology is defined directly on $\mathrm{L}^\infty(G)$ by the family $(p_f)_{f\in \mathrm{L}^1(G)}$ of semi-norms indexed by the set of functions $\mathrm{L}^1(G)$, such that for every $f \in \mathrm{L}^1(G)$,

$$p_f(g) = |(f, g)|, \qquad \text{for every } g \in \mathrm{L}^\infty(G).$$

(See Section 2.7 and Dieudonné [24] (Chapter XII, Section 15).)

It is proven in Folland [33] (Chapter 3, Theorem 3.31) that the topology induced on $\mathcal{P}_1 \subseteq \mathrm{L}^\infty(G)$ by the weak *-topology of $\mathrm{L}^\infty(G)$ coincides with the topology induced on $\mathcal{P}_1$ by the topology of compact convergence in $\mathbb{C}^G$ (see Definition 2.9). This result is one of the key facts in the proof of the Gelfand–Raikov theorem.

In Dieudonné [22] (Chapter XXII, Section 1, Theorem 2.1.11), it is shown that the topology induced on $\mathcal{P}_1 \subseteq \mathcal{C}(G;\mathbb{C})$ by the topology of Fréchet space of $\mathcal{C}(G;\mathbb{C})$ (see Section 2.7) and the topology induced on $\mathcal{P}_1$ by the weak *-topology of $\mathrm{L}^\infty(G)$ coincide.

An important theorem due to Gelfand and Raikov shows that there is vast supply of irreducible unitary representations for any locally compact group. This is far from obvious a priori. For example, $\mathbf{SL}(2,\mathbb{R})$ does not have finite-dimensional unitary representations, and it is not that easy to find irreducible unitary representations.

Theorem 12.23. *(Gelfand–Raikov) If G is a locally compact group, then the irreducible unitary representations of G separate points. This means that for any $s, t \in G$, if $s \neq t$, then there is an irreducible representation U such that $U(s) \neq U(t)$.*

Theorem 12.23 is proven in Folland [33] (Chapter 3, Theorem 3.34).

The notion of function of positive type is closely related to the notion of positive semidefinite function defined below, which came up during the proof of Theorem 12.19.

Definition 12.17. A function $p\colon G \to \mathbb{C}$ (not necessarily continuous) is *positive semidefinite* if for all $s_1, \ldots, s_n \in G$ and all $\xi_1, \ldots, \xi_n \in \mathbb{C}$, we have

$$\sum_{j,k=1}^{n} p(s_j^{-1}s_k)\xi_k\overline{\xi_j} \geq 0.$$

As we showed during the proof of Theorem 12.19, the matrix $(p(s_j^{-1}s_k))$ is hermitian positive semidefinite. We also have $p(s^{-1}) = \overline{p(s)}$ and $|p(s)| \leq p(e)$, so p is bounded, but examples of discontinuous or even nonmeasurable positive semidefinite p can be given. However, if p is continuous, then p is actually a function of positive type. The following result is shown in Folland [33] (Chapter 3, Proposition 3.35).

Proposition 12.24. *Let G be a locally compact group. For any bounded continuous function $p\colon G \to \mathbb{C}$, the following are equivalent:*

(1) *The function p is of positive type.*

(2) *The function p is positive semidefinite.*

(3) *We have $\int (f^* * f)(s)p(s)\, d\lambda(s) \geq 0$ for all $f \in \mathcal{K}_{\mathbb{C}}(G)$.*

In Section 17.8 we will need the notion of measure of positive type, a natural generalization of the notion of function of positive type.

12.7 Measures of Positive Type and Unitary Representations

As in the previous section, we assume that G is a separable, metrizable, locally compact group.

For any complex or σ-Radon measure μ, and any function $f \in \mathcal{K}_{\mathbb{C}}(G)$, we have

$$\begin{aligned} \int (f^* * f)(s)\, d\mu(s) &= \int\!\!\int \Delta(t^{-1})\overline{f(t^{-1})} f(t^{-1}s)\, d\lambda(t)\, d\mu(s) \\ &= \int\!\!\int \overline{f(t)} f(ts)\, d\lambda(t)\, d\mu(s). \end{aligned}$$

This suggests defining a measure of positive type as follows.

Definition 12.18. A complex or σ-Radon measure μ is of *positive type* if

$$\int (f^* * f)(s)\, d\mu(s) = \int\!\!\int \overline{f(t)} f(ts)\, d\lambda(t)\, d\mu(s) \geq 0, \qquad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G).$$

Observe that if $\mu = p\, d\lambda$ for some $p \in \mathrm{L}^1(G)$, then

$$\int (f^* * f)(s)\, d\mu(s) = \int\!\!\int \overline{f(t)} f(ts) p(s)\, d\lambda(t)\, d\lambda(s) = \int\!\!\int \overline{f(t)} f(s) p(t^{-1}s)\, d\lambda(t)\, d\lambda(s),$$

which is exactly the expression defining a function of positive type in Definition 12.14. But here $p \in \mathrm{L}^1(G)$ is not necessarily continuous, so Definition 12.18 yields a generalization of the notion of function of positive type.

Proposition 12.25. *For every complex measure $\nu \in \mathcal{M}^1(G)$, the measure $\mu = \check{\overline{\nu}} * \nu$ is of positive type on G.*

Proof. For every function $f \in \mathcal{K}_{\mathbb{C}}(G)$, we have

$$\begin{aligned}
\int (f^* * f)(s) d(\check{\overline{\nu}} * \nu)(s) &= \int\left(\int \overline{f(t)} f(ts)\, d\lambda(t)\right) d(\check{\overline{\nu}} * \nu)(s) \\
&= \int\int\int \overline{f(t)} f(tyz)\, d\lambda(t)\, d\check{\overline{\nu}}(y)\, d\nu(z) \\
&= \int\int\int \overline{f(t)} f(ty^{-1}z)\, d\lambda(t)\, d\overline{\nu}(y)\, d\nu(z) \\
&= \int\int\int \Delta(t^{-1})\overline{f(t^{-1})} f(t^{-1}y^{-1}z)\, d\lambda(t)\, d\overline{\nu}(y)\, d\nu(z) \\
&= \int\int\int \Delta(t^{-1}y)\overline{f(t^{-1}y)} f(t^{-1}z)\, d\overline{\nu}(y)\, d\nu(z)\, d\lambda(t) \\
&= \int \Delta(t^{-1}y) \left| \int f(t^{-1}x)\, d\nu(x)\right|^2 d\lambda(t) \geq 0,
\end{aligned}$$

since the modular function Δ is strictly positive. $\square$

As a corollary of Proposition 12.25, since $\delta_e = \check{\overline{\delta_e}} = \check{\overline{\delta_e}} * \delta_e$, we see that the Dirac measure δ_e is of positive type.

In the special case where $\nu = f\, d\lambda$, with $f \in \mathcal{L}^1(G)$, we see that $f^* * f$ is of positive type. It should be noted that if $f \in \mathcal{L}^1(G)$, the function of positive type $f^* * f \in \mathcal{L}^1(G)$ may not be bounded.

Example 12.9. For example, if $G = \mathbb{R}$ and if $f(x) = x^{-1/2}$ for $0 < x < 1$ and $f(x) = 0$ otherwise, then $f^* * f$ is not bounded.

Observe that $f^*(x) = f(x^{-1}) = \sqrt{x}$ if $x > 1$ and $f^*(x) = 0$ otherwise. See Figure 12.1.

We have

$$g(x) = (f^* * f)(x) = \int_{\mathbb{R}} f^*(t) f(x - t)\, dt,$$

and this integral is not zero if

$$t > 1, \qquad x - 1 < t < x.$$

If $x \leq 1$, then $g(x) = 0$. If $1 < x \leq 2$, then

$$g(x) = (f^* * f)(x) = \int_1^x \frac{\sqrt{t}}{\sqrt{x - t}}\, dt,$$

and if $x > 2$, then

$$g(x) = (f^* * f)(x) = \int_{x-1}^x \frac{\sqrt{t}}{\sqrt{x - t}}\, dt.$$

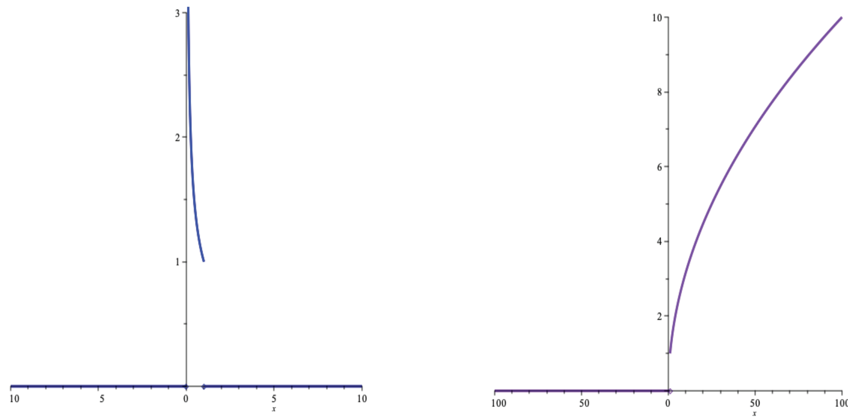


Figure 12.1: The left figure is the graph of $f(x) = x,^{-1/2}$ for $0 < x < 1$ while the right figure is the graph of $f^*(x) = f(x^{-1}) = \sqrt{x}$ if $x > 1$.

Let us compute the integral

$$\int_{x-1}^{x} \frac{\sqrt{t}}{\sqrt{x-t}}dt = \int_{x-1}^{x} \frac{1}{\sqrt{\frac{x}{t}-1}}dt.$$

If we do the change of variable

$$u = \frac{x}{t} - 1,$$

we get

$$t = \frac{x}{u+1}, \qquad dt = -\frac{xdu}{(u+1)^2},$$

so

$$\int_{x-1}^{x} \frac{dt}{\sqrt{\frac{x}{t}-1}} = \int_0^{\frac{1}{x-1}} \frac{xdu}{\sqrt{u}(u+1)^2}.$$

Next we make the change of variable

$$u = w^2,$$

so we have

$$w = \sqrt{u}, \qquad du = 2wdw,$$

and we get

$$\int_0^{\frac{1}{x-1}} \frac{xdu}{\sqrt{u}(u+1)^2} = 2x \int_0^{\frac{1}{\sqrt{x-1}}} \frac{dw}{(w^2+1)^2}.$$

But

$$\int \frac{dw}{(w^2+1)^2} = \int \frac{w^2+1-w^2}{(w^2+1)^2}dw = \int \frac{dw}{(w^2+1)} - \int \frac{w^2dw}{(w^2+1)^2} = \arctan w - \int w\frac{wdw}{(w^2+1)^2},$$

and by integrating the second term by parts, we get

$$\int \frac{dw}{(w^2+1)^2} = \frac{w}{2(w^2+1)} + \frac{1}{2}\arctan w.$$

We finally obtain

$$\begin{aligned} g(x) &= 2x\left[\frac{w}{2(w^2+1)} + \frac{1}{2}\arctan w\right]_0^{\frac{1}{\sqrt{x-1}}} \\ &= \sqrt{x-1} + x\arctan\left(\frac{1}{\sqrt{x-1}}\right), \qquad x > 2. \end{aligned}$$

When $x > 2$ goes to infinity, the second term remains positive (in fact, goes to infinity, as we can see by using the power series for $\arctan y$ with $|y| < 1$), and the first term goes to infinity.

For the sake of completeness, if $1 < x \leq 2$, we have

$$\int_1^x \frac{dt}{\sqrt{\frac{x}{t}-1}} = \int_0^{x-1} \frac{x du}{\sqrt{u}(u+1)^2},$$

and

$$\int_0^{x-1} \frac{x du}{\sqrt{u}(u+1)^2} = 2x\int_0^{\sqrt{x-1}} \frac{dw}{(w^2+1)^2}.$$

It follows that for $1 < x \leq 2$, we have

$$\begin{aligned} g(x) &= 2x\left[\frac{w}{2(w^2+1)} + \frac{1}{2}\arctan w\right]_0^{\sqrt{x-1}} \\ &= \sqrt{x-1} + x\arctan\left(\sqrt{x-1}\right). \end{aligned}$$

Therefore, the function $g = f^* * f$ is given by

$$g(x) = \begin{cases} 0 & \text{if } x \leq 1 \\ \sqrt{x-1} + x\arctan\left(\sqrt{x-1}\right) & \text{if } 1 < x \leq 2 \\ \sqrt{x-1} + x\arctan\left(\frac{1}{\sqrt{x-1}}\right) & \text{if } x > 2. \end{cases}$$

See Figure 12.2.

We showed that a continuous function p of positive type satisfies the property $p(s^{-1}) = \overline{p(s)}$, equivalently, $\check{\overline{p}} = p$. This property generalizes to measures of positive type.

Proposition 12.26. *For every measure μ of positive type on G, we have $\check{\overline{\mu}} = \mu$.*

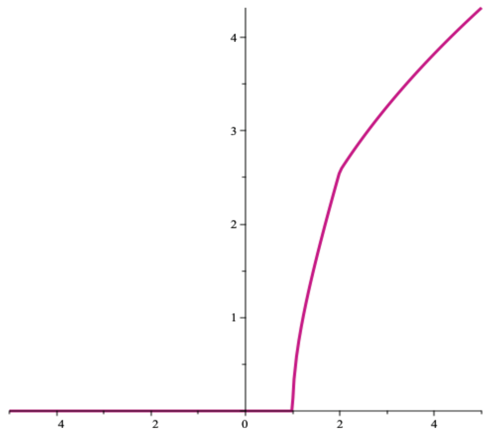


Figure 12.2: The graph of $g = f^* * f$.

Proof. Observe that if we can prove for all $f, g \in \mathcal{K}_{\mathbb{C}}(G)$ that

$$\int (f^* * g)(s)\, d\mu(s) = \int (f^* * g)(s)\, d\check{\overline{\mu}}(s), \tag{$*_3$}$$

then by regularization (Proposition 8.50), we will have

$$\int h(s)\, d\mu(s) = \int h(s)\, d\check{\overline{\mu}}(s), \qquad \text{for all } h \in \mathcal{K}_{\mathbb{C}}(G),$$

which proves that $\check{\overline{\mu}} = \mu$. By polarization, to prove $(*_3)$, it suffices to prove it for $g = f$, since

$$4g^* * f = (f+g)^* * (f+g) - (f-g)^* * (f-g) + i(f+ig)^* * (f+ig) - i(f-ig)^* * (f-ig).$$

If $f \in \mathcal{K}_{\mathbb{C}}(G)$, with $\nu = f\, d\lambda$, by Proposition 12.25 we see that $f^* * f \in \mathcal{K}_{\mathbb{C}}(G)$ is of positive type, so by Theorem 12.19(c),

$$\overline{(f^* * f)(s^{-1})} = (f^* * f)(s),$$

and since μ is of positive type we have $\int (f^* * f)(s)\, d\mu(s) \geq 0$. By $(*)$ just before Proposition 8.46,

$$\int g(s)\, d\check{\overline{\mu}}(s) = \overline{\int \check{\overline{g}}(s)\, d\mu(s)},$$

and we obtain

$$\begin{aligned}
\int (f^* * f)(s)\, d\check{\overline{\mu}}(s) &= \overline{\int \overline{(f^* * f)(s^{-1})}\, d\mu(s)} \\
&= \overline{\int (f^* * f)(s)\, d\mu(s)} \\
&= \int (f^* * f)(s)\, d\mu(s),
\end{aligned}$$

as claimed. $\square$

We conclude this section by showing that a measure μ of positive type defines a unitary representation U_μ of G. This construction will be used in Section 17.8 to define the Plancherel transform.

The vector space $\mathcal{K}_{\mathbb{C}}(G)$ is a nonunital algebra under convolution with involution $f \mapsto f^*$, with $f^*(s) = \Delta(s^{-1})\check{\overline{f}}(s)$. Because μ is of positive type, the linear form $\varphi_\mu \colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$ given by

$$\varphi_\mu(f) = \int f(s)\, d\mu(s)$$

is a positive linear form in the sense of Definition 11.10. As in Section 12.5, the set

$$\mathfrak{n} = \{f \in \mathcal{K}_{\mathbb{C}}(G) \mid \varphi_\mu(f^* * f) = 0\}$$

is a left ideal in $\mathcal{K}_{\mathbb{C}}(G)$, and $H_0 = \mathcal{K}_{\mathbb{C}}(G)/\mathfrak{n}$ is a hermitian space with the hermitian inner product

$$\langle \pi(f), \pi(g)\rangle_\mu = \varphi_\mu(g^* * f) = \int (g^* * f)(s)\, d\mu(s), \qquad (\dagger_1)$$

where $\pi \colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G)/\mathfrak{n}$ is the quotient map. Since

$$\begin{aligned}\int (g^* * f)(s)\, d\mu(s) &= \int\!\!\int \Delta(t^{-1})\overline{g(t^{-1})} f(t^{-1}s)\, d\lambda(t)\, d\mu(s)\\ &= \int\!\!\int \overline{g(t)} f(ts)\, d\lambda(t)\, d\mu(s),\end{aligned}$$

we have

$$\langle \pi(f), \pi(g)\rangle_\mu = \varphi_\mu(g^* * f) = \int\!\!\int \overline{g(t)} f(ts)\, d\lambda(t)\, d\mu(s). \qquad (\dagger_2)$$

We claim that $H_0 = \mathcal{K}_{\mathbb{C}}(G)/\mathfrak{n}$ is separable. Recall from Proposition 2.16 that since G is a locally compact separable metric space, the space $\mathcal{K}_{\mathbb{C}}(G)$ is separable. If (f_n) is a sequence of functions in $\mathcal{K}_{\mathbb{C}}(G)$ converging uniformly to a function $f \in \mathcal{K}_{\mathbb{C}}(G)$, with the supports of the f_n remaining within some fixed compact subset, then $f_n^* * f_n$ converges uniformly to $f^* * f$, the supports of the $f_n^* * f_n$ remaining with some fixed compact set, thus $\|\pi(f_n) - \pi(f)\|_\mu$ tends to zero as n tends to infinity. This shows that H_0 is separable, and we let H (or H_μ) denote the separable Hilbert space which is its completion.

Instead of first defining a nondegenerate representation V of the algebra $\mathcal{K}_{\mathbb{C}}(G)$ and then the unitary representation U of G such that $U_{\mathrm{ext}} = V$, we define $U_\mu \colon G \to \mathbf{GL}(H_0)$ as follows:

$$U_\mu(s)(\pi(f)) = \pi(\delta_s * f), \qquad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_{\mathbb{C}}(G). \qquad (*_{U_\mu})$$

Recall $(\delta_s * f)(t) = f(s^{-1}t)$, and that if $f \in \mathcal{K}_{\mathbb{C}}(G)$, then $\delta_s * f \in \mathcal{K}_{\mathbb{C}}(G)$.

Theorem 12.27. *For any measure μ of positive type, with the notation as above, if $U_\mu\colon G \to \mathbf{GL}(H_0)$ is the map defined by*

$$U_\mu(s)(\pi(f)) = \pi(\delta_s * f), \qquad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_{\mathbb{C}}(G),$$

then each linear map $U_\mu(s)$ from H_0 to itself is continuous and unitary, thus the map $U_\mu(s)$ extends to a unitary map of H, and we obtain a homomorphism $U_\mu\colon G \to \mathbf{U}(H)$. For each $x \in H$, the map $s \mapsto U_\mu(s)(x)$ is continuous, therefore, $U_\mu\colon G \to \mathbf{U}(H)$ is a unitary representation of G in H.

Proof. By $(\dagger_2)$, we have

$$\begin{aligned}
\|U_\mu(s)(\pi(f))\|_\mu^2 &= \langle U_\mu(s)(\pi(f)), U_\mu(s)(\pi(f))\rangle_\mu = \langle \pi(\delta_s * f), \pi(\delta_s * f)\rangle_\mu \\
&= \iint \overline{f(s^{-1}t)} f(s^{-1}tu)\, d\lambda(t)\, d\mu(u) \\
&= \iint \overline{f(t)} f(tu)\, d\lambda(t)\, d\mu(u) \\
&= \langle \pi(f), \pi(f)\rangle_\mu = \|\pi(f)\|_\mu^2 .
\end{aligned}$$

Thus $U_\mu(s)$ is unitary and continuous.

For any $f \in \mathcal{K}_{\mathbb{C}}(G)$, and $s \in G$, and any sequence (s_n) in G converging to $s \in G$, the sequence $(\lambda_{s_n} f)$ converges uniformly to $\lambda_s f$ (recall that $(\lambda_s f)(t) = f(s^{-1}t)$), the support of each $\lambda_s f$ remaining in a fixed compact set, so as before, the sequence $(\delta_{s_n} * f)$ converges to $\delta_s * f$, and since $U_\mu(s_n)(\pi(f)) = \pi(\delta_{s_n} * f) = \pi(\lambda_{s_n} f)$, the sequence $(U_\mu(s_n)(\pi(f)))$ converges to $U_\mu(s)(\pi(f)) \in H$. Since H_0 is dense in H, and since the set of maps $\{U_\mu(s) \mid s \in G\}$ from H to itself is equicontinuous (see Proposition 2.13 or Dieudonné [24], Chapter XII, Section 15, Theorem 12.15.7.1), as a consequence, each map $s \mapsto U_\mu(s)(x)$ is continuous (see Proposition 2.12 or Dieudonné [25], Chapter VII, Section 5, Theorem 7.5.5). ☐

Remark: According to Definition 12.10, the algebra representation $(U_\mu)_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ is defined such that for all $f \in \mathrm{L}^1(G)$ and all $g \in \mathcal{K}_{\mathbb{C}}(G)$, the vector $(U_\mu)_{\mathrm{ext}}(f)(\pi(g))$ is uniquely determined by the equation

$$\langle (U_\mu)_{\mathrm{ext}}(f)(\pi(g)), \pi(h)\rangle_\mu = \int f(s)\langle U_\mu(s)(\pi(g)), \pi(h)\rangle_\mu\, d\lambda(s) \qquad \text{for all } h \in \mathcal{K}_{\mathbb{C}}(G),$$

and since

$$U_\mu(s)(\pi(g)) = \pi(\delta_s * g)$$

and by $(\dagger_2)$,

$$\begin{aligned}
\langle U_\mu(s)(\pi(g)), \pi(h)\rangle_\mu = \langle \pi(\delta_s * g), \pi(h)\rangle_\mu &= \iint \overline{h(u)} (\delta_s * g)(ut)\, d\lambda(u)\, d\mu(t) \\
&= \iint g(s^{-1}ut)\overline{h(u)}\, d\lambda(u)\, d\mu(t),
\end{aligned}$$

we obtain

$$\begin{aligned}\langle (U_\mu)_{\mathrm{ext}}(f)(\pi(g)), \pi(h)\rangle_\mu &= \iiint f(s)g(s^{-1}ut)\overline{h(u)}\, d\lambda(u)\, d\mu(t)\, d\lambda(s)\\ &= \iiint f(s)g(s^{-1}ut)\overline{h(u)}\, d\lambda(s)\, d\lambda(u)\, d\mu(t),\end{aligned}$$

as in Dieudonné [22] (Chapter XXII, Section 7, no. 22.7.2.1), except that in Dieudonné [22], u^{-1} occurs instead of u. But Dieudonné assumes that G is unimodular, so this does not make any difference. To deal wih the case where G is not unimodular, we need to replace $\check{\overline{f}}$ by $f^* = \Delta^{-1}\check{\overline{f}}$, as we did, following Folland [33].

If $f \in \mathcal{K}_{\mathbb{C}}(G)$, then for all $g \in \mathcal{K}_{\mathbb{C}}(G)$ we have the simpler expression

$$(U_\mu)_{\mathrm{ext}}(f)(\pi(g)) = \pi(f * g),$$

as in Proposition 11.37.

Representation theory is a vast area of mathematics and we will only give a few references. A classic on the general theory is Kirillov [53]. Kirillov's survey [55] gives an excellent panorama of the field. Another encyclopedic source that covers a lot of the general theory is Hewitt and Ross [48]. Another good source for the general theory is Folland [33]. Bröcker and tom Dieck [16], Dieudonné [21], and Knapp [57] cover the representation theory of compact groups in great depth. Fulton and Harris [36], Humphreys [51], Knapp [56], Taylor [96], Varadarajan [98, 99], and Vilenkin [101] cover the representation theory of Lie groups.

We are now ready to prove the famous Peter–Weyl theorem.

Chapter 13

Analysis on Compact Groups and Representations

Chapter 10 is devoted to harmonic analysis on *abelian* locally compact (not necessarily compact) groups. In this chapter we consider the case of a compact *not necessarily abelian* group G. Noncommutativity causes trouble. In particular, the characters no longer form a group. Irreducible representations of the group G become a substitute for the group characters of abelian groups. Fortunately, compactness also has a positive influence.

The structure of the algebra $\mathrm{L}^2(G)$ is described by a Hilbert sum of *finite-dimensional* matrix algebras which are representation spaces of irreducible unitary representations of G (in fact, up to equivalence, all of them). These results constitute a deep and beautiful theorem due to Peter and Weyl, and most of this chapter is devoted to its proof.

If the (metrizable) group G is compact, then some remarkable things happen:

(1) The involutive algebra (under convolution) $\mathrm{L}^2(G)$ is a complete Hilbert algebra, and as a consequence of Theorem 11.31, the algebra $\mathrm{L}^2(G)$ is a finite or countably infinite Hilbert sum $\bigoplus_{\rho\in R} \mathfrak{a}_\rho$ of topologically simple algebras, but because G is compact, each $\mathfrak{a}_\rho$ is isomorphic to a *finite-dimensional matrix algebra* $\mathrm{M}_{n_\rho}(\mathbb{C})$. The elements of $\mathfrak{a}_\rho$ are continuous functions on G. This is the first half of the first part of a theorem due to Peter and Weyl (1927); see Theorem 13.2.

Since each minimal two-sided ideal $\mathfrak{a}_\rho$ is finite-dimensional, it can be expressed as a finite direct sum
$$\mathfrak{a}_\rho = \bigoplus_{1\le j\le n_\rho} \mathfrak{a}_\rho * m_j,$$
of orthogonal minimal left ideals, where the m_j are self-adjoint irreducible idempotents. We can pick a Hilbert basis $(a_j)_{1\le j\le n_\rho}$ in $\mathfrak{l}_1 = \mathfrak{a}_\rho * m_1$, such that $a_j \in m_j * \mathfrak{a}_\rho * m_1$, and then it turns out that there is some $\gamma > 0$ such that
$$a_j * \check{\overline{a}}_j = \gamma m_j, \qquad \check{\overline{a}}_j * a_j = \gamma m_1, \qquad 1 \le j \le n_\rho.$$

In fact, we will show that $\gamma = n_\rho^{-1}$. Finally, for all j, k with $1 \leq j, k \leq n_\rho$, let

$$m_{jk} = \gamma^{-1} a_j * \check{\overline{a}}_k,$$

which we also denote by $m_{jk}^{(\rho)}$. We have $m_{jj} = m_j$.

Then the family

$$\left(\frac{1}{\sqrt{n_\rho}} m_{ij}^{(\rho)}\right)_{1 \leq i,j \leq n_\rho,\, \rho \in R}$$

is a Hilbert basis of $\mathrm{L}^2(G)$.

Furthermore, for every $s \in G$, if we define the $n_\rho \times n_\rho$ matrix $M_\rho(s)$ by

$$M_\rho(s) = \left(\frac{1}{n_\rho} m_{ij}(s)\right),$$

then these matrices are invertible and satisfy the equations

$$M_\rho(st) = M_\rho(s) M_\rho(t) \qquad \text{and} \qquad M_\rho(s^{-1}) = (M_\rho(s))^*.$$

Thus, the map $s \mapsto M_\rho(s)$ is a continuous unitary representation in matrix form $M_\rho \colon G \to \mathbf{U}(n_\rho)$ of G in $\mathbb{C}^{n_\rho}$.

The above results are parts of Theorem 13.6, which constitutes the second half of the part of the Peter–Weyl theorem dealing with the structure of $\mathrm{L}^2(G)$ as Hilbert sum of finite-dimensional matrix algebras. But already, representations show their nose.

The unit of every two-sided ideal $\mathfrak{a}_\rho$ is $u_\rho = \sum_{j=1}^{n_\rho} m_{jj}$, and we show that the center of the Hilbert algebra $\mathrm{L}^2(G)$ is the Hilbert sum of the one-dimensional spaces $\mathbb{C}u_\rho$. The above results are shown in Section 13.1.

(2) Besides characters of groups and characters of algebras, there is one more kind of characters, namely, characters of finite-dimensional representations. For every $\rho \in R$, define the *character* χ_ρ *of* G *associated with the ideal* $\mathfrak{a}_\rho$ as the function given by

$$\chi_\rho(s) = \frac{1}{n_\rho}\, u_\rho(s) = \mathrm{tr}(M_\rho(s)), \qquad \text{for all } s \in G.$$

The character χ_{ρ_0} associated with $\mathfrak{a}_{\rho_0}$ is the constant function $\chi_{\rho_0}(s) = 1$ for all $s \in G$, called the *trivial character* of G. One of the main properties of the characters is that the family of characters $(\chi_\rho)_{\rho \in R}$ forms a Hilbert basis of the center of $\mathrm{L}^2(G)$; see Proposition 13.10. Other properties of the characters χ_ρ are shown in Section 13.2. In particular, if G is compact and abelian, then the characters are continuous homomorphisms of G to $\mathbf{U}(1)$, and they form a Hilbert basis for $\mathrm{L}^2(G)$.

(3) The second part of the Peter–Weyl theorem (Theorem 13.16) deals with unitary representations and is discussed in Section 13.3. This theorem asserts the following facts. Let $V : G \to \mathbf{U}(H)$ be a unitary representation of G in a separable Hilbert space H. Then H is a Hilbert sum of subspaces E_ρ invariant under V, and each nontrivial E_ρ is the Hilbert sum of invariant subspaces corresponding to irreducible representations of G. More precisely:

(1) For every $\rho \in R$, there is an orthogonal projection of H onto a closed subspace E_ρ (which may be reduced to (0)), and H is the Hilbert sum of the $E_\rho \neq (0)$.

(2) Every subspace $E_\rho \neq (0)$ is invariant under V, and the restriction V_ρ of V to E_ρ is a finite or countably infinite Hilbert sum of irreducible representations, all equivalent to M_ρ.

In particular, all the representations $M_\rho : G \to \mathbf{U}(n_\rho)$ occurring in Peter–Weyl I *are irreducible*, and since every unitary irreducible representations is equivalent to some representation of the form $M_{\overline{\rho}}$, the index set R corresponds to a complete set of unitary representations of G. Now because G is compact, there is a normalized Haar measure λ_G on G such that $\lambda_G(G) = 1$, and it can be shown that for *any finite-dimensional* representation $V : G \to \mathbf{GL}(H)$, there is an inner product on H such that V becomes a unitary representation. Then we define the *character χ_V of the representation* V by

$$\chi_V(s) = \mathrm{tr}(V(s)), \qquad s \in G.$$

Theorem 13.19 shows that two finite-dimensional unitary representations $V_1 : G \to \mathbf{U}(H_2)$ and $V_2 : G \to \mathbf{U}(H_2)$ of G are equivalent if and only if $\chi_{V_1} = \chi_{V_2}$. This confirms the importance of the characters; they determine the equivalence classes of finite-dimensional representations of a metrizable compact group.

The Fourier transform and the Fourier cotransform can also be generalized, but they involve the unitary irreducible representations of G which are usually very difficult to determine, so they are not so useful in practice.

In Section 13.4 we define a notion of Fourier transform and Fourier cotransform for a (metrizable) compact group G. Since for a nonabelian compact group the set of characters is not a group, the definition of the spaces $\mathrm{L}^p(\widehat{G})$ is more complicated. The *Fourier transform* $\mathcal{F}f$ of a function $f \in \mathrm{L}^1(G)$ is now a function with domain R, a complete set of irreducible unitary representations of G, such that for every $\rho \in R$,

$$\mathcal{F}(f)(\rho) = \int f(t)(M_\rho(t))^* \, d\lambda_g(t).$$

The Fourier transform defined above is the natural generalization of the definition of the Fourier transform when G is an abelian compact group (Definition 10.3),

$$\mathcal{F}(f)(\chi) = \int f(s)\overline{\chi(s)} \, d\lambda(s) = \int f(s)\chi(s^{-1}) \, d\lambda(s);$$

the character χ is replaced by the irreducible representation M_ρ.

The definition of $\mathcal{F}(f)(\rho)$ implies that $\mathcal{F}(f)(\rho)$ is a linear map from $\mathbb{C}^{n_\rho}$ to itself (since $(M_\rho(t))^*$ is a matrix). Thus, $\mathcal{F}(f) \in \prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$. Every element $F \in \prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$ is an R-indexed sequence $F = (F(\rho))_{\rho\in R}$ of $n_\rho \times n_\rho$ matrices $F(\rho)$. These sequences can be added and rescaled componentwise, so we obtain a vector space.

It is natural to define $\widehat{G}$ as R, but the vector space $\prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$ is too big. Thus, we define some normed vector spaces $\mathrm{L}^p(\widehat{G})$ which are subspaces of $\prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$. For this, we need to define some norms due to von Neumann.

We obtain some Banach spaces $\mathrm{L}^1(\widehat{G})$, $\mathrm{L}^2(\widehat{G})$, and $\mathrm{L}^\infty(\widehat{G})$; the space $\mathrm{L}^2(\widehat{G})$ is a Hilbert space. The following result is obtained (Theorem 13.25). Let G be a compact group.

(1) The map $f \mapsto \mathcal{F}(f)$ is a non norm-increasing injective involutive algebra homomorphism from $\mathrm{L}^1(G)$ into $\mathrm{L}^\infty(\widehat{G})$. In particular, for all $f, g \in \mathrm{L}^1(G)$, for all $\rho \in R$, we have
$$(\mathcal{F}(f * g))(\rho) = \mathcal{F}(g)(\rho) \circ \mathcal{F}(f)(\rho).$$

(2) For every $\rho \in R$, the map $f \mapsto \mathcal{F}(f)(\rho)$ is an algebra representation of $\mathrm{L}^1(G)$ in $\mathbb{C}^{n_\rho}$.

We also have a version of Plancherel's theorem (see Theorem 13.28). If G is a compact group, then the map $f \mapsto \mathcal{F}(f)$ is an isometric isomorphism between the Hilbert space $\mathrm{L}^2(G)$ and the Hilbert space $\mathrm{L}^2(\widehat{G})$.

We can also define a notion of Fourier cotransform and there are versions of Fourier inversion. For any $F \in \prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$, the *Fourier cotransform* $\overline{\mathcal{F}}(F)$ of F is the function on G given by
$$\overline{\mathcal{F}}(F)(s) = \sum_{\rho\in R} n_\rho \operatorname{tr}(F(\rho)M_\rho(s)), \qquad s \in G.$$
Of course, there are convergence issues. It can be shown (Theorem 13.29) that if $F \in \mathrm{L}^1(\widehat{G})$, then the map
$$s \mapsto (\overline{\mathcal{F}}(F))(s) = \sum_{\rho\in R} n_\rho \operatorname{tr}(F(\rho)M_\rho(s))$$
converges uniformly to a continuous function f. Furthermore, we have the Fourier inversion formula
$$(\overline{\mathcal{F}}(\mathcal{F}(f)))(s) = \sum_{\rho\in R} n_\rho \operatorname{tr}(\mathcal{F}(f)(\rho)M_\rho(s)), \qquad s \in G.$$
Also, Fourier inversion holds for $\mathrm{L}^2(G)$. (Theorem 13.31). The Fourier cotransform $\overline{\mathcal{F}}(F) \in \mathrm{L}^2(G)$ of any $F \in \mathrm{L}^2(\widehat{G})$ converges in the L^2-norm, and for every $f \in \mathrm{L}^2(G)$, we have
$$f(s) = (\overline{\mathcal{F}}(\mathcal{F}(f)))(s) = \sum_{\rho\in R} n_\rho \operatorname{tr}\left(\mathcal{F}(f)(\rho)M_\rho(s)\right), \qquad s \in G$$
in the L^2-norm.

13.1 The Peter–Weyl Theorem, I

The theorem below is the first of several theorems describing the structure of the involutive Banach algebra $\mathrm{L}^2(G)$, where G is a metrizable compact group. By Proposition 11.17, the Banach algebra $\mathrm{L}^2(G)$ is a complete separable Hilbert algebra, so Theorem 11.31 is applicable and yields most of a deep theorem first proved by Peter and Weyl (1927); see Theorem 13.2.

No matter how it is approached, the proof of the Peter–Weyl theorem (Theorem 13.2) is hard. We follow Dieudonné's exposition [21] (Sections 1-4). The disadvantage in doing so is that it requires some material on Hilbert algebras from Chapter 11, in particular, Theorem 11.31, whose proof is long. The advantage is that we obtain a sharper and more informative version of the Peter–Weyl theorem.

Since G is compact, it is unimodular, and so it has a Haar measure λ which is both left and right invariant. We also assume that λ is normalized so that $\lambda(G) = 1$.

When we describe operations on elements of $\mathrm{L}^2(G)$, such as $f*g$ for the convolution $[f]*[g]$ of $[f]$ and $[g]$ in $\mathrm{L}^2(G)$, we mean the equivalence class $[f*g]$ of $f*g$, where $f, g \in \mathcal{L}^2(G)$ are representatives in the equivalence classes $[f], [g] \in \mathrm{L}^2(G)$ (where two functions are equivalent iff they are equal almost everywhere). To be perfectly rigorous, we should check that these constructions do not depend on the representatives chosen in these equivalence classes, but we will not inflict such verifications on the reader.

Recall that by Proposition 8.49, if $f, g \in \mathrm{L}^2(G)$, then $f * g \in \mathcal{C}_0(G; \mathbb{C})$. In particular, $f * g$ is continuous.

Definition 13.1. A function $h \in \mathcal{L}^2(G)$ is *central* if its class in $\mathrm{L}^2(G)$ belongs to the center of $\mathrm{L}^2(G)$. This means that for every $f \in \mathcal{L}^2(G)$, we have $f * h = h * f$ almost everywhere.

The following auxiliary result is needed.

Proposition 13.1. *Let G be a compact group. A continuous function $h \in \mathcal{L}^2(G)$ is central if and only if $h(sts^{-1}) = h(t)$ for all $s, t \in G$. The class of every central function $f \in \mathcal{L}^2(G)$ belongs to the center of $\mathcal{M}^1(G)$.*

Proof. If $f*h = h*f$ almost everywhere, since $f*h$ and $h*f$ are continuous, we must have $f * h = h * f$ everywhere. Since G is compact, it is unimodular, so have

$$f * h(s) = \int f(t)h(t^{-1}s)\, d\lambda(t) = \int f(t^{-1})h(ts)\, d\lambda(t),$$

and

$$h * f(s) = \int h(t)f(t^{-1}s)\, d\lambda(t) = \int h(st)f(t^{-1})\, d\lambda(t).$$

Thus, for every $s \in G$, we have

$$\int_G f(t^{-1})(h(st) - h(ts))\, d\lambda(t) = 0.$$

The above implies that $h(st) = h(ts)$ for all t in the complement of a set of measure zero (depending on s), but since h is continuous, this subset must be empty. It follows that $h(st) = h(ts)$ for all $s, t \in G$, and if we replace t by ts^{-1}, we obtain $h(sts^{-1}) = h(t)$ for all $s, t \in G$.

By the formula

$$(\mu * f)(s) = \int f(t^{-1}s)\, d\mu(t)$$

in Definition 8.24, and the formula

$$(f * \mu)(s) = \int f(st^{-1})\Delta(t^{-1})\, d\mu(t)$$

in Definition 8.25, since G is compact, it is unimodular, so $\Delta(t^{-1}) = 1$, and if f is a central function, then it is easy to show that $\mu * (fd\lambda) = (fd\lambda) * \mu$, which shows that $fd\lambda$ is in the center of $\mathcal{M}^1(G)$. Recall that $\mathcal{L}^1(G)$ is embedded in $\mathcal{M}^1(G)$ by mapping f to the measure $fd\lambda$ and we usually identify f and $fd\lambda$. $\square$

Theorem 13.2. *(Peter–Weyl theorem, I) Let G be a metrizable compact group. The complete Hilbert algebra $\mathrm{L}^2(G)$ is the Hilbert sum of a finite or countably infinite family $(\mathfrak{a}_\rho)_{\rho \in R}$ of topologically simple Hilbert algebras $\mathfrak{a}_\rho$ of finite dimension n_ρ^2, where each $\mathfrak{a}_\rho$ is a minimal two-sided ideal of $\mathrm{L}^2(G)$ isomorphic to the matrix algebra $\mathrm{M}_{n_\rho}(\mathbb{C})$, and $\mathfrak{a}_h\mathfrak{a}_k = (0)$ for all $h \neq k$. The elements of $\mathfrak{a}_\rho$ are classes of continuous functions on G; the unit element of $\mathfrak{a}_\rho$ is the class of a continuous function u_ρ such that $\check{\overline{u}}_\rho = u_\rho$, and the orthogonal projection of $\mathrm{L}^2(G)$ onto $\mathfrak{a}_\rho$ is the map $f \mapsto f * u_\rho = u_\rho * f$, for every $f \in \mathcal{L}^2(G)$. Furthermore, for every $f \in \mathcal{L}^2(G)$, we have*

$$f = \sum_{\rho \in R} f * u_\rho,$$

where the series on the right-hand side is commutatively convergent.

Proof. We follow Dieudonné's proof [21] (Chapter XXI, Section 2, Theorem 21.2.3). Since By Proposition 11.17, the Banach algebra $A = \mathrm{L}^2(G)$ is a complete separable Hilbert algebra, Theorem 11.31 shows that $\mathrm{L}^2(G)$ is the Hilbert sum of a finite or countably infinite family $(\mathfrak{a}_\rho)_{\rho \in R}$ of two-sided ideals which are topologically simple Hilbert algebras, and $\mathfrak{a}_h\mathfrak{a}_k = (0)$ for all $h \neq k$. If we can prove that every $\mathfrak{a}_\rho$ is finite-dimensional, we will be done because then, by Theorem 11.32, each $\mathfrak{a}_\rho$ will be a finite Hilbert sum of isomorphic minimal left ideals $\mathfrak{l}_j$, each generated by a self-adjoint irreducible idempotent e_j, and the sum of these idempotents will be the unit $\mathbf{1}_\rho$ of the algebra $\mathfrak{a}_\rho$. If u_ρ is a function whose class is $\mathbf{1}_\rho$, every element of $\mathfrak{a}_\rho$ will be the class of a function of the form $f * u_\rho$, with $f \in \mathcal{L}^2(G)$, which is a continuous function by Proposition 8.49. The other assertions of the theorem follow from Theorem 11.28, since the orthogonal projection of f onto $\mathfrak{a}_\rho$ is of the form $\sum_{j=1}^{n_\rho} f * e_j = f * u_\rho$, since $u_\rho = \sum_{j=1}^{n_\rho} e_j$.

By Proposition 11.33, if we can prove that there is a nonzero element in the center of $\mathfrak{a}_\rho$, then $\mathfrak{a}_\rho$ will be finite-dimensional. This is a consequence of the following proposition.

Proposition 13.3. *For every closed two-sided ideal $\mathfrak{b} \neq (0)$ in $\mathrm{L}^2(G)$, there is some nonzero element $c \in \mathfrak{b}$ in the center of $\mathrm{L}^2(G)$.*

Proof. The proof of Proposition 13.3 makes use of the following result.

Proposition 13.4. *Let $\mathfrak{b}$ be a closed subspace (as a vector space) of $\mathrm{L}^2(G)$. Then the following conditions are equivalent.*

(1) $\mathfrak{b}$ is a left ideal in $\mathrm{L}^2(G)$.

(2) $\mathfrak{b}$ is invariant under the regular representation $\mathbf{R}_{\mathrm{ext}}$ of $\mathrm{L}^1(G)$ in $\mathrm{L}^2(G)$ (see Definition 12.12).

(3) For every function f whose class is in $\mathfrak{b}$ and for all $s \in G$, the class of $\delta_s * f = \lambda_s(f)$ is in $\mathfrak{b}$.

Proof. The equivalence of (2) and (3) follows from Theorem 12.15 applied to the regular representation $\mathbf{R}_{\mathrm{ext}}$. It is clear that (2) implies (1). On the other hand, Theorem 7.10 implies that $\mathcal{L}^2(G)$ is dense in $\mathcal{L}^1(G)$, and by Proposition 8.48, the map $f \mapsto f * g$ from $\mathcal{L}^1(G)$ to $\mathcal{L}^2(G)$ is continuous for every $g \in \mathcal{L}^2(G)$, so (1) implies (2). ☐

Proposition 13.4 also applies to right ideals in (1) and to $f * \delta_s = \rho_{s^{-1}}(f)$ in (3).

We can now prove Proposition 13.3. First, let us prove that $\mathfrak{b}$ contains the class of a continuous function f such that f is not the zero function (we need a continuous function, because we want to construct a central function, and to apply Proposition 13.1 such a function must be continuous). Indeed, let $g \in \mathfrak{b}$ be a function not zero almost everywhere. Then the class of the function $g * \check{\overline{g}}$ also belongs to $\mathfrak{b}$. But $g * \check{\overline{g}}$ is continuous (by Proposition 8.49), and since the definition of the convolution of functions implies that

$$(g * \check{\overline{g}})(e) = \int g(s)\check{\overline{g}}(s^{-1})\, d\lambda(s) = \int g(s)\overline{g}(s)\, d\lambda(s) = \|g\|_2^2 > 0,$$

we can pick $f = g * \check{\overline{g}}$ in $\mathfrak{b}$. Consider the function h given by

$$h(t) = \int_G f(sts^{-1})\, d\lambda(s).$$

Since the function $(x, y, z) \mapsto f(xyz)$ is uniformly continuous on $G \times G \times G$, we see that h is continuous, and since $h(e) = f(e) \neq 0$, it is not the zero function. For all $s \in G$, we have

$$h(xtx^{-1}) = \int_G f((sx)t(sx)^{-1})\, d\lambda(s) = h(t),$$

since the Haar measure on a compact group is left and right invariant. By Proposition 13.1, the function h belongs to the center of $\mathrm{L}^2(G)$. It remains to show that the class of h belongs to $\mathfrak{b}$. Since $\mathrm{L}^2(G) = \mathfrak{b} \oplus \mathfrak{b}^{\perp}$ as a Hilbert sum, and $\mathfrak{b}^{\perp}$ is also a two-sided ideal by Proposition 11.19, it suffices to prove that $\langle h, w\rangle = 0$ for all $w \in \mathfrak{b}^{\perp}$ (we are abusing notation, h and

w should be equivalence classes). Using the fact that the Haar measure is left and right invariant, and Fubini's theorem, we have

$$\begin{aligned}\langle h, w\rangle &= \int \overline{w(t)} \int f(sts^{-1})\, d\lambda(s)\, d\lambda(t)\\ &= \int \left(\int \overline{w(t)} f(sts^{-1})\, d\lambda(t)\right) d\lambda(s)\\ &= \int \left(\int \overline{w(s^{-1}ts)} f(t)\, d\lambda(t)\right) d\lambda(s).\end{aligned}$$

Since $w \in \mathfrak{b}^{\perp}$, by Proposition 13.4 and its version for right ideals, the class of $\delta_s * w * \delta_{s^{-1}}$ also belongs to $\mathfrak{b}^{\perp}$. Since G is unimodular (see after Definition 8.24 and Definition 8.25), we have

$$(\delta_s * w * \delta_{s^{-1}})(t) = w(s^{-1}ts),$$

and since $f \in \mathfrak{b}$ and $\delta_s * w * \delta_{s^{-1}} \in \mathfrak{b}^{\perp}$,

$$\int \overline{w(s^{-1}ts)} f(t)\, d\lambda(t) = 0,$$

which concludes the proof of Proposition 13.3. $\square$

This also concludes the proof of Theorem 13.2. $\square$

We will identify every element of $\mathfrak{a}_\rho$ with the unique continuous function belonging to this class.

Our next goal is to get a better understanding of the structure of the algebras $\mathfrak{a}_\rho$ by decomposing them as finite Hilbert sums of minimal left ideals, and by choosing some Hilbert bases in these ideals.

For every $\rho \in R$, we assume that we have chosen a decomposition of $\mathfrak{a}_\rho$ as a finite Hilbert sum of n_ρ minimal left ideals $\mathfrak{l}_j = \mathfrak{a}_\rho * m_j$,

$$\mathfrak{a}_\rho = \bigoplus_{j=1}^{n_\rho} \mathfrak{l}_j = \bigoplus_{1 \le j \le n_\rho} \mathfrak{a}_\rho * m_j,$$

where the $\mathfrak{a}_\rho * m_j$, also denoted $\mathfrak{l}_j^{(\rho)}$, are pairwise isomorphic and orthogonal, and where m_j is a self-adjoint irreducible idempotent ($1 \le j \le n_\rho$), so that the unit of $\mathfrak{a}_\rho$ is

$$u_\rho = \sum_{j=1}^{n_\rho} m_j; \tag{u_ρ}$$

see Theorem 11.32.[1] By Proposition 11.21, since m_j and m_j are orthogonal when $i \neq j$, we have $m_i * m_j = 0$ if $i \neq j$. Let $(a_j)_{1 \le j \le n_\rho}$ be a Hilbert basis of $\mathfrak{l}_1 = \mathfrak{a}_\rho * m_1$, such that $a_j \in m_j * \mathfrak{a}_\rho * m_1$.

[1]Note that the m_j are the e_j used in the proof of Theorem 11.32.

Since $a_j \in m_j * \mathfrak{a}_\rho * m_1$, we have

$$a_j * m_1 = a_j, \qquad m_j * a_j = a_j, \quad 1 \leq j \leq n_\rho.$$

We have the following proposition, which is in fact part of the proof of Theorem 11.32. To simplify notation, write $a^* = \check{\overline{a}}$.

Proposition 13.5. *The inner products $\langle m_j, m_j \rangle$ have the same value $\gamma > 0$, and we have*

$$a_j * a_j^* = \gamma m_j, \qquad a_j^* * a_j = \gamma m_1, \qquad 1 \leq j \leq n_\rho.$$

Proof. Since each $\mathfrak{a}_\rho$ is a Hilbert algebra, $\mathfrak{a}_\rho = \mathfrak{a}_\rho^*$ and $\mathfrak{a}_\rho \mathfrak{a}_\rho = \mathfrak{a}_\rho$. Since the m_j are self-adjoint idempotent ($m_j * m_j = m_j$ and $m_j^* = m_j$) and $(a * b)^* = b^* * a^*$, as $a_j \in m_j * \mathfrak{a}_\rho * m_1$, we have $a_j * a_j^* \in m_j * \mathfrak{a}_\rho * m_j$. By Theorem 11.30(2), we must have $a_j * a_j^* = \lambda_j m_j$, for some $\lambda_j \in \mathbb{C}$, with $\lambda_j \neq 0$. Similarly, $a_j^* * a_j \in m_1 * \mathfrak{a}_\rho * m_1$, so $a_j^* * a_j = \lambda_j' m_1$ for some $\lambda_j' \in \mathbb{C}$, with $\lambda_j' \neq 0$. We claim that $\lambda_j = \lambda_j'$.

First, we have

$$a_j * a_j^* * a_j * a_j^* = \lambda_j m_j * \lambda_j m_j = \lambda_j^2 m_j * m_j = \lambda_j^2 m_j,$$

and second

$$a_j * a_j^* * a_j * a_j^* = a_j * \lambda_j' m_1 * a_j^* = \lambda_j' a_j * m_1 * a_j^* = \lambda_j' a_j * a_j^* = \lambda_j' \lambda_j m_j,$$

since $a_j * m_1 = a_j$. Therefore, $\lambda_j^2 = \lambda_j' \lambda_j$, and since λ_j and λ_j' are nonzero, we deduce that $\lambda_j = \lambda_j'$, for $j = 1, \ldots, n_\rho$.

We also have

$$1 = \langle a_j, a_j \rangle = \langle a_j, m_j * a_j \rangle = \langle a_j * a_j^*, m_j \rangle = \lambda_j \langle m_j, m_j \rangle,$$

and

$$1 = \langle a_j, a_j \rangle = \langle a_j, a_j * m_1 \rangle = \langle a_j^* * a_j, m_1 \rangle = \lambda_j \langle m_1, m_1 \rangle.$$

Since $\lambda_j \neq 0$, we deduce that

$$\langle m_j, m_j \rangle = \langle m_1, m_1 \rangle, \qquad 1 \leq j \leq n_\rho.$$

and so

$$\lambda_j = \langle m_1, m_1 \rangle^{-1} = \gamma, \qquad 1 \leq j \leq n_\rho. \qquad \square$$

Thus there is some $\gamma > 0$ such that

$$a_j * \check{\overline{a}}_j = \gamma m_j, \qquad \check{\overline{a}}_j * a_j = \gamma m_1, \qquad 1 \leq j \leq n_\rho,$$

Since $a_i \in m_i * \mathfrak{a}_\rho * m_1$ and $\check{\overline{a}}_j \in m_1 * \mathfrak{a}_\rho * m_j$, we have $a_i * \check{\overline{a}}_j \in m_i * \mathfrak{a}_\rho * m_j$ and since $\mathfrak{l}_j = \mathfrak{a}_\rho * m_j$ is a left ideal, $a_i * \check{\overline{a}}_j \in \mathfrak{l}_j$.

Definition 13.2. For all j, k with $1 \leq j, k \leq n_\rho$, let

$$m_{jk} = \gamma^{-1} a_j * \check{\overline{a}}_k \in \mathfrak{l}_k.$$

In particular, $m_{jj} = m_j$.

Since $a_h \in m_h * \mathfrak{a}_\rho * m_1$, $\check{\overline{a}}_k \in m_1 * \mathfrak{a}_\rho * m_k$, and $m_k * m_h = 0$ whenever $h \neq k$, we have $\check{\overline{a}}_k * a_h = 0$ whenever $h \neq k$, so

$$m_{jk} * a_h = \delta_{kh} a_j. \tag{$*$}$$

Remark: Observe that the m_{ij} are the e_{mn} introduced during the proof of Theorem 11.32.

We will also write $m_{ij}^{(\rho)}$ instead of m_{ij}. The following result reveals that some representations are hidden in the Hilbert sum of the $\mathfrak{a}_\rho$.

Theorem 13.6. *With the above notation, the following properties hold.*

(1) For every j with $1 \leq j \leq n_\rho$, the $(m_{ij})_{1 \leq i \leq n_\rho}$ form an orthogonal basis of $\mathfrak{l}_j$, and the $(m_{ij})_{1 \leq i,j \leq n_\rho}$ form an orthogonal basis of $\mathfrak{a}_\rho = \bigoplus_{j=1}^{n_\rho} \mathfrak{l}_j$.

(2) We have $m_{ji} = \check{\overline{m}}_{ij}$ and $m_{ij} * m_{hk} = \delta_{jh}\, m_{ik}$.

(3) We have $\langle m_{ij}, m_{ij} \rangle = n_\rho$ and $m_{ij}(e) = n_\rho\, \delta_{ij}$, for all i, j with $1 \leq i, j \leq n_\rho$ (in other words, $\gamma = (n_\rho)^{-1}$). Thus the family of functions

$$\left(\frac{1}{\sqrt{n_\rho}} m_{ij}^{(\rho)} \right)_{1 \leq i,j \leq n_\rho,\, \rho \in R}$$

is a Hilbert basis of $\mathrm{L}^2(G)$.

(4) For every $s \in G$, if we define the $n_\rho \times n_\rho$ matrix $M_\rho(s)$ by

$$M_\rho(s) = \left(\frac{1}{n_\rho} m_{ij}(s) \right),$$

then these matrices are invertible and satisfy the equations

$$M_\rho(st) = M_\rho(s) M_\rho(t) \qquad \text{and} \qquad M_\rho(s^{-1}) = (M_\rho(s))^*.$$

Thus, the map $s \mapsto M_\rho(s)$ is a continuous unitary representation in matrix form $M_\rho \colon G \to \mathbf{U}(n_\rho)$ of G in $\mathbb{C}^{n_\rho}$, for the standard hermitian inner product $\sum_{j=1}^{n_\rho} \alpha_j \overline{\beta}_j$.

Proof. (1) Since $m_{ij} = \gamma^{-1} a_i * \check{\overline{a}}_j = \gamma^{-1} a_i * a_j^*$, it suffices to prove that for any fixed j,

$$\langle a_i * a_j^*, a_k * a_j^* \rangle = 0 \qquad \text{for all } i \neq k.$$

Since $a_j^* * a_j = \gamma m_1$ (by Proposition 13.5), $m_1^* = m_1$, and $a_i * m_1 = a_i$, we have

$$\begin{aligned}\langle a_i * a_j^*, a_k * a_j^*\rangle &= \langle a_i * a_j^* * a_j, a_k\rangle \\ &= \langle a_i * \gamma m_1, a_k^*\rangle \\ &= \gamma\langle a_i, a_k\rangle = 0,\end{aligned}$$

since the a_i are pairwise orthogonal.

(2) We have

$$\begin{aligned}m_{ji}(s) &= \gamma^{-1}(a_j * \check{\overline{a}}_i)(s) \\ &= \gamma^{-1}\int a_j(t)\check{\overline{a}}_i(t^{-1}s)\, d\lambda(t) \\ &= \gamma^{-1}\overline{\int \overline{a_j(t)} a_i(s^{-1}t)\, d\lambda(t)} \\ &= \gamma^{-1}\overline{\int a_i(t)\overline{a_j(st)}\, d\lambda(t)} \\ &= \gamma^{-1}\overline{\int a_i(t)\check{\overline{a}}_j(t^{-1}s^{-1})\, d\lambda(t)} \\ &= \gamma^{-1}\overline{(a_i * \check{\overline{a}}_j)}(s^{-1}) \\ &= \overline{m}_{ij}(s^{-1}) = \check{\overline{m}}_{ij}(s),\end{aligned}$$

where Proposition 7.24 was used to derive the third equation. Since $a_i \in m_i * \mathfrak{a}_\rho * m_1$, we have $a_i^* * a_j \in m_1 * \mathfrak{a}_\rho * m_i * m_j * \mathfrak{a}_\rho * m_1 = 0$ whenever $i \neq j$, since the m_i are pairwise orthogonal self-adjoint irreducible idempotents, and thus $m_i * m_j = 0$ whenever $i \neq j$. Consequently

$$a_j^* * a_h = \delta_{jh}\gamma m_1,$$

and thus

$$m_{ij} * m_{hk} = \gamma^{-1} a_i * a_j^* * \gamma^{-1} a_h * a_k^* = \gamma^{-2} a_i * \delta_{jh}\gamma m_1 * a_k^* = \delta_{jh}\gamma^{-1} a_i * a_k^* = \delta_{jh} m_{ik}.$$

(3) Since $\mathfrak{a}_\rho$ is a Hilbert algebra, by (2) and (2') (see Definition 11.14), we have

$$\langle m_{ij}, m_{ij}\rangle = \gamma^{-2}\langle a_i * \check{\overline{a}}_j, a_i * \check{\overline{a}}_j\rangle = \gamma^{-2}\langle \check{\overline{a}}_i * a_i, \check{\overline{a}}_j * a_j\rangle = \langle m_1, m_1\rangle. \tag{$*_1$}$$

To compute this value, observe that for every k, by Proposition 13.4, the function $t \mapsto m_{ik}(st)$ belongs to $\mathfrak{l}_k$ for all $s \in G$. Thus we can write

$$m_{ik}(st) = \sum_{j=1}^{n_\rho} c_{ij}(s) m_{jk}(t). \tag{$*_2$}$$

On the other hand, using the fact that $m_{1k} = \check{\overline{m}}_{k1}$ and (2), we have

$$m_{jk}(t) = (m_{j1} * m_{1k})(t) = \int_G m_{j1}(tx)m_{1k}(x^{-1})\, d\lambda(x) = \int_G m_{j1}(tx)\overline{m_{k1}(x)}\, d\lambda(x),$$

which yields

$$m_{jk}(e) = \langle m_{j1}, m_{k1}\rangle, \tag{$*_3$}$$

and if we let $t = e$ in $(*_2)$, using the orthogonality properties of the m_{ij} and the fact that $m_{jj} = m_j$ and $\langle m_j, m_j\rangle = \langle m_1, m_1\rangle$, we get

$$m_{ik}(s) = \langle m_1, m_1\rangle c_{ik}(s). \tag{$*_4$}$$

If we let $s = t^{-1}$ and $i = k = 1$ in $(*_2)$ and $j = k = 1$ in $m_{jk}(e) = \langle m_{j1}, m_{k1}\rangle$, we get

$$\langle m_1, m_1\rangle = m_1(e) = \sum_{j=1}^{n_\rho} c_{1j}(s)m_{j1}(s^{-1}) = \sum_{j=1}^{n_\rho} c_{1j}(s)\overline{m_{1j}(s)},$$

and using $(*_4)$,

$$c_{1j} = \frac{1}{\langle m_1, m_1\rangle}\, m_{1j},$$

we obtain

$$\sum_{j=1}^{n_\rho} m_{1j}(s)\overline{m_{1j}(s)} = \langle m_1, m_1\rangle^2. \tag{$*_5$}$$

Since by $(*_1)$ we have

$$\langle m_1, m_1\rangle = \langle m_{1j}, m_{1j}\rangle = \int m_{1j}(s)\overline{m_{1j}(s)}\, d\lambda(s) \tag{$*_6$}$$

and since $\langle m_1, m_1\rangle^2$ is a constant and λ is the normalized Haar measure, if we integrate both sides of $(*_5)$ we obtain

$$\sum_{j=1}^{n_\rho} \int_G m_{1j}(s)\overline{m_{1j}(s)}\, d\lambda(s) = \int_G \langle m_1, m_1\rangle^2\, d\lambda(s) = \langle m_1, m_1\rangle^2 \int_G d\lambda(s) = \langle m_1, m_1\rangle^2,$$

and by $(*_6)$ we have

$$\sum_{j=1}^{n_\rho} \langle m_1, m_1\rangle = n_\rho\langle m_1, m_1\rangle = \langle m_1, m_1\rangle^2,$$

so $\langle m_{ij}, m_{ij}\rangle = \langle m_1, m_1\rangle = n_\rho$, which proves (3). The equations in (4) follow immediately from (2), $(*_2)$, and $(*_4)$. $\square$

As in Definition 12.2 the unitary matrix representation $M_\rho\colon G \to \mathbf{U}(n_\rho)$ defines (with a small abuse of notation) the representation $M_\rho\colon G \to \mathbf{U}(\mathbb{C}^{n_\rho})$ given by

$$(M_\rho(s))(z) = M_\rho(s)z, \quad z \in \mathbb{C}^{n_\rho},\ s \in G.$$

We usually identify these two variants. We will see later that the representations M_ρ are irreducible. The center of $\mathrm{L}^2(G)$ is characterized as follows.

Proposition 13.7. *Let G be a metrizable compact group. The center of the Hilbert algebra $\mathrm{L}^2(G)$ is the Hilbert sum of the one-dimensional spaces $\mathbb{C}u_\rho$ (with $\rho \in R$). In particular, if G is commutative, then every ideal $\mathfrak{a}_\rho$ is one-dimensional ($n_\rho = 1$).*

Proof. Since u_ρ is the unit element of $\mathfrak{a}_\rho$ and since $\mathfrak{a}_\rho * \mathfrak{a}_{\rho'} = (0)$ whenever $\rho \neq \rho'$, we see that u_ρ belongs to the center of $\mathrm{L}^2(G)$. If the class of a function $f \in \mathcal{L}^2(G)$ belongs to the center of $\mathrm{L}^2(G)$, since u_ρ also belongs to this center, we deduce that the class of $f * u_\rho \in \mathfrak{a}_\rho$ belongs to the center $\mathrm{L}^2(G)$, but since $\mathfrak{a}_\rho$ is topologically simple, complete, separable algebra, by Proposition 11.33, the center of $\mathfrak{a}_\rho$ is one-dimensional, so $f * u_\rho = c_\rho u_\rho$ for some $c_\rho \in \mathbb{C}$. Since by Theorem 13.2, we have

$$f = \sum_{\rho \in R} f * u_\rho$$

for every $f \in \mathcal{L}^2(G)$, we must have $f = \sum_{\rho \in R} c_\rho u_\rho$, as claimed. ☐

Since the group G is compact, for every $f \in \mathrm{L}^2(G)$ and every constant function α, we have

$$f * \alpha = \alpha * f = \alpha \left(\int_G f(s)\, d\lambda(s) \right).$$

Therefore the (complex) constant functions form a two-sided ideal in $\mathrm{L}^2(G)$, and thus must be an ideal of the form $\mathfrak{a}_{\rho_0}$.

Definition 13.3. The ideal $\mathfrak{a}_{\rho_0}$ is called the *trivial ideal*.

The corresponding representation M_{ρ_0} is one-dimensional, and $M_{\rho_0}(s) = 1$ for all $s \in G$. In other words, M_{ρ_0} is the trivial representation of G. For all $\rho \neq \rho_0$, since the spaces $\mathfrak{a}_\rho$ and $\mathfrak{a}_{\rho_0}$ are orthogonal, we have

$$\int_G m_{ij}^{(\rho)}(s)\, d\lambda(s) = \int_G m_{ij}^{(\rho)}(s)\overline{1}\, d\lambda(s) = 0.$$

Therefore, for every $\rho \neq \rho_0$, we have

$$\int_G m_{ij}^{(\rho)}(s)\, d\lambda(s) = 0. \tag{$*_{\rho\neq\rho_0}$}$$

We also have the following results.

Proposition 13.8. *Let G be a metrizable compact group. With the notation as above, the following properties hold.*

(1) If f and g are two continuous functions in $\mathcal{C}(G;\mathbb{C})$, then we have

$$f * g = \sum_{\rho \in R} \left(\sum_{1 \leq i,j \leq n_\rho} \frac{1}{n_\rho} \langle g, m_{ij}^{(\rho)} \rangle (f * m_{ij}^{(\rho)}) \right),$$

where the family on the right-hand side converges for the topology of uniform convergence.

(2) The family of continuous functions $\{m_{ij}^{(\rho)} \mid \rho \in R, 1 \leq i,j \leq n_\rho\}$ is an orthogonal system that is dense in $\mathcal{C}(G;\mathbb{C})$ for the topology of uniform convergence.

Proof. (1) From Theorem 13.6 which says that the family of functions

$$\left(\frac{1}{\sqrt{n_\rho}} m_{ij}^{(\rho)} \right)_{1 \leq i,j \leq n_\rho,\, \rho \in R}$$

is a Hilbert basis of $\mathrm{L}^2(G)$, we have

$$g = \sum_{\rho \in R} \sum_{1 \leq i,j \leq n_\rho} \frac{1}{n_\rho} \langle g, m_{ij}^{(\rho)} \rangle \, m_{ij}^{(\rho)}.$$

Since the map $h \mapsto f * h$ is a continuous map from $\mathrm{L}^2(G)$ to $\mathcal{C}(G;\mathbb{C})$, (since by Proposition 8.49, $\|f * h\|_\infty \leq \|f\|_2 \, \|h\|_2$), we can apply convolution to both sides, and we get the equation in (1).

(2) By Proposition 8.50, for every continuous function $g \in \mathcal{C}(G;\mathbb{C})$, there is some continuous function f such that $\|f * g - g\|_\infty$ can be made arbitrarily small. But, for every $\rho \in R$, the functions $f * m_{ij}^{(\rho)}$ belong to the two-sided ideal $\mathfrak{a}_\rho$, and so they are (complex) linear combinations of the $m_{hk}^{(\rho)}$ with $1 \leq h,k \leq n_\rho$. Therefore, the formula of (1) for $f * g$ shows that $f * g$ can be expressed in terms of the $m_{hk}^{(\rho)}$, which proves (2). $\square$

In Section 15.10 we will need the following result.

Proposition 13.9. *For any unitary $n_\rho \times n_\rho$ matrix P, for every $s \in G$, let $Q_\rho(s) = P^* M_\rho(s) P$. The matrices $Q_\rho(s) = (q_{ij}(s))$ define n_ρ^2 functions $q_{ij} \in \mathfrak{a}_\rho$ which are linear combinations of the functions m_{ij}, where $m_{ij}(s) \in M_\rho(s)$, and satisfy the following properties:*

(1) The $(q_{ij})_{1 \leq i,j \leq n_\rho}$ form an orthogonal basis of $\mathfrak{a}_\rho$.

(2) We have $q_{ji} = \check{\overline{q}}_{ij}$ and $q_{ij} * q_{hk} = \delta_{jh} \, q_{ik}$.

(3) *We have* $\langle q_{ij}, q_{ij}\rangle = n_\rho$ *and* $q_{ij}(e) = n_\rho\,\delta_{ij}$*, for all* i, j *with* $1 \leq i, j \leq n_\rho$*,*

(4) *The map* $s \mapsto Q_\rho(s)$ *is unitary representation in matrix form* $Q_\rho \colon G \to \mathbf{U}(n_\rho)$ *of* G *in* $\mathbb{C}^{n_\rho}$*, equivalent to the unitary representation* $M_\rho \colon G \to \mathbf{U}(n_\rho)$*.*

(5) *If* $\mathfrak{l}_j$ *is the minimal left ideal of* $\mathfrak{a}_\rho$ *spanned by the* j*th column* M_ρ^j *of* M_ρ*,*

$$\mathfrak{l}_j = \bigoplus_{i=1}^{n_\rho} \mathbb{C} m_{ij}^{(\rho)},$$

then the j*th column of* $Q_\rho = P^* M_\rho P$ *spans a minimal ideal* $\mathfrak{l}_j^Q$ *of* $\mathfrak{a}_\rho$ *(of dimension* n_ρ*) given by*

$$\mathfrak{l}_j^Q = \left\{ \sum_{h=1}^{n_\rho} p_{hj} \left(\sum_{k=1}^{n_\rho} \mu_k m_{kh} \right) \;\middle|\; \mu = (\mu_1, \ldots, \mu_{n_\rho}) \in \mathbb{C}^{n_\rho} \right\},$$

where every $\sum_{k=1}^{n_\rho} \mu_k m_{kh} \in \mathfrak{l}_h$ *is a linear combination of the entries of the* h*th column of* M_ρ *involving the same scalars* $(\mu_1, \ldots, \mu_{n_\rho})$ *for all* $h = 1, \ldots, n_\rho$*.*

Proof. The (i, h) entry of the matrix $P^* M_\rho(s)$ is

$$\sum_{k=1}^{n_\rho} \overline{p_{ki}}\, m_{kh}(s),$$

and $q_{ij}(s)$ (the (i, j) entry in $Q_\rho(s) = P^* M_\rho(s) P$) is given by

$$q_{ij}(s) = \sum_{h,k=1}^{n_\rho} \overline{p_{ki}}\, p_{hj}\, m_{kh}(s). \tag{q_{ij}}$$

The $q_{ij}(s)$ are indeed linear combinations of the $m_{ij}(s)$. Let us compute the inner product

$$\langle q_{i_1 j_1},\, q_{i_2 j_2}\rangle = \int_G q_{i_1 j_1}(s) \overline{q_{i_2 j_2}(s)}\, d\lambda(s).$$

We have

$$q_{i_1 j_1} \overline{q_{i_2 j_2}} = \sum_{h,k=1}^{n_\rho} \sum_{h',k'=1}^{n_\rho} \overline{p_{k i_1}}\, p_{h j_1} m_{kh}(s) p_{k' i_2} \overline{p_{h' j_2}}\, \overline{m_{k'h'}(s)},$$

and so

$$\begin{aligned}
\langle q_{i_1 j_1},\, q_{i_2 j_2}\rangle &= \int_G q_{i_1 j_1}(s) \overline{q_{i_2 j_2}(s)}\, d\lambda(s) \\
&= \sum_{h,k=1}^{n_\rho} \sum_{h',k'=1}^{n_\rho} \overline{p_{k i_1}}\, p_{h j_1} p_{k' i_2} \overline{p_{h' j_2}} \int_G m_{kh}(s) \overline{m_{k'h'}(s)}\, d\lambda(s) \\
&= \sum_{h,k=1}^{n_\rho} \sum_{h',k'=1}^{n_\rho} p_{k' i_2} \overline{p_{k i_1}}\, p_{h j_1} \overline{p_{h' j_2}}\, \langle m_{kh},\, m_{k'h'}\rangle.
\end{aligned}$$

Since the m_{ij} form an orthogonal family and $\langle m_{kh},\, m_{kh}\rangle = n_\rho$, we obtain

$$\langle q_{i_1j_1},\, q_{i_2j_2}\rangle = \sum_{h,k=1}^{n_\rho} p_{ki_2}\overline{p_{ki_1}}\, p_{hj_1}\overline{p_{hj_2}}\, n_\rho = n_\rho \sum_{k=1}^{n_\rho} p_{ki_2}\overline{p_{ki_1}} \sum_{h=1}^{n_\rho} p_{hj_1}\overline{p_{hj_2}}.$$

If $i_1 \neq i_2$, since P is a unitary matrix the columns of index i_1 and i_2 are orthogonal and so $\sum_{k=1}^{n_\rho} p_{ki_2}\overline{p_{ki_1}} = 0$, and similarly, if $j_1 \neq j_2$, the columns of index j_1 and j_2 are orthogonal and $\sum_{h=1}^{n_\rho} p_{hj_1}\overline{p_{hj_2}} = 0$. Thus $q_{i_1j_1}$ and $q_{i_2j_2}$ are orthogonal if $(i_1, j_1) \neq (i_2, j_2)$.

If $i_1 = i_2$ and $j_1 = j_2$, since P is unitary, its columns are unit vectors, so $\sum_{k=1}^{n_\rho} p_{ki_1}\overline{p_{ki_1}} = 1$ and $\sum_{h=1}^{n_\rho} p_{hj_1}\overline{p_{hj_1}} = 1$, and thus

$$\langle q_{i_1j_1},\, q_{i_1j_1}\rangle = n_\rho.$$

This concludes the proof of (1) and part of (3).

For $s = e$, since $m_{kh}(e) = n_\rho\, \delta_{kh}$, we have

$$q_{ij}(e) = \sum_{h,k=1}^{n_\rho} \overline{p_{ki}}\, p_{hj}\, m_{kh}(e) = n_\rho \sum_{h,k=1}^{n_\rho} p_{kj}\overline{p_{ki}}\, \delta_{kh} = n_\rho \sum_{k=1}^{n_\rho} p_{kj}\overline{p_{ki}}.$$

Since P is a unitary matrix, if $i \neq j$, then $\sum_{k=1}^{n_\rho} p_{kj}\overline{p_{ki}} = 0$, and if $i = j$, then $\sum_{k=1}^{n_\rho} p_{ki}\overline{p_{ki}} = 1$, so we have

$$q_{ij}(e) = n_\rho \delta_{ij}.$$

This finishes the proof of (3).

Since $m_{kh}(s) = \overline{m_{hk}(s^{-1})}$, using (q_{ij}) twice, we have

$$\begin{aligned} q_{ji}(s) &= \sum_{h,k=1}^{n_\rho} \overline{p_{kj}}\, p_{hi}\, m_{kh}(s) \\ &= \sum_{h,k=1}^{n_\rho} \overline{\overline{p_{hi}}\, p_{kj}}\; \overline{m_{hk}(s^{-1})} \\ &= \overline{q_{ij}(s^{-1})}. \end{aligned}$$

Since $m_{kh} * m_{k'h'} = \delta_{hk'} m_{kh'}$, we have

$$\begin{aligned} q_{i_1j_1} * q_{i_2j_2} &= \sum_{h,k=1}^{n_\rho} \sum_{h',k'=1}^{n_\rho} \overline{p_{ki_1}}\, p_{hj_1} \overline{p_{k'i_2}}\, p_{h'j_2}\, m_{kh} * m_{k'h'} \\ &= \sum_{h,k=1}^{n_\rho} \sum_{h',k'=1}^{n_\rho} \overline{p_{ki_1}}\, p_{hj_1} \overline{p_{k'i_2}}\, p_{h'j_2}\, \delta_{hk'} m_{kh'} \\ &= \sum_{h=1}^{n_\rho} p_{hj_1}\overline{p_{hi_2}} \sum_{k,h'=1}^{n_\rho} \overline{p_{ki_1}}\, p_{h'j_2} m_{kh'}. \end{aligned}$$

If $j_1 \neq i_2$, since P is unitary we have $\sum_{h=1}^{n_\rho} p_{hj_1}\overline{p_{hi_2}} = 0$ and then $q_{i_1j_1} * q_{i_2j_2} = 0$. If $j_1 = i_2$, since P is unitary we have $\sum_{h=1}^{n_\rho} p_{hj_1}\overline{p_{hj_1}} = 1$, in which case, using (q_{ij}),

$$q_{i_1j_1} * q_{j_1j_2} = \sum_{k,h'=1}^{n_\rho} \overline{p_{ki_1}}\, p_{h'j_2}\, m_{kh'} = q_{i_1j_2},$$

which concludes the proof of (2).

Since $Q_\rho(s) = P^* M_\rho(s) P$, Part (4) is trivial.

Since the (i,j) entry q_{ij} of $Q_\rho = P^* M_\rho P$ is given by

$$q_{ij}(s) = \sum_{h,k=1}^{n_\rho} \overline{p_{ki}}\, p_{hj}\, m_{kh}(s),$$

any linear combination of the entries of the jth column of Q_ρ is of the form

$$\begin{aligned}\sum_{i=1}^{n_\rho} \lambda_i \sum_{h,k=1}^{n_\rho} \overline{p_{ki}}\, p_{hj}\, m_{kh}(s) &= \sum_{h=1}^{n_\rho} p_{hj} \left(\sum_{k=1}^{n_\rho} \sum_{i=1}^{n_\rho} \overline{p_{ki}}\, \lambda_i m_{kh} \right) \\ &= \sum_{h=1}^{n_\rho} p_{hj} \left(\sum_{k=1}^{n_\rho} (P^*\lambda)_k m_{kh} \right)\end{aligned}$$

with $\lambda = (\lambda_1, \ldots, \lambda_{n_\rho}) \in \mathbb{C}^{n_\rho}$ and where $(P^*\lambda)_k$ is the kth component of the vector $P^*\lambda$, and since P^* is invertible, we deduce that the subspace $\mathfrak{l}_j^Q$ of all linear combinations of entries in the jth column of Q_ρ is given by

$$\mathfrak{l}_j^Q = \left\{ \sum_{h=1}^{n_\rho} p_{hj} \left(\sum_{k=1}^{n_\rho} \mu_k m_{kh} \right) \;\middle|\; \mu = (\mu_1, \ldots, \mu_{n_\rho}) \in \mathbb{C}^{n_\rho} \right\}.$$

To check that $\mathfrak{l}_j^Q$ is a left ideal we need to check that $m_{ij'} * \mathfrak{l}_j^Q \subseteq \mathfrak{l}_j^Q$ for all $m_{ij'} \in \mathfrak{a}_\rho$. Since $m_{ij'} * m_{kh} = \delta_{j'k} m_{ih}$, we have

$$\begin{aligned}m_{ij'} * \mathfrak{l}_j^Q &= \left\{ \sum_{h=1}^{n_\rho} p_{hj} \left(\sum_{k=1}^{n_\rho} \mu_k m_{ij'} * m_{kh} \right) \;\middle|\; \mu = (\mu_1, \ldots, \mu_{n_\rho}) \in \mathbb{C}^{n_\rho} \right\} \\ &= \left\{ \sum_{h=1}^{n_\rho} p_{hj} \left(\sum_{k=1}^{n_\rho} \mu_k \delta_{j'k} m_{ih} \right) \;\middle|\; \mu = (\mu_1, \ldots, \mu_{n_\rho}) \in \mathbb{C}^{n_\rho} \right\} \\ &= \left\{ \sum_{h=1}^{n_\rho} p_{hj} (\mu_{j'} m_{ih}) \;\middle|\; \mu_{j'} \in \mathbb{C} \right\} \subseteq \mathfrak{l}_j^Q.\end{aligned}$$

Thus $\mathfrak{l}_j^Q$ is indeed a left ideal. This left ideal has dimension n_ρ since $(q_{1j}, \ldots, q_{n_\rho j})$ is a basis for it, so it is a minimal ideal because all minimal ideals of $\mathfrak{a}_\rho$ are isomorphic to $\mathfrak{l}_1$, which has dimension n_ρ. ∎

In summary the functions q_{ij} in the matrix $Q_\rho = P^* M_\rho P$ provide another isomorphism of the minimal two-sided ideal $\mathfrak{a}_\rho$ of $\mathrm{L}^2(G)$ with the matrix algebra $\mathrm{M}_{n_\rho}(\mathbb{C})$ and the family of functions

$$\left(\frac{1}{\sqrt{n_\rho}} q_{ij}\right)_{1 \leq i,j \leq n_\rho}$$

is an orthonormal basis of $\mathfrak{a}_\rho$.

13.2 Characters of Compact Groups

Besides characters of groups and characters of algebras, there is one more kind of characters, namely, characters of finite-dimensional representations. As in the previous section, we work with metrizable compact groups. Since we have the Peter–Weyl theorem and Theorem 13.6 at our disposal, it will be fairly easy to prove the properties of characters of these groups.

Definition 13.4. Let G be a metrizable compact group. With the notations of Section 13.2, for every $\rho \in R$, define the *character* χ_ρ *of* G *associated with the ideal* $\mathfrak{a}_\rho$ as the function given by

$$\chi_\rho(s) = \frac{1}{n_\rho}\, u_\rho(s) = \frac{1}{n_\rho} \sum_{j=1}^{n_\rho} m_{jj}^{(\rho)}(s) = \mathrm{tr}(M_\rho(s)), \qquad \text{for all } s \in G.$$

The character χ_{ρ_0} associated with $\mathfrak{a}_{\rho_0}$ is the constant function $\chi_{\rho_0}(s) = 1$ for all $s \in G$, called the *trivial character* of G.

The properties stated in the following proposition are immediate consequences of Theorem 13.2 and Theorem 13.6.

Proposition 13.10. *The following properties hold.*

(1) Every character χ_ρ is a continuous central function, which means that

$$\chi_\rho(sts^{-1}) = \chi_\rho(s) \qquad \textit{for all } s,t \in G.$$

(2) We have

$$\chi_\rho(s^{-1}) = \overline{\chi_\rho(s)} \qquad \textit{for all } s \in G.$$

(3) We have

$$\chi_\rho * \chi_{\rho'} = 0 \quad \textit{whenever } \rho \neq \rho', \quad \textit{and} \quad \chi_\rho * \chi_\rho = \frac{1}{n_\rho}\, \chi_\rho.$$

(4) The family of characters $(\chi_\rho)_{\rho \in R}$ forms a Hilbert basis of the center of $\mathrm{L}^2(G)$, which means that:

(a) *We have*

$$\langle \chi_\rho, \chi_{\rho'} \rangle = \int \chi_\rho(s)\overline{\chi_{\rho'}(s)}\, d\lambda(s) = 0 \qquad\qquad \text{whenever } \rho \neq \rho'$$
$$\langle \chi_\rho, \chi_\rho \rangle = \int |\chi_\rho(s)|^2\, d\lambda(s) = 1.$$

(b) *For every central function* $f \in \mathrm{L}^2(G)$,

$$f = \sum_{\rho \in R} \langle f, \chi_\rho \rangle \chi_\rho = \sum_{\rho \in R} n_\rho\, (f * \chi_\rho).$$

(c) *We have*

$$\int \chi_\rho(s)\, d\lambda(s) = 0 \qquad \text{for all } \rho \neq \rho_0.$$

(d) *For all* $s \in G$, *we have*

$$\chi_\rho(s) = \mathrm{tr}(M_\rho(s)),$$

and

$$\chi_\rho(e) = n_\rho.$$

The only nontrivial proof is the proof of Property (b). We use Property (3) to write

$$\chi_\rho = n_{\rho'} \sum_{\rho' \in R} \chi_\rho * \chi_{\rho'},$$

so that

$$\begin{aligned} f = \sum_{\rho \in R} \langle f, \chi_\rho \rangle \chi_\rho &= \sum_{\rho \in R} \sum_{\rho' \in R} n_{\rho'} \langle f, \chi_\rho \rangle \chi_\rho * \chi_{\rho'} \\ &= \sum_{\rho' \in R} n_{\rho'} \sum_{\rho \in R} \langle f, \chi_\rho \rangle \chi_\rho * \chi_{\rho'} \\ &= \sum_{\rho' \in R} n_{\rho'} \left(\sum_{\rho \in R} \langle f, \chi_\rho \rangle \chi_\rho \right) * \chi_{\rho'} \\ &= \sum_{\rho' \in R} n_{\rho'} (f * \chi_{\rho'}). \end{aligned}$$

Observe that unlike the characters of a locally compact *abelian* group G, which take their values in $\mathbf{U}(1) \cong \mathbf{T}$, the characters χ_ρ of a compact *not necessarily abelian* group G take their values in $\mathbb{C}$. For instance $\chi_\rho(e) = n_\rho$, and in general, $n_\rho > 1$.

The next proposition is needed to prove Theorem 13.12.

Proposition 13.11. *For any two continuous central functions f, g in $\mathcal{C}(G; \mathbb{C})$, we have*

$$f * g = \sum_{\rho \in R} \langle g, \chi_\rho \rangle (f * \chi_\rho),$$

where the family on the right-hand side converges in the topology of uniform convergence.

Proof. This follows from the fact shown in Proposition 13.10(4) that the family of characters $(\chi_\rho)_{\rho \in R}$ is a Hilbert basis of the center of $\mathrm{L}^2(G)$, and the fact that $\|f * g\|_\infty \leq \|f\|_2 \|g\|_2$. □

The next theorem will require an auxiliary proposition.

Theorem 13.12. *The family of continuous central functions $(\chi_\rho)_{\rho \in R}$ constitutes an orthonormal system which is dense in the space of continuous central functions in $\mathcal{C}(G; \mathbb{C})$ for the topology of uniform convergence.*

Proof. For every continuous central function f, since $\chi_\rho = \frac{1}{n_\rho} u_\rho$, by Proposition 13.7, the function $f * \chi_\rho$ is a scalar multiple of χ_ρ. In view of the formula of Proposition 13.11, it suffices to show that for every continuous central function g, there exists a continuous central function f such that $\|f * g - g\|_\infty$ is arbitrarily small. Recall that for any element $t \in G$ the *inner automorphism* C_t is defined by $C_t(s) = tst^{-1}$, for all $s \in G$. The following result is needed.

Proposition 13.13. *The following properties hold.*

(1) *Let G be a metrizable topological group, and let K be a compact subset of G. For every neighborhood U of the identity element e of G, there is a neighborhood $V \subseteq U$ of e such that, $tVt^{-1} \subseteq U$ for all $t \in K$.*

(2) *In a metrizable compact group G, there exists a neighborhood base of neighborhoods of e invariant under all inner automorphisms of G. For such a neighborhood T, there exists a continuous central function $h \geq 0$, of support contained in T, such that $\int_G h(s)\, d\lambda(s) = 1$.*

Proof. (1) Using the technique used to prove Proposition 8.2 applied to the continuous map $(g_1, g_2, g_3) \mapsto g_1 g_2 g_3$, we can find a neighborhood U_0 of e such that $U_0^3 \subseteq U$. Using this technique again, for every $s \in G$, by continuity of the map $g \mapsto sgs^{-1}$, there is a neighborhood V_s of e in G such that $V_s^{-1} = V_s$ and $sV_s s^{-1} \subseteq U_0$. Note that

$$sV_s^3 s^{-1} = sV_s s^{-1} sV_s s^{-1} sV_s s^{-1} \subseteq U_0^3 \subseteq U.$$

Since

$$tV_s t^{-1} = ss^{-1} tV_s t^{-1} ss^{-1},$$

for any $t \in G$, if $s^{-1}t \in V_s$, which implies that $t^{-1}s \in V_s^{-1} = V_s$, then $s^{-1}tV_s t^{-1}s \in V_s^3$, so

$$tV_s t^{-1} = ss^{-1} tV_s t^{-1} ss^{-1} \in sV_s^3 s^{-1} \subseteq U.$$

Therefore if we let $W_s = sV_s$, then for all $t \in W_s = sV_s$, we have $s^{-1}t \in V_s$ and so $tV_st^{-1} \subseteq U$. Since K is compact, there exists a finite number of m elements $s_j \in K$ such that $W_{s_1} \cup \cdots \cup W_{s_m}$ covers K. If we let $V = \bigcap_{j=1}^m V_{s_j}$, we have $tVt^{-1} \subseteq U$ for all $t \in K$.

(2) Apply (1) with $K = G$. Then

$$T = \bigcup_{t \in G} tVt^{-1}$$

is a neighborhood of e contained in U, obviously invariant under the inner automorphisms of G. To define h, start with a continuous function $f \geq 0$ of support contained in T, such that $f(e) > 0$. Let

$$h(t) = c \int_G f(sts^{-1})\, d\lambda(s),$$

where the constant $c > 0$ is chosen in a suitable way, and then the proof that h works is the same as in the proof of Theorem 13.2. This finishes the proof of Proposition 13.13. $\square$

In view of Proposition 13.13(2), Proposition 8.50(i) shows that for every continuous central function g, there is some continuous central function h such that $\|g * h - g\|_\infty$ can be made arbitrarily small, which finishes the proof of Theorem 13.12. $\square$

The following result shows certain independence results.

Theorem 13.14. *The following properties hold.*

(1) For every $s \in G$, if $s \neq e$, then there is some $\rho \in R$ such that $\chi_\rho(s) \neq \chi_\rho(e)$.

(2) We have

$$\bigcap_{\rho \in R} N_\rho = \{e\},$$

where N_ρ is the kernel of the (group) homomorphism $s \mapsto M_\rho(s)$.

Proof. (1) If there is some $s \neq e$ such that $\chi_\rho(s) = \chi_\rho(e)$ for all $\rho \in R$, then by Theorem 13.12 we would have $f(s) = f(e)$ for all continuous central functions, but this contradicts Proposition 13.13(2), since we can find a continuous central function h with $h(e) \neq 0$ whose support T does not contain $s \neq 0$.

(2) If $s \in N_\rho$, then $M_\rho(s) = I_{n_\rho}$, and since $\chi_\rho(s) = \mathrm{tr}(M_\rho(s)) = \mathrm{tr}(I_{n_\rho}) = n_\rho$, by Proposition 13.10(4)(d) we have $\chi_\rho(s) = \chi_\rho(e) = n_\rho$ and (1) implies that

$$\bigcap_{\rho \in R} N_\rho = \{e\},$$

as claimed. $\square$

The following is a product formula for the characters.

Proposition 13.15. *For every character χ of G we have*

$$\chi(s)\chi(t) = \chi(e)\int_G \chi(usu^{-1}t)\,d\lambda(u).$$

Proof. By definition and since $M_\rho(st) = M_\rho(s)M_\rho(t)$ for all $s, t \in G$, we have

$$\chi_\rho(usu^{-1}t) = \frac{1}{n_\rho}\sum_i m_{ii}^{(\rho)}(usu^{-1}t) = \frac{1}{n_\rho^4}\sum_{i,j,h,k} m_{ij}(u)m_{jh}(s)m_{hk}(u^{-1})m_{ki}(t).$$

Using Theorem 13.6 (2) and (3) (convolution evaluated at e) , it follows that

$$\begin{aligned}\int_G \chi_\rho(usu^{-1}t)\,d\lambda(u) &= \frac{1}{n_\rho^4}\sum_{i,j,h,k} m_{jh}(s)m_{ki}(t)\int m_{ij}(u)m_{hk}(u^{-1})\,d\lambda(u)\\ &= \frac{1}{n_\rho^3}\sum_{i,j,h,k}\delta_{jh}\delta_{ik}m_{jh}(s)m_{ki}(t)\\ &= \frac{1}{n_\rho^3}\sum_{i,j} m_{jj}(s)m_{ii}(t)\\ &= \frac{1}{n_\rho}\,\chi_\rho(s)\chi_\rho(t),\end{aligned}$$

as claimed. ☐

Since by Proposition 8.47, $\overline{f * g} = \overline{f} * \overline{g}$, the function which maps the class of a function $f \in \mathcal{L}^2(G)$ to the class of its complex conjugate $\overline{f}$ is a semilinear bijection of $\mathrm{L}^2(G)$, and an automorphism of its ring structure (under convolution).

Definition 13.5. The above automorphism of $\mathrm{L}^2(G)$ maps every ideal $\mathfrak{a}_\rho$ into the minimal two-sided ideal $\overline{\mathfrak{a}_\rho} = \{\overline{f} \mid f \in \mathfrak{a}_\rho\}$ that we denote by $\mathfrak{a}_{\overline{\rho}}$.

The map $\mathfrak{a}_\rho \mapsto \mathfrak{a}_{\overline{\rho}}$ permutes the indices of R but leaves the Hilbert sum unchanged, namely $\mathrm{L}^2(G)$ is the Hilbert sum of both families $(\mathfrak{a}_\rho)_{\rho\in R}$ and $(\mathfrak{a}_{\overline{\rho}})_{\rho\in R} = (\overline{\mathfrak{a}_\rho})_{\rho\in R}$.

If (as usual), given a complex matrix $X = (x_{ij})$, we denote by $\overline{X}$ the matrix $(\overline{x}_{ij})$, then we have

$$\overline{M_\rho(s)} = M_{\overline{\rho}}(s) \qquad \text{for all } s \in G,$$

and as a consequence, since $u_\rho(s) = n_\rho \mathrm{tr}(M_\rho(s)) = n_\rho\chi_\rho(s)$, we have

$$\overline{u_\rho} = u_{\overline{\rho}} \quad \text{and} \quad \overline{\chi_\rho} = \chi_{\overline{\rho}}.$$

Thus the equation $\mathfrak{a}_\rho = \mathfrak{a}_{\overline{\rho}}$ is equivalent to saying that the character χ_ρ only takes *real values*.

Let us now consider the special cases where either G is compact and abelian or G is finite.

Example 13.1. We will consider the case where G is a compact (metrizable) *abelian* group, but before doing this, let G be a not necessarily abelian compact group. Let $f \in \mathcal{L}^2(G)$ be a function not zero everywhere such that for every $s \in G$,

$$f(st) = f(s)f(t) \qquad \text{for almost all } t \in G.$$

Then

$$f(s^{-1}t) = f(s^{-1})f(t),$$

so the subgroup $\mathbb{C}f$ of $\mathrm{L}^2(G)$ is invariant under the map $g \mapsto \lambda_s(g)$ for every $s \in G$, and by Proposition 13.4, this subgroup is a closed minimal left ideal of dimension 1. This is only possible if this left ideal is one of the $\mathfrak{a}_\rho$ for which $n_\rho = 1$, and then f is equal to the character χ_ρ almost everywhere. Such characters are called *abelian characters*. The above reasoning shows that these are the only continuous homomorphisms of G into $\mathbb{C}^*$. Since the image of G by such a character χ_ρ is a compact subgroup of $\mathbb{C}^*$, it must be contained in $\mathbf{U}(1)$.

If G is compact *and abelian*, then every character is obviously abelian, since the algebras $\mathfrak{a}_\rho$ are commutative. Then by Theorem 13.6(3), the characters of G form a Hilbert basis of $\mathrm{L}^2(G)$, and every continuous function is a uniform limit of linear combinations of characters (by Theorem 13.12). We have determined the characters of several compact abelian groups, such as $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{T}^n$, in Proposition 10.9 and Corollary 10.11.

Example 13.2. Let G be a *finite* (not necessarily abelian) group of order $|G| = g$. In this case, the algebras $\mathrm{L}^1(G)$ and $\mathrm{L}^2(G)$ are the same and equal to the group algebra $\mathbb{C}[G] = [G \to \mathbb{C}]$ of formal linear combinations $\sum_{s\in G} x_s\, s$ with $x_s \in \mathbb{C}$, with the convolution

$$\frac{1}{g}\left(\sum_{s_1\in G} x_{s_1}\, s_1\right) * \left(\sum_{s_2\in G} y_{s_2}\, s_2\right) = \frac{1}{g}\sum_{s\in G}\left(\sum_{s_1 s_2 = s} x_{s_1} y_{s_2}\right) s = \frac{1}{g}\sum_{s\in G}\left(\sum_{t\in G} x_t y_{t^{-1}s}\right) s.$$

Recall that two elements $a, b \in G$ are *conjugate* if $b = sas^{-1}$ for some $s \in G$. Conjugation is an equivalence relation in G, and its classes, called the *conjugacy classes* of G, are the sets

$$C_a = \{sas^{-1} \mid s \in G\}.$$

Since G is finite, it has finite number r of conjugacy classes $C_1, \ldots, C_r$, and we assume that $C_1 = \{e\}$.

The central functions (also called *class functions*) are constant on the conjugacy classes. Also, since G is finite, by Theorem 13.12, every central function is a linear combination of characters, and since they are linearly independent, the number r of conjugacy classes is equal to the number of characters, and to the dimension of the center of the algebra $Z(\mathbb{C}[G])$ of $\mathbb{C}[G]$.

Let $R = \{\rho_1, \ldots, \rho_r\}$, and write χ_{ij} for the value of the character χ_i on the conjugacy class C_j $(1 \le i, j \le r)$. If g is the order of the group G and h_j is the number of elements in

the conjugacy class C_j, then the orthogonality relations in Proposition 13.10(4) become

$$\frac{1}{g}\sum_{k=1}^{r} h_k \chi_{ik}\overline{\chi_{jk}} = \delta_{ij} \qquad (1 \leq i, j \leq r).$$

In other words, the matrix

$$\left(\frac{h_k}{g}\,\chi_{ik}\right)_{1\leq i,k\leq r}$$

is unitary. We get more identities by expressing that the transpose of the above matrix is unitary, namely:

$$\sum_{i=1}^{r} \chi_{ik}\overline{\chi_{il}} = 0 \qquad \text{if } k \neq l$$

and

$$\sum_{i=1}^{r} |\chi_{ik}|^2 = \frac{g}{h_k}.$$

These formulae can also be written as

$$\sum_{\rho\in R} \chi_\rho(s)\chi_\rho(t^{-1}) = 0 \tag{$\dagger_1$}$$

if s and t are not conjugate in G, and

$$\sum_{\rho\in R} |\chi_\rho(s)|^2 = \frac{g}{h_k}, \qquad \text{if } s \in C_k. \tag{$\dagger_2$}$$

Since e is not conjugate to any other element in G, if we let $t = e$ in ($\dagger_1$), using the fact that $\chi_\rho(e) = n_\rho$, we obtain

$$\sum_{\rho\in R} n_\rho \chi_\rho(s) = 0 \qquad \text{if } s \neq e.$$

If we let $s = e$ in ($\dagger_2$), we obtain the relation

$$\sum_{\rho\in R} n_\rho^2 = g. \tag{$\dagger_3$}$$

The above equation confirms that $\mathrm{L}^2(G)$ is the direct sum of the $\mathfrak{a}_\rho$.

13.3 The Peter–Weyl Theorem, II

In this section we prove the second part of the Peter–Weyl theorem which has to do with unitary representations. In particular, we prove the important result that the unitary representations $s \mapsto M_\rho(s)$ of G discussed in Theorem 13.6 are irreducible (in fact, all of them, up to equivalence).

Given a unitary representation $V\colon G \to \mathbf{U}(H)$, recall from Definition 12.10 specialized to measures of the form $f\,d\lambda$, where $f \in \mathrm{L}^1(G)$, that for every $x \in H$, the unique vector $\widetilde{V}(f\,d\lambda)(x) \in H$ such that

$$\langle \widetilde{V}(f\,d\lambda)(x), y\rangle = \int \langle V(s)(x), y\rangle f(s)\,d\lambda(s) \qquad \text{for all } y \in H$$

is called the *weak integral* of the function $s \mapsto V(s)(x)$ from G to H with respect to $fd\lambda$, and is denoted by

$$\int_G V(s)(x) f(x)\,d\lambda(s).$$

We write $V_{\mathrm{ext}}(f)$ instead of $\widetilde{V}(f\,d\lambda)$. We know from by Proposition 12.13 that $\|V_{\mathrm{ext}}(f)\| \leq \|f\|_1$. Recall fom Definition 13.4 that $u_\rho(s) = n_\rho \chi_\rho(s)$.

Theorem 13.16. *(Peter–Weyl theorem, II) Let G be a metrizable compact group, and let $V\colon G \to \mathbf{U}(H)$ be a unitary representation of G in a separable Hilbert space H.*

(1) For every $\rho \in R$, the map $V_{\mathrm{ext}}(\overline{u_\rho})$ given by

$$V_{\mathrm{ext}}(\overline{u_\rho})(x) = \int_G \overline{u_\rho}(s) V(s)(x)\,d\lambda(s) = n_\rho \int_G \overline{\chi_\rho}(s) V(s)(x)\,d\lambda(s) \tag{proj}$$

is an orthogonal projection of H onto a closed subspace E_ρ (which may be reduced to (0)), and H is the Hilbert sum of the $E_\rho \neq (0)$.

(2) Every subspace $E_\rho \neq (0)$ is invariant under V, and the restriction V_ρ of V to E_ρ is a finite or countably infinite Hilbert sum of irreducible representations, all equivalent to M_ρ, viewed as a representation $M_\rho\colon G \to \mathbf{U}(\mathbb{C}^{n_\rho})$.

Proof. To help the reader navigate through the flow of this proof we provide the following proof outline. By Theorem 13.2 we have the Hilbert sum

$$\mathrm{L}^2(G) = \bigoplus_{\rho \in R} \mathfrak{a}_\rho = \bigoplus_{\rho \in R} \mathfrak{a}_{\overline{\rho}},$$

with

$$\mathfrak{a}_\rho = \mathfrak{l}_1 \oplus \cdots \oplus \mathfrak{l}_{n_\rho}.$$

We are given a unitary representation $V\colon G \to \mathbf{U}(H)$ of G and we use Theorem 12.14 to form the algebra representation $V_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ of $\mathrm{L}^1(G)$. For Part 1, we define

$$E_\rho = \{V_{\mathrm{ext}}(\overline{u_\rho})(x) \mid x \in H\},$$

and we show that H is the Hilbert sum

$$H = \oplus_{\rho \in R} E_\rho.$$

In Part (2), Step 1, we prove that E_ρ is invariant under V. For Step 2 we consider the restriction $V_\rho\colon G \to \mathbf{U}(E_\rho)$ of V to E_ρ and its extension $(V_\rho)_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(E_\rho)$ to $\mathrm{L}^1(G)$. Since G is compact, $L^2(G) \subseteq L^1(G)$, and we can show that $(V_\rho)_{\mathrm{ext}}$ is zero on every $\mathfrak{a}_{\overline{\rho'}}$ with $\rho' \neq \rho$. Consequently the restriction of $(V_\rho)_{\mathrm{ext}}$ to $\mathrm{L}^2(G)$ can be viewed as a nondegenerate representation

$$(V_\rho)_{\mathrm{ext}}\colon \mathfrak{a}_{\overline{\rho}} \to \mathcal{L}(E_\rho)$$

of the topologically simple algebra $\mathfrak{a}_{\overline{\rho}}$ in E_ρ. This allows us to use Theorem 11.34(2) to obtain a finite or countably infinite Hilbert sum

$$E_\rho = \bigoplus_k E_\rho^{(k)}$$

such that every representation $(V_\rho)_{\mathrm{ext}}^{(k)}\colon \mathfrak{a}_{\overline{\rho}} \to \mathcal{L}(E_\rho^{(k)})$ is equivalent to the irreducible representation $U_{\overline{\mathfrak{l}_1}}\colon \mathfrak{a}_{\overline{\rho}} \to \mathcal{L}(\overline{\mathfrak{l}_1})$.

In Step 3 we observe that $U_{\overline{\mathfrak{l}_1}}$ is the restriction of $\mathbf{R}_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(\mathrm{L}^2(G))$ to $\mathfrak{a}_{\overline{\rho}}$, where $\mathbf{R}\colon G \to \mathbf{U}(\mathrm{L}^2(G))$ is the left regular representation of G in $\mathrm{L}^2(G)$. Thus we can view $U_{\overline{\mathfrak{l}_1}}$ as the nondegenerate topologically irreducible representation $\widetilde{U_{\overline{\mathfrak{l}_1}}}\colon \mathrm{L}^1(G) \to \mathcal{L}(\overline{\mathfrak{l}_1})$ obtained by extending the nondegenerate representation $U_{\overline{\mathfrak{l}_1}}\colon \mathfrak{a}_{\overline{\rho}} \to \mathcal{L}(\overline{\mathfrak{l}_1})$ to $\mathrm{L}^1(G)$ (we set $\widetilde{U_{\overline{\mathfrak{l}_1}}}$ to zero on the orthogonal complement of $\mathfrak{a}_{\overline{\rho}}$) which is equal to $\mathbf{R}_{\mathrm{ext}}$ on $\mathfrak{a}_{\overline{\rho}}$. The corresponding representation of G is an irreducible unitary representation of G in $\overline{\mathfrak{l}_1}$ that agrees with $\mathbf{R}$, so we compute the matrix of $\mathbf{R}(s)$ in the basis of $\overline{\mathfrak{l}_1}$ consisting of the vectors $(\frac{1}{n_\rho}\overline{m_{i1}})_{1\le i\le n_\rho}$, and we find M_ρ.

And now comes the detailed proof. (1) By the Peter–Weyl theorem (Theorem 13.2), $\mathrm{L}^2(G)$ is the Hilbert sum of both families $(\mathfrak{a}_\rho)_{\rho\in R}$ and $(\mathfrak{a}_{\overline{\rho}})_{\rho\in R} = (\overline{\mathfrak{a}_\rho})_{\rho\in R}$, but to obtain the matrix representations M_ρ we need to use the Hilbert sum $(\mathfrak{a}_{\overline{\rho}})_{\rho\in R}$. We observed just after Definition 13.5 that

$$\overline{M_\rho(s)} = M_{\overline{\rho}}(s), \quad \overline{u_\rho} = u_{\overline{\rho}} \quad \overline{\chi_\rho} = \chi_{\overline{\rho}}.$$

The first equation implies that if $m_{ij}^{(\rho)}(s)$ are the elements of the matrix $M_\rho(s)$, then the elements $m_{ij}^{(\overline{\rho})}(s)$ of the matrix $M_{\overline{\rho}}(s)$ are given by $m_{ij}^{(\overline{\rho})}(s) = \overline{m_{i,j}^{(\rho)}}$. The $u_{\overline{\rho}}$ are self-adjoint idempotents, that is, $u_{\overline{\rho}} * u_{\overline{\rho}} = u_{\overline{\rho}}$ and $u_{\overline{\rho}}^* = \check{\overline{u}}_{\overline{\rho}} = u_{\overline{\rho}}$, and since $V_{\mathrm{ext}}\colon \mathrm{L}^1(G) \to \mathcal{L}(H)$ is an algebra homomorphism, we have

$$\begin{aligned} V_{\mathrm{ext}}(u_{\overline{\rho}}) &= V_{\mathrm{ext}}(u_{\overline{\rho}} * u_{\overline{\rho}}) \\ &= V_{\mathrm{ext}}(u_{\overline{\rho}}) \circ V_{\mathrm{ext}}(u_{\overline{\rho}}), \end{aligned}$$

and

$$\begin{aligned} V_{\mathrm{ext}}(u_{\overline{\rho}}) &= V_{\mathrm{ext}}(u_{\overline{\rho}}^*) \\ &= (V_{\mathrm{ext}}(u_{\overline{\rho}}))^*, \end{aligned}$$

so the continuous linear map $V_{\mathrm{ext}}(u_{\overline{\rho}})$ is idempotent and hermitian, and by Proposition 11.5, it is an orthogonal projection. Since $u_{\overline{\rho}} * u_{\overline{\rho'}} = 0$ if $\rho \neq \rho'$, we have $V_{\mathrm{ext}}(u_{\overline{\rho}}) \circ V_{\mathrm{ext}}(u_{\overline{\rho'}}) = 0$, which implies that the images E_ρ of the projections $V_{\mathrm{ext}}(u_{\overline{\rho}})$ are closed subspaces that are pairwise orthogonal. To prove that H is the Hilbert sum of the the family $(E_\rho)_{\rho \in R}$, we need to show that the algebraic direct sum $\bigoplus_{\rho \in R} E_\rho$ is dense in H. We know from the proof of Theorem 12.15 that the linear span E of the set $\{V_{\mathrm{ext}}(f)(x) \mid f \in \mathcal{C}(G; \mathbb{C}), x \in H\}$ is dense in H, and by Proposition 12.13, we have $\|V_{\mathrm{ext}}(f)\| \leq \|f\|_1$. By Proposition 13.8(2), if f is continuous, then for every $\epsilon > 0$, there is a (finite) linear combination $\sum_{i,j,\rho} c_{ij}^{(\overline{\rho})} m_{ij}^{(\overline{\rho})}$ such that

$$\left\| f - \sum_{i,j,\rho} c_{ij}^{(\overline{\rho})} m_{ij}^{(\overline{\rho})} \right\|_\infty \leq \epsilon,$$

and since $\|V_{\mathrm{ext}}(f)\| \leq \|f\|_1$, G is compact, and $\lambda(G) = 1$, we have $\|g\|_1 \leq \|g\|_\infty$ for any $g \in \mathrm{L}^1(G)$, and this implies that

$$\left\| V_{\mathrm{ext}}(f) - \sum_{i,j,\rho} c_{ij}^{(\overline{\rho})} V_{\mathrm{ext}}(m_{ij}^{(\overline{\rho})}) \right\| \leq \left\| f - \sum_{i,j,\rho} c_{ij}^{(\overline{\rho})} m_{ij}^{(\overline{\rho})} \right\|_1 \leq \left\| f - \sum_{i,j,\rho} c_{ij}^{(\overline{\rho})} m_{ij}^{(\overline{\rho})} \right\|_\infty \leq \epsilon.$$

Since $m_{ij}^{(\overline{\rho})} = u_{\overline{\rho}} * m_{ij}^{(\overline{\rho})}$, we have $V_{\mathrm{ext}}(m_{ij}^{(\overline{\rho})}) = V_{\mathrm{ext}}(u_{\overline{\rho}}) \circ V_{\mathrm{ext}}(m_{ij}^{(\overline{\rho})})$, thus the vector $\left(\sum_{i,j,\rho} c_{ij}^{(\overline{\rho})} V_{\mathrm{ext}}(m_{ij}^{(\overline{\rho})}) \right)(x)$ belongs to the sum of the E_ρ, which proves that E is dense in H. We delete the summands E_ρ such that $E_\rho = (0)$.

(2) *Step 1*. We prove that the subspaces E_ρ which are invariant under V_{ext} are also invariant under V. Since by Proposition 13.7, the $u_{\overline{\rho}}$ belong to the center of $\mathrm{L}^2(G)$, by Proposition 13.1, the $u_{\overline{\rho}}$ belong to the center of $\mathcal{M}^1(G)$. Recall that Theorem 12.14 actually yields an algebra representation $\widetilde{V} \colon \mathcal{M}^1(G) \to \mathcal{L}(H)$ of the unital involutive Banach algebra $\mathcal{M}^1(G)$ and that V_{ext} is the restriction of $\widetilde{V}$ to $\mathrm{L}^1(G)$. In particular, even though $\delta_s \notin \mathrm{L}^1(G)$ (unless G is discrete), $\widetilde{V}(\delta_s)$ makes sense, and

$$\widetilde{V}(\delta_s) = V(s),$$

as stated in ($\widetilde{U}(\delta_s)$) just after Definition 12.10. Also recall (†) from the proof of Theorem 12.15,

$$V(s)(V_{\mathrm{ext}}(f)(x)) = V_{\mathrm{ext}}(\delta_s * f)(x), \qquad \text{for all } f \in \mathcal{L}^1(G), \text{ and all } x \in H. \tag{†}$$

Then as any $E_\rho \neq (0)$ is the image of $V_{\mathrm{ext}}(u_{\overline{\rho}})$, we have

$$\begin{aligned} V(s)(V_{\mathrm{ext}}(u_{\overline{\rho}})(x)) &= V_{\mathrm{ext}}(\delta_s * u_{\overline{\rho}})(x) \\ &= \widetilde{V}(\delta_s * u_{\overline{\rho}})(x) \\ &= \widetilde{V}(u_{\overline{\rho}} * \delta_s)(x) \\ &= \widetilde{V}(u_{\overline{\rho}})(\widetilde{V}(\delta_s)(x)) \\ &= V_{\mathrm{ext}}(u_{\overline{\rho}})(V(s)(x)), \end{aligned}$$

we conclude that the E_ρ are invariant under V.

Step 2. We want to prove that if V_ρ is the restriction of the representation V of G to E_ρ we obtain a Hilbert sum decomposition into irreducible representations for the restriction of the nondegenerate representation $(V_\rho)_{\mathrm{ext}}$ to $\mathfrak{a}_{\overline{\rho}}$ in E_ρ.

If V_ρ is the restriction of the representation V of G to E_ρ, then $(V_\rho)_{\mathrm{ext}}(u_{\overline{\rho'}}) = 0$ for all $\rho' \neq \rho$, since $u_{\overline{\rho}} * u_{\overline{\rho'}} = 0$. The representation $(V_\rho)_{\mathrm{ext}}$ is a representation of $\mathrm{L}^1(G)$ in E_ρ, and since $\mathrm{L}^2(G) \subseteq \mathrm{L}^1(G)$ is the Hilbert sum $\mathrm{L}^2(G) = \bigoplus_{\rho\in R} \mathfrak{a}_{\overline{\rho}}$, and the projection of $\mathrm{L}^2(G)$ onto $\mathfrak{a}_{\overline{\rho}}$ is the map $f \mapsto f * u_{\overline{\rho}}$, we have $(V_\rho)_{\mathrm{ext}}(f * u_{\overline{\rho'}}) = (V_\rho)_{\mathrm{ext}}(f) \circ (V_\rho)_{\mathrm{ext}}(u_{\overline{\rho'}}) = 0$, which means that $(V_\rho)_{\mathrm{ext}}$ is zero on every $\mathfrak{a}_{\overline{\rho'}}$ with $\rho' \neq \rho$. Consequently the restriction of $(V_\rho)_{\mathrm{ext}}$ to $\mathrm{L}^2(G)$ can be viewed as a nondegenerate representation of the topologically simple algebra $\mathfrak{a}_{\overline{\rho}}$ in E_ρ. Since by Proposition 12.13, the representation $(V_\rho)_{\mathrm{ext}}$ is continuous, by Theorem 11.34(2), the nondegenerate representation $(V_\rho)_{\mathrm{ext}}$ of $\mathfrak{a}_{\overline{\rho}}$ in E_ρ is a finite or countably infinite (if E_ρ is infinite dimensional) Hilbert sum of topologically irreducible representations all equivalent to the representation $U_{\overline{\mathfrak{l}_1}} \colon \mathfrak{a}_{\overline{\rho}} \to \mathcal{L}(\overline{\mathfrak{l}_1})$.

Step 3. We observe that $U_{\overline{\mathfrak{l}_1}}$ is the restriction of $\mathbf{R}_{\mathrm{ext}} \colon \mathrm{L}^1(G) \to \mathcal{L}(\mathrm{L}^2(G))$ to $\mathfrak{a}_{\overline{\rho}}$, where $\mathbf{R} \colon G \to \mathbf{U}(\mathrm{L}^2(G))$ is the left regular representation of G in $\mathrm{L}^2(G)$ given by

$$(\mathbf{R}(s)(f))(t) = \lambda_s(f)(t) = f(s^{-1}t), \qquad f \in \mathrm{L}^2(G),\ s,t \in G;$$

see Definition 12.11. This is because by definition of $U_{\overline{\mathfrak{l}_1}}$ in Proposition 11.18,

$$U_{\overline{\mathfrak{l}_1}}(f)(g) = f * g, \qquad f \in \mathfrak{a}_{\overline{\rho}},\ g \in \overline{\mathfrak{l}_1},$$

and Definition 12.12 of the left regular representation $\mathbf{R}_{\mathrm{ext}}$ of $\mathrm{L}^1(G)$ in $\mathrm{L}^2(G)$,

$$(\mathbf{R}_{\mathrm{ext}}(f))(g) = f * g, \qquad f \in \mathrm{L}^1(G),\ g \in \mathrm{L}^2(G),$$

so $U_{\overline{\mathfrak{l}_1}}$ is the restriction of $\mathbf{R}_{\mathrm{ext}}$ to $\mathfrak{a}_{\overline{\rho}}$. Thus we can view $U_{\overline{\mathfrak{l}_1}}$ as the nondegenerate topologically irreducible representation $\widetilde{U_{\overline{\mathfrak{l}_1}}} \colon \mathrm{L}^1(G) \to \mathcal{L}(\overline{\mathfrak{l}_1})$ obtained by extending the nondegenerate representation $U_{\overline{\mathfrak{l}_1}} \colon \mathfrak{a}_{\overline{\rho}} \to \mathcal{L}(\overline{\mathfrak{l}_1})$ to $\mathrm{L}^1(G)$ (we set $\widetilde{U_{\overline{\mathfrak{l}_1}}}$ to zero on the orthogonal complement of $\mathfrak{a}_{\overline{\rho}}$) which is equal to $\mathbf{R}_{\mathrm{ext}}$ on $\mathfrak{a}_{\overline{\rho}}$. The corresponding representation of G is an irreducible unitary representation of G in $\overline{\mathfrak{l}_1}$ that agrees with $\mathbf{R}$, so we compute the matrix of $\mathbf{R}(s)$ in the basis of $\overline{\mathfrak{l}_1}$ consisting of the vectors $(\frac{1}{n_\rho}\overline{m_{i1}})_{1\leq i\leq n_\rho}$.

Using Theorem 13.6(4), we have

$$\mathbf{R}(s)(\overline{m_{j1}}) = \overline{m_{j1}}(s^{-1}t) = \frac{1}{n_\rho}\sum_{i=1}^{n_\rho} \overline{m_{ji}}(s^{-1})\overline{m_{i1}}(t)$$

and since by Theorem 13.6(2), $\overline{m_{ji}}(s^{-1}) = m_{ij}(s)$ we recognize that the matrix of $\mathbf{R}(s)$ is $M_\rho(s)$, as claimed. $\square$

The above proof is an adaptation of Dieudonné's proof [21] (Section 4, Theorem 21.4.1). Dieudonné's proof uses the projection $V_{\mathrm{ext}}(u_\rho)$ instead of the projection $V_{\mathrm{ext}}(\overline{u_\rho})$. The second option is the projection used by Serre in his short section on the representation of compact groups and also in Hewitt and Ross; see Serre [90] (Section 4.3) and Hewitt and Ross [48] (Chapter VII, Theorem 27.44). The advantage of Dieudonné's choice is that we avoid a plethora of indices $\overline{\rho}$, but the disadvantage is that the irreducible representations that occur in a given representation are the $\overline{M_\rho} = M_{\overline{\rho}}$. With the second option (as in Serre and Hewitt and Ross), the irreducible representations that occur are the M_ρ; no conjugation needed. Even though using the second option causes an additional notational burden in the proof (most indices are $\overline{\rho}$ instead of ρ), in the long term this simplifies matters because the representations that occur are the M_ρ's.

Let us emphasize that Theorem 13.16 proves that *every* representation M_ρ is irreducible, which is not at all obvious from their definition. Theorem 13.16 also shows that every irreducible unitary representation of G is equivalent to some representation of the form M_ρ, and M_ρ is not equivalent to $M_{\rho'}$ for $\rho \neq \rho'$.

Definition 13.6. Let G be a locally compact (metrizable) group. A sequence of unitary representations $(U_\rho \colon G \to \mathbf{U}(H_\rho))_{\rho \in R}$ of G where R is some index set (possibly infinite) is called a *complete set of irreducible unitary representations of* G if

(1) Each unitary representation $U_\rho \colon G \to \mathbf{U}(H_\rho)$ is irreducible.

(2) Any two representations U_ρ and $U_{\rho'}$ with $\rho \neq \rho'$ are inequivalent.

(3) Every irreducible unitary representation $V \colon G \to \mathbf{U}(H)$ of G is equivalent to some representation U_ρ (necessarily unique).

Consequently $(M_\rho)_{\rho \in R}$ is a complete set of unitary irreducible representations of G in a separable Hilbert space. When we deal with more than one group G (say also a closed subgroup of G) we use the notation $R(G)$ instead of R.

Remark: It would be tempting to say that each ρ corresponds to an equivalence class of unitary representations (under equivalence) but there is a set-theoretic difficulty since the collection of unitary representations is not a set. This sticky point appears to be ignored by most authors, who do not hesitate to refer to the "set of equivalence classes" of irreducible representations of a group G, and even to the set of *all* representations of G. Some authors are more careful and avoid the term "equivalence classes of irreducible representations." The only source we are aware of that brings up this issue is Hewitt and Ross [48] (Chapter VII, second footnote on Page 2). They suggest that a way to circumvent this set-theoretic difficulty is to observe that for a given group G, the cardinality of the vector spaces involved in irreducible representations of G is bounded. In fact, by Proposition 12.1, it is bounded by $\aleph_1^{|G|}$, where $|G|$ denotes the cardinality of G. Then by Riesz–Fischer (Theorem D.19), we can pick representatives for the Hilbert spaces of cardinality at most $\aleph_1^{|G|}$ among $\ell^p(K)$-spaces with K of cardinality bounded by $\aleph_1^{|G|}$. For compact groups, we just showed that the

irreducible unitary representations are finite-dimensional so we can pick these vector spaces as the spaces $\mathbb{C}^n$ (countably many). Hewitt and Ross's footnote ends with the sentence: "The exact details are of little interest for the purposes of the present book." We tend to agree! Definition 13.6 is designed to avoid set-theoretic difficulties. With a small abuse of language, we may still say that the unitary representations equivalent to the representation M_ρ are *of class* ρ.

If the compact group G is abelian, then every algebra $\mathfrak{a}_\rho$ is abelian, and since it is simple, it must be one-dimensional. Therefore, every unitary representation of a metrizable compact abelian group is a finite or a countably infinite Hilbert sum of *one-dimensional* representations.

It is customary to introduce the following terminology.

Definition 13.7. With the notations of Theorem 13.16, if $V\colon G \to \mathbf{U}(H)$ is a unitary representation of G in a separable Hilbert space H, and if $H = \bigoplus_{\rho\in R} E_\rho$ is the Hilbert sum induced by the projections $\pi_\rho^V = V_{\mathrm{ext}}(\overline{u_\rho})$, with

$$\pi_\rho^V(x) = n_\rho \int_G \overline{\chi_\rho}(s)V(s)(x)\, d\lambda(s), \quad x \in H$$

whenever $E_\rho \neq (0)$ and $V_\rho\colon G \to \mathbf{U}(E_\rho)$ is the corresponding representation, we say that the irreducible representation M_ρ is *contained* in the representation V. If E_ρ is finite-dimensional of dimension $d_\rho n_\rho > 0$ we say that M_ρ is *contained* d_ρ *times* in V (or *infinitely many times* if E_ρ is infinite-dimensional). We also call d_ρ the *multiplicity* of M_ρ in V_ρ. The representations M_ρ such that $d_\rho > 0$ are called the *irreducible components* of the representation V.

If we consider the left regular representation $\mathbf{R}$ of G in $\mathrm{L}^2(G)$, then the projection $\pi_\rho^{\mathbf{R}}$ is given by

$$\pi_\rho^{\mathbf{R}}(f) = \int \overline{u_\rho}(s)\mathbf{R}_s(f)\, d\lambda(s) = \int \overline{u_\rho}(s)\lambda_s(f)\, d\lambda(s) = \overline{u_\rho}(s) * f,$$

so $E_\rho = \mathfrak{a}_{\overline{\rho}} = \overline{\mathfrak{a}_\rho}$ for all $\rho \in R$, and Theorem 13.16 says that on $\mathfrak{a}_{\overline{\rho}}$, the representation $\mathbf{R}$ splits into n_ρ irreducible representations all equivalent to M_ρ. We can view these representation as acting on the columns of $M_{\overline{\rho}} = \overline{M_\rho}$, which span n_ρ minimal left ideals $\mathfrak{l}_j^{(\overline{\rho})}$ of $\mathfrak{a}_{\overline{\rho}}$; that is,

$$\mathfrak{a}_{\overline{\rho}} = \bigoplus_{j=1}^{n_\rho} \mathfrak{l}_j^{(\overline{\rho})} \quad \text{and} \quad \mathfrak{l}_j^{(\overline{\rho})} = \bigoplus_{k=1}^{n_{\overline{\rho}}} \mathbb{C} m_{kj}^{(\overline{\rho})}.$$

Remark: The statement $E_\rho = \mathfrak{a}_{\overline{\rho}} = \overline{\mathfrak{a}_\rho}$ may seem wrong, but it is correct. It is a consequence of the definition of the projection π_ρ^V. The exact same fact is noted in Hewitt and Ross [48] (Chapter VII, Section 27.49).

The above fact is worth recording as a proposition.

Proposition 13.17. *The left regular representation* $\mathbf{R}\colon G \to \mathbf{U}(\mathrm{L}^2(G))$ *of a compact (metrizable) group* G *in* $\mathrm{L}^2(G)$ *contains every irreducible unitary representation* M_ρ *of* G*, and each one is contained* n_ρ *times, where* n_ρ *is the dimension of the space of the representation.*

Proposition 13.17 is a generalization to compact groups of a property holding for finite groups for which the proof is much easier; see Serre [90] (Section 2.4).

If V is a finite-dimensional unitary representation, then the trace of the linear map $V(s)$ plays a crucial role. In fact, it determines this representation up to equivalence.

Proposition 13.18. *Let* G *be a metrizable compact group. For any unitary representation* $V\colon G \to \mathbf{U}(H)$ *of* G *in a finite-dimensional hermitian space* H *of dimension* d*, assume that for every* $\rho \in R$*, the irreducible representation* M_ρ *is contained* d_ρ *times in* V*, so that*

$$d = \sum_{\rho \in R} d_\rho n_\rho,$$

where $d_\rho \neq 0$ *for only finitely many* $\rho \in R$*. Then we have*

$$\mathrm{tr}(V(s)) = \sum_{\rho \in R} d_\rho \chi_\rho(s), \qquad \text{for all } s \in G.$$

Proof. We can write H as the direct sum of finite-dimensional spaces, and by picking bases, we can express $V(s)$ as a sum of matrices similar to some of the $M_\rho(s)$. Then the above formula follows from the fact that $\chi_\rho(s) = \mathrm{tr}(M_\rho(s))$ and the fact that the trace is invariant under conjugation, $\mathrm{tr}(PVP^{-1}) = \mathrm{tr}(V)$. ◻

Theorem 13.19. *Let* G *be a metrizable compact group. Two unitary representations* $V_1\colon G \to \mathbf{U}(H_2)$ *and* $V_2\colon G \to \mathbf{U}(H_2)$ *of* G *in finite-dimensional hermitian spaces* H_1 *and* H_2 *of dimensions* d_1 *and* d_2 *are equivalent if and only if*

$$\mathrm{tr}(V_1(s)) = \mathrm{tr}(V_2(s)) \qquad \text{for all } s \in G.$$

In particular, if V_1 *and* V_2 *are equivalent, then* $d_1 = d_2$*. Moreover, if* V_1 *and* V_2 *are any two equivalent irreducible unitary representations, then*

$$\mathrm{tr}(V_1(s)) = \mathrm{tr}(V_2(s)) = \chi_\rho(s), \quad s \in G,$$

where M_ρ *is the irreducible representation from Theorem 13.16 to which* V_2 *and* V_2 *are equivalent.*

Proof. Clearly, if V_1 and V_2 are equivalent, the formula of the theorem holds. Conversely, by Proposition 13.18, since

$$\mathrm{tr}(V_1(s)) = \sum_{\rho \in R_1} d_\rho \chi_\rho(s) \quad \text{and} \quad \mathrm{tr}(V_2(s)) = \sum_{\rho \in R_2} d_\rho \chi_\rho(s)$$

for some finite subsets R_1 and R_2 of R, and since by Theorem 13.6(3) the characters are linearly independent, we must have $R_1 = R_2$ and $d_1 = d_2$. ◻

Theorem 13.19 suggests the following (standard) definition.

Definition 13.8. Let G be a metrizable compact group. For any unitary representation $V\colon G \to \mathbf{U}(H)$ of G in a finite-dimensional hermitian space H of dimension d, we define the *character χ_V of the representation V* as the map $\chi_V\colon G \to \mathbb{C}$ given by

$$\chi_V(s) = \mathrm{tr}(V(s)) \qquad \text{for all } s \in G.$$

The characters of a finite-dimensional unitary representation are central functions, and in view of Proposition 13.18, they have many of the properties of the characters χ_ρ.

By definition, the character χ_ρ of the compact group G is identical to the character χ_{M_ρ} of the special representation M_ρ, which is irreducible by Peter–Weyl II, and by Theorem 13.19, it is also the character of *all* equivalent irreducible unitary representations of G equivalent to M_ρ. Thus the set $(\chi_\rho)_{\rho\in R}$ is the set of characters of *all* irreducible unitary representations of G. If we have some complete set of irreducible unitary representations for G, we can determine the characters of G. If the group G is finite, then there are finitely many irreducible representations up to equivalence, so this method can be used practically.

Example 13.3. Let G be a finite group and assume thatthat $(\rho_1, \ldots, \rho_r)$ is a complete set of irreducible unitary representations $\rho_i\colon G \to \mathbf{U}(W_i)$ of G (where r is the number of conjugacy classes of G) so that $R = \{\rho_1, \ldots, \rho_r\}$, write $n_i = \dim(W_i)$, and let $\chi_1, \ldots, \chi_r$ be the characters of G (which are equal to the characters of the ρ_i). If $U\colon G \to \mathbf{U}(E)$ is any unitary representation of G (where E is finite-dimensional), then by Peter–Weyl II, we have a direct sum

$$E = E_{i_1} \oplus \cdots \oplus E_{i_h} \tag{$\dagger_1$}$$

for some subset $\{\rho_{i_1}, \ldots, \rho_{i_h}\}$ of R ($h \leq r$), and each E_{i_j} ($1 \leq j \leq h$) is a direct sum

$$E_{i_j} = E_1^{(i_j)} \oplus \cdots \oplus E_{d_j}^{(i_j)} \quad (d_j \geq 1) \tag{$\dagger_2$}$$

such that for $k = 1, \ldots, d_j$, each representation $U\colon G \to \mathbf{U}(E_k^{(i_j)})$ is equivalent to the irreducible representation $\rho_{i_j}\colon G \to \mathbf{U}(W_{i_j})$. Each subspace E_{i_j} is the projection of E by the projection $\pi_{i_j}^U$ given by

$$\pi_{i_j}^U(x) = \frac{n_{i_j}}{|G|} \sum_{s\in G} \overline{\chi_{i_j}}(s) U(s)(x) \quad x \in E. \tag{$\dagger_3$}$$

The E_{i_j} in ($\dagger_1$) are uniquely determined by U (in terms of the projections $\pi_{i_j}^U$), but the splitting of E_{i_j} as a direct sum as above in ($\dagger_2$) is not.

The decomposition of $U\colon G \to \mathbf{U}(E)$ into the h unitary representations $U\colon G \to \mathbf{U}(E_{i_j})$ ($1 \leq j \leq h$) is called the *canonical decomposition of U*. For finite groups, these results can be obtained more directly; see Serre [90] (Section 2.6, in particular, Theorem 8). Each representation $U\colon G \to \mathbf{U}(E_{i_j})$ ($1 \leq j \leq h$) contains the irreducible representation $\rho_{i_j}\colon G \to$

$\mathbf{U}(W_{i_j})$ d_j times, so it is not irreducible unless $d_j = 1$. It is actually possible to obtain a specific decomposition of each E_{i_j} into some subspaces $E_k^{(i_j)}$ as in ($\dagger_2$) given by projections expressed in terms of matrix representations for the irreducible representations $\rho_{i_j} \colon G \to \mathbf{U}(W_{i_j})$; See Serre [90] (Section 2.7).

Example 13.4. Recall from Example 12.1 that the group $\mathfrak{S}_3$ consists of the permutations on the set $\{1,2,3\}$. There are $3! = 6$ permutations

$$\sigma_1 = (1,2,3),\ \ \sigma_2 = (1,3,2),\ \ \sigma_3 = (2,1,3),\ \ \sigma_4 = (2,3,1),\ \ \sigma_5 = (3,1,2),\ \ \sigma_6 = (3,2,1),$$

three conjugacy classes, $C_1 = \{\sigma_1\}$, $C_2 = \{\sigma_2, \sigma_3, \sigma_6\}$, $C_3 = \{\sigma_4, \sigma_5\}$, and three irreducible representations (up to equivalence). The two one-dimensional irreducible representations are the trivial representation $\rho_1 \colon \mathfrak{S}_3 \to \mathbf{U}(1)$ with

$$\rho_1(\sigma_i) = 1, \quad i = 1, \ldots, 6,$$

and the signature representation $\rho_2 \colon \mathfrak{S}_3 \to \mathbf{U}(1)$ from Example 12.5, with

$$\rho_2(\sigma_i) = \begin{cases} +1 & \sigma_i \in C_1 \\ -1 & \sigma_i \in C_2 \\ +1 & \sigma_i \in C_3. \end{cases}$$

The third irreducible representation ρ_3 is two-dimensional and is obtained from Example 12.4. We obtained the matrix representation of $\rho_3 \colon \mathfrak{S}_3 \to \mathbf{U}(2)$ by 3×3 matrices with respect to the basis (w_1, w_2, w_3) and we just have two consider the 2×2 matrices obtained by deleting the first row and the first column since w_1 is invariant. We get

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix},$$
$$\begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}, \quad \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}.$$

In Example 12.6 we found the canonical decomposition of the regular representation $\rho_{\mathbf{R}} \colon \mathfrak{S}_3 \to \mathbf{GL}(\mathbb{C}^6)$ of $\mathfrak{S}_3$ from Example 12.2. We have

$$\mathbb{C}^6 = V_1 \oplus V_2 \oplus V_3 = V_1 \oplus V_2 \oplus V_1^3 \oplus V_2^3,$$

where V_1 is spanned by the vector $e_1 + e_2 + e_3 + e_4 + e_5 + e_6$, V_2 is spanned by the vector $e_1 - e_2 - e_3 + e_4 + e_5 - e_6$, V_1^3 is spanned by the vectors $e_1 + e_2 - e_3 - e_4$, $e_3 + e_4 - e_5 - e_6$, and V_2^3 is spanned by the vectors $e_1 - e_3 - e_4 + e_6$, $e_2 + e_4 - e_5 - e_6$. The representation $\rho_{\mathbf{R}} \colon \mathfrak{S}_3 \to \mathbf{GL}(V_1)$ is equivalent to the irreducible representation $\rho_1 \colon \mathfrak{S}_3 \to \mathbf{U}(1)$, the representation $\rho_{\mathbf{R}} \colon \mathfrak{S}_3 \to \mathbf{GL}(V_2)$ is equivalent to the irreducible representation $\rho_2 \colon \mathfrak{S}_3 \to \mathbf{U}(1)$, and both representations $\rho_{\mathbf{R}} \colon \mathfrak{S}_3 \to \mathbf{GL}(V_k^3)$, $k = 1, 2$, are equivalent to the irreducible representation $\rho_3 \colon \mathfrak{S}_3 \to \mathbf{U}(2)$.

Theorem 13.19 shows that two finite-dimensional unitary representations $V_1 \colon G \to \mathbf{U}(H_2)$ and $V_2 \colon G \to \mathbf{U}(H_2)$ of G are equivalent if and only if $\chi_{V_1} = \chi_{V_2}$. This confirms the importance of the characters; they determine the equivalence classes of finite-dimensional unitary representations of a metrizable compact group.

Observe that the definition of the character of a representation makes sense even if the representation is not unitary (it only needs to be finite-dimensional). In view of Theorem 12.6 and the discussion following it, every *finite-dimensional* representation of G (not necessarily unitary) can be viewed as a unitary representation for some suitable hermitian inner product, so Proposition 13.18 and Theorem 13.19 also apply to such representations. Consequently, the characters also determine the equivalence classes of all finite-dimensional, not necessarily unitary, representations of a metrizable compact group.

If G is finite, it may be possible to build a character table for G by determining a complete set of irreducible representations of G (in view of the above remarks, not necessarily unitary). In general, this is difficult. It should be noted that for finite groups, using Peter–Weyl II to introduce characters and obtain some of their properties is a very heavy-handed method. A more gentle (and standard) approach is to define the characters of finite-dimensional representations and to derive their properties directly, singling out the role played by the characters of irreducible representations. Such an approach is presented in the excellent texts of Serre [90] and Simon [93]. Here is an example of the computation of the character table of the symmetric group $\mathfrak{S}_3$. Since we determined the irreducible representations of the symmetric group $\mathfrak{S}_3$ in Section 12.1, we can build its table of characters.

Example 13.5. In Example 13.4 we found the three irreducible unitary representations (up to equivalence) of the group $\mathfrak{S}_3$ consisting of the permutations on the set $\{1, 2, 3\}$. Recall that there are $3! = 6$ permutations

$$\sigma_1 = (1,2,3),\;\; \sigma_2 = (1,3,2),\;\; \sigma_3 = (2,1,3),\;\; \sigma_4 = (2,3,1),\;\; \sigma_5 = (3,1,2),\;\; \sigma_6 = (3,2,1),$$

and three conjugacy classes, $C_1 = \{\sigma_1\}$, $C_2 = \{\sigma_2, \sigma_3, \sigma_6\}$, $C_3 = \{\sigma_4, \sigma_5\}$. The two one-dimensional irreducible unitary representations are the trivial representation ρ_1 with

$$\rho_1(\sigma_i) = 1, \quad i = 1, \ldots, 6,$$

and the signature representation ρ_2 from Example 12.5, with

$$\rho_2(\sigma_i) = \begin{cases} +1 & \sigma_i \in C_1 \\ -1 & \sigma_i \in C_2 \\ +1 & \sigma_i \in C_3. \end{cases}$$

The corresponding characters χ_1, χ_2 are the central functions obtained by taking traces, in this scalar case the identity, so we get

$$\chi_1(\sigma_i) = \begin{cases} 1 & \sigma_i \in C_1 \\ 1 & \sigma_i \in C_2 \\ 1 & \sigma_i \in C_3 \end{cases} \quad \chi_2(\sigma_i) = \begin{cases} +1 & \sigma_i \in C_1 \\ -1 & \sigma_i \in C_2 \\ +1 & \sigma_i \in C_3. \end{cases}$$

The third irreducible unitary representation $\rho_3 \colon \mathfrak{S}_3 \to \mathbf{U}(2)$ is given by the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix},$$
$$\begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}, \quad \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}.$$

By taking traces we obtain

$$\chi_3(\sigma_i) = \begin{cases} 2 & \sigma_i \in C_1 \\ 0 & \sigma_i \in C_2 \\ -1 & \sigma_i \in C_3. \end{cases}$$

Thus we obtain the following character table for $\mathfrak{S}_3$.

	C_1	C_2	C_3
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

For much more about the representation of finite groups, see Serre [90], Simon [93] and Fulton and Harris [36]. The conjugacy classes and the characters of the symmetric group are discussed in Fulton and Harris [36] and Simon [93]. This beautiful theory makes use of Young tableaux.

Operations on (finite-dimensional) vector space induce operations on finite-dimensional, not necessarily unitary, representations, which in turn induce operations on their characters.

Given two finite-dimensional representations $U_1 \colon G \to \mathbf{U}(H_2)$ and $U_2 \colon G \to \mathbf{U}(H_2)$, with $d_1 = \dim(H_1)$ and $d_2 = \dim(H_2)$, we already defined their *direct sum* as the representation $U_1 \oplus U_2$ of G in $H_1 \oplus H_2$ given by

$$(U_1 \oplus U_2)(s)(x_1 + x_2) = U_1(s)(x_1) + U_2(s)(x_2), \quad s \in G,\ x_1 \in H_1,\ x_2 \in H_2.$$

If H_1 and H_2 are two finite-dimensional vector spaces, following Serre, the tensor product of H_1 and H_2 can be defined in a way that avoids the rather abstract universal mapping property.

Definition 13.9. If H_1 and H_2 are two finite-dimensional (real or complex) vector spaces, a tensor product $H_1 \otimes H_2$ of H_1 and H_2 is a (real or complex) vector space together with a map $\iota_\otimes \colon H_1 \times H_1 \to H_1 \otimes H_2$ such that the following two conditions hold:

(1) The map $\iota_\otimes \colon H_1 \times H_1 \to H_1 \otimes H_2$ is bilinear. For any $u \in H_1$ and any $v \in H_2$, we denote $\iota_\otimes(u, v)$ by $u \otimes v$.

(2) For any basis $(u_1, \ldots, u_m)$ of H_1 and any basis $(v_1, \ldots, v_n)$ of H_2, the $m \times n$ vectors $u_i \otimes v_j$ form a basis of $H_1 \otimes H_2$.

By standard methods of linear algebra it can shown that such a space $H_1 \otimes H_2$ exists and is unique up to isomorphism. The tensor product $H_1 \otimes H_2$ has the following *universal mapping property*: for every vector space F and every bilinear map $f\colon H_1 \times H_2 \to F$, there is a *unique linear map* $f_\otimes\colon H_1 \otimes H_2 \to F$ such that

$$f = f_\otimes \circ \iota_\otimes,$$

as illustrated in the following diagram:

$$\begin{tikzcd} H_1 \times H_2 \arrow[r, "\iota_\otimes"] \arrow[dr, "f"'] & H_1 \otimes H_2 \arrow[d, "f_\otimes"] \\ & F. \end{tikzcd}$$

Given two linear maps $f\colon E \to E'$ and $g\colon F \to F'$, there is a unique linear map

$$f \otimes g\colon E \otimes F \to E' \otimes F'$$

such that

$$(f \otimes g)(u \otimes v) = f(u) \otimes g(v), \quad \text{for all } u \in E \text{ and all } v \in F.$$

This is because we can define $h\colon E \times F \to E' \otimes F'$ by

$$h(u,\, v) = f(u) \otimes g(v).$$

It is immediately verified that h is bilinear, and thus by the universal mapping property it induces a unique linear map

$$f \otimes g\colon E \otimes F \to E' \otimes F'$$

making the following diagram commutes

$$\begin{tikzcd} E \times F \arrow[r, "\iota_\otimes"] \arrow[dr, "h"'] & E \otimes F \arrow[d, "f\otimes g"] \\ & E' \otimes F', \end{tikzcd}$$

such that $(f \otimes g)(u \otimes v) = f(u) \otimes g(v)$, for all $u \in E$ and all $v \in F$. For proofs of the above facts (in a more general framework) see Gallier and Quaintance [39] (Chapter 2).

In terms of matrices, given a basis $(u_1, \ldots, u_{d_1})$ of H_1 and a basis $(v_1, \ldots, v_{d_2})$ of H_2, assume f_1 is represented by the matrix A_1 and f_2 is represented by the matrix A_2. Then with respect to the basis $(u_i \otimes v_j)_{1 \leq i \leq d_1, 1 \leq j \leq d_2}$, the linear map $f_1 \otimes f_2$ is defined by a $(d_1 d_2) \times (d_1 d_2)$ matrix; as a block matrix, it is the $d_1 \times d_1$ matrix of $d_2 \times d_2$ blocks where the (i, j) block is the matrix $(A_1)_{ij} A_2$ $(1 \leq i, j \leq d_1)$. This matrix is called the *Kronecker product* of A_1 and A_2.

Given a complex vector space H, recall that $\overline{H}$ is the complex vector space with the same additive operation $+$ but with multiplication by a scalar defined by

$$(\lambda, u) \mapsto \overline{\lambda} u, \quad u \in H, \lambda \in \mathbb{C}.$$

Then a map $f\colon H \to \mathbb{C}$ is *semilinear* iff $f\colon \overline{H} \to \mathbb{C}$ is linear, which means that

$$\begin{aligned} f(u+v) &= f(u) + f(v) \\ f(\lambda u) &= \overline{\lambda} u, \end{aligned}$$

for all $u, v \in H$ and all $\lambda \in \mathbb{C}$. Observe that a map $\varphi\colon H \times H \to \mathbb{C}$ is *sesquilinear*, which means linear in its first argument and semilinear in its second argument, iff $\varphi\colon H \times \overline{H} \to \mathbb{C}$ is bilinear.

If $(H_1, \langle -, - \rangle_1)$ and $(H_2, \langle -, - \rangle_2)$ are finite-dimensional complex vector spaces each equipped with a hermitian inner product $\langle -, - \rangle_i$ $(i = 1, 2)$, the map $\langle -, - \rangle\colon (H_1 \times H_2) \times (\overline{H_1} \times \overline{H_2}) \to \mathbb{C}$ is defined as follows: for all $u_1, u_2 \in H_1$ and all $v_1, v_2 \in H_2$,

$$\langle (u_1, v_1), (u_2, v_2) \rangle = \langle u_1, u_2 \rangle_1 \langle v_1, v_2 \rangle_2. \tag{$\langle\,\rangle$}$$

It is immediately verified that this map is linear in each of its arguments. By the universal mapping property, the above map extends to a unique bilinear map $\langle -, - \rangle_{\otimes}\colon (H_1 \otimes H_2) \times (\overline{H_1} \otimes \overline{H_2}) \to \mathbb{C}$ such that

$$\langle u_1 \otimes v_1, u_2 \otimes v_2 \rangle_{\otimes} = \langle u_1, u_2 \rangle_1 \langle v_1, v_2 \rangle_2$$

for all $u_1, u_2 \in H_1$ and all $v_1, v_2 \in H_2$. However, $\overline{H_1} \otimes \overline{H_2}$ is isomorphic to $\overline{H_1 \otimes H_2}$, so we obtain a sesquilinear map $\langle -, - \rangle\colon (H_1 \otimes H_2) \times (H_1 \otimes H_2) \to \mathbb{C}$. Since

$$\begin{aligned} \langle u_2 \otimes v_2, u_1 \otimes v_1 \rangle_{\otimes} &= \langle u_2, u_1 \rangle_1 \langle v_2, v_1 \rangle_2 = \overline{\langle u_1, u_2 \rangle_1}\, \overline{\langle v_1, v_2 \rangle_2} \\ &= \overline{\langle u_1, u_2 \rangle_1 \langle v_1, v_2 \rangle_2} = \overline{\langle u_1 \otimes v_1, u_2 \otimes v_2 \rangle_{\otimes}}, \end{aligned}$$

the sesquilinear form $\langle -, - \rangle\colon (H_1 \otimes H_2) \times (H_1 \otimes H_2) \to \mathbb{C}$ is hermitian. Finally, observe that

$$\langle u_1 \otimes v_1, u_1 \otimes v_1 \rangle_{\otimes} = \langle u_1, u_1 \rangle_1 \langle v_1, v_1 \rangle_2,$$

and $\langle u_1, u_1 \rangle_1 \langle v_1, v_1 \rangle_2 > 0$ iff $\langle u_1, u_1 \rangle_1 > 0$ and $\langle v_1, v_1 \rangle_2 > 0$ iff $u_1 \neq 0$ and $v_1 \neq 0$, which means that our inner product is positive definite. Therefore the map $\langle -, - \rangle\colon (H_1 \otimes H_2) \times (H_1 \otimes H_2) \to \mathbb{C}$ uniquely defined by

$$\langle u_1 \otimes v_1, u_2 \otimes v_2 \rangle_{\otimes} = \langle u_1, u_2 \rangle_1 \langle v_1, v_2 \rangle_2$$

for all $u_1, u_2 \in H_1$ and all $v_1, v_2 \in H_2$, is a hermitian inner product on $H_1 \otimes H_2$.

Definition 13.10. If $(H_1, \langle -,-\rangle_1)$ and $(H_2, \langle -,-\rangle_2)$ are two finite-dimensional complex vector spaces each equipped with a hermitian inner product $\langle -,-\rangle_i$ $(i = 1, 2)$, there is a unique hermitian inner product $\langle -,-\rangle_\otimes \colon (H_1 \otimes H_2) \times (H_1 \otimes H_2) \to \mathbb{C}$ on $H_1 \otimes H_2$ satisfying the equation

$$\langle u_1 \otimes v_1, u_2 \otimes v_2\rangle_\otimes = \langle u_1, u_2\rangle_1 \langle v_1, v_2\rangle_2 \qquad (\langle\,\rangle_\otimes)$$

for all $u_1, u_2 \in H_1$ and all $v_1, v_2 \in H_2$.

Observe that if $(u_1, \ldots, u_{d_1})$ is an othonormal basis of H_1 and $(v_1, \ldots, v_{d_2})$ is an orthonormal basis of H_2, then $(u_i \otimes v_j)_{1\leq i\leq d_1, 1\leq j\leq d_2}$ is an orthonormal basis of $H_1 \otimes H_2$ with respect to the inner product $\langle -,-\rangle_\otimes$.

If $f_1 \colon H_1 \to H_1$ and $f_2 \colon H_2 \to H_2$ are unitary linear maps, then for all $u_1, u_2 \in H_1$ and all $v_1, v_2 \in H_2$, we have

$$\begin{aligned}
\langle (f_1 \otimes f_2)(u_1 \otimes v_1),\ (f_1 \otimes f_2)(u_2 \otimes v_2)\rangle_\otimes &= \langle f_1(u_1) \otimes f_2(v_1),\ f_1(u_2) \otimes f_2(v_2)\rangle \\
&= \langle f_1(u_1),\ f_1(u_2)\rangle_1 \langle f_2(v_1),\ f_2(v_2)\rangle_2 \\
&= \langle u_1, u_2\rangle_1 \langle v_1, v_2\rangle_2 \\
&= \langle u_1 \otimes v_1,\ u_2 \otimes v_2\rangle_\otimes,
\end{aligned}$$

which proves that $f_1 \otimes f_2 \colon H_1 \otimes H_2 \to H_1 \otimes H_2$ is unitary for the hermitian inner product $\langle -,-\rangle_\otimes$ on $H_1 \otimes H_2$. As a consequence of all this we can make the following definition.

Definition 13.11. Given two finite-dimensional unitary representations $U_1 \colon G \to \mathbf{U}(H_2)$, $U_2 \colon G \to \mathbf{U}(H_2)$, we define the *tensor product* $U_1 \otimes U_2$ of U_1 and U_2 as the unitary representation $U_1 \otimes U_2 \colon G \to \mathbf{U}(H_1 \otimes H_2)$ of G in $H_1 \otimes H_2$ (with the hermitian inner product $\langle -,-\rangle_\otimes$ on $H_1 \otimes H_2$ defined in $(\langle\,\rangle_\otimes)$) given by

$$(U_1 \otimes U_2)(s) = U_1(s) \otimes U_2(s), \quad s \in G,$$

where $U_1(s) \otimes U_2(s)$ is the tensor product linear map given by

$$(U_1(s) \otimes U_2(s))(x_1 \otimes x_2) = U(s)(x_1) \otimes U(s)(x_2), \quad \text{for all } x_1 \in H_1, x_2 \in H_2.$$

In terms of matrices, the linear map $U_1(s) \otimes U_2(s)$ is defined by a $(d_1 d_2) \times (d_1 d_2)$ matrix, namely the Kronecker product of $U_1(s)$ and $U_2(s)$. As a block matrix, it is the $d_1 \times d_1$ matrix of $d_2 \times d_2$ blocks where the (i, j) block is the matrix $U_1(s)_{ij} U_2(s)$ $(1 \leq i, j \leq d_1)$.

It is well known that

$$\begin{aligned}
\mathrm{tr}(U_1(s) \oplus U_2(s)) &= \mathrm{tr}(U_1(s)) + \mathrm{tr}(U_2(s)) \\
\mathrm{tr}(U_1(s) \otimes U_2(s)) &= \mathrm{tr}(U_1(s))\mathrm{tr}(U_2(s)).
\end{aligned}$$

In particular, if $U_1 = M_{\rho'}$ and $U_2 = M_{\rho''}$, two irreducible representations, then since $\chi_{\rho'}\chi_{\rho''} = \mathrm{tr}(M_{\rho'} \otimes M_{\rho''})$, which is finite-dimensional, Proposition 13.18 implies that

$$\chi_{\rho'}\chi_{\rho''} = \sum_{\rho \in R} c^{\rho}_{\rho', \rho''} \chi_\rho, \qquad (\otimes)$$

where $c^{\rho}_{\rho',\rho''} \geq 0$ is an integer, the number of times that the representations M_ρ is contained in $M_{\rho'} \otimes M_{\rho''}$ (this is d_ρ).

The determination of the $c^{\rho}_{\rho',\rho''}$ is usually very difficult. When $G = \mathbf{SU}(2)$, the irreducible representations can be completely determined and the $c^{\rho}_{\rho',\rho''}$ turn out to be either 1 or 0; see Chapter 14. They play an important role in physics.

Since the characters are linearly independent, we see that they form a subring of $\mathcal{C}(G;\mathbb{C})$ spanned by the characters, which is a $\mathbb{Z}$-algebra having the trivial character as identity, where the characters form a basis over $\mathbb{Z}$, and whose multiplication table is given as above.

Remark: For every $\rho \in R$, the trivial representation is contained in $M_\rho \otimes M_{\overline{\rho}} = M_\rho \otimes \overline{M}_\rho$. Otherwise, by Proposition 13.10(4)(a,c) and by $(\otimes)$, we would have

$$0 = \sum_{\rho'} c^{\rho'}_{\rho,\overline{\rho}} \int_G \chi_{\rho'}\, d\lambda(s) = \int \chi_\rho(s)\overline{\chi_\rho(s)}\, d\lambda(s) = \int |\chi_\rho(s)|^2\, d\lambda(s),$$

which is absurd.

Since any irreducible representation V of G is equivalent to a unique representation M_ρ, we call ρ the *class* of V and we write $\rho = \mathrm{cl}(V)$. Any finite-dimensional representation V of G corresponds uniquely to the formal linear combinations $\mathrm{cl}(V) = \sum_{\rho\in R} d_\rho \rho$, over those ρ for which M_ρ occurs d_ρ times. The $\mathbb{Z}$-module $\mathbb{Z}^{(R)}$ of formal linear combinations $\sum_{\rho\in R_1} m_\rho \rho$, with $m_\rho \in \mathbb{Z}$ and R_1 a finite subset of R, is isomorphic to the subring of $\mathcal{C}(G;\mathbb{C})$ spanned by the characters, and we can give it a multiplication operation using formula $(\otimes)$. With this multiplication, we have

$$\mathrm{cl}(U_1 \otimes U_2) = \mathrm{cl}(U_1)\mathrm{cl}(U_2).$$

This ring is the *ring of linear representations of* G. It is a substitute for the group of characters $\widehat{G}$, when G is abelian.

13.4 The Fourier Transform for Compact Groups

First we need to discuss the notion of weak integral a bit more. The reasoning used in Section 12.3 can be immediately adapted to show the following fact. Let G be a locally compact group equipped with a Haar measure λ and let $A\colon G \to \mathbf{U}(H)$ be a map such that $s \mapsto A(s)(x)$ is continuous for any fixed $x \in H$, where H is a Hilbert space. For any function $f \in \mathrm{L}^1(G)$, for all $x, y \in H$, the function $s \mapsto f(s)\langle A(s)(x), y\rangle d\lambda(s)$ is integrable and the functional Φ_x given by

$$\Phi_x(y) = \int f(s)\langle A(s)(x), y\rangle d\lambda(s)$$

is a bounded linear functional on H, so by the Riesz representation theorem, there is a unique vector in H called a weak integral and denoted $\int f(s)A(s)(x)d\lambda(s)$ (or even $A(f)(x)$), such that

$$\left\langle \int f(s)A(s)(x)d\lambda(s), y \right\rangle = \int f(s)\langle A(s)(x), y\rangle d\lambda(s), \quad \text{for all } x, y \in H.$$

Note that here, A is not necessarily a representation of G.

In the special case where H is a finite-dimensional space of dimensional n, we can pick an orthonormal basis in H and we can view $A(s)$ as an $n \times n$ matrix whose entry $a(s)_{ij}$ is a function on G. In this case, x and y are vectors of dimension n, so we have

$$\begin{aligned}\int f(s)\langle A(s)(x), y\rangle d\lambda(s) &= \int \sum_{i,j}^{n} f(s)a(s)_{ij}x_j\overline{y_i}d\lambda(s) \\ &= \sum_{i,j}^{n} \int f(s)a(s)_{ij}d\lambda(s)x_j\overline{y_i} \\ & \left\langle \left(\int f(s)a(s)_{ij}d\lambda(s)\right)x, y\right\rangle.\end{aligned}$$

The above shows that the weak integral $\int f(s)A(s)(x)d\lambda(s)$ is equal to the product by x of the $n \times n$ matrix $\left(\int f(s)a(s)_{ij}d\lambda(s)\right)$ obtained by integrating every entry in the matrix $f(s)A(s)$. We also denote the matrix $\left(\int f(s)a(s)_{ij}d\lambda(s)\right)$ by $\int f(s)A(s)d\lambda(s)$, or even $A(f)$.

Remark: More generally, let $\mu \in \mathcal{M}^1(G)$ be a complex regular Borel measure and let $h\colon G \to H$ be a function from G to a Hilbert space H such that:

(1) For every $y \in H$, the map $s \mapsto \langle h(s), y\rangle$ belongs to $\mathrm{L}^1(G)$.

(2) The map $s \mapsto \|h(s)\|$ belongs to $\mathrm{L}^1(G)$.

Then there is a unique vector in H, denoted $\int h(s)d\mu$, such that

$$\left\langle \int h(s)d\mu, y\right\rangle = \int \langle h(s), y\rangle d\mu, \quad \text{for all } y \in H;$$

see Dieudonné [24] (Chapter XIII, Section 10). The quantity $\int h(s)d\mu$ is called the weak integral of h. If we have a map $A\colon G \to \mathbf{U}(H)$ as before, for every fixed $x \in H$, if we let $h(s) = A(s)(x)$ and $\mu = fd\lambda$, we obtain the weak integral $\int f(s)A(s)(x)d\lambda(s)$ as a special case. Even more general notions of weak integrals are discussed in Folland [33] (Appendix 3).

We now return to the case where G is a compact group. Recall that

$$M_\rho(s) = \left(\frac{1}{n_\rho}m_{ij}(s)\right),$$

We now apply the above discussion to the matrix

$$A(t) = M_\rho(t^{-1}s).$$

Note that due to the presence of t^{-1}, for s fixed, the map $t \mapsto M_\rho(t^{-1}s)$ is not a representation.

Using the notations introduced just after Theorem 13.2, the formula

$$f = \sum_{\rho \in R} f * u_\rho, \qquad f \in \mathcal{L}^2(G)$$

given by this theorem can be written as

$$f = \sum_{\rho \in R} \left(\sum_{j=1}^{n_\rho} (f * m_{jj}^{(\rho)}) \right). \tag{$*_1$}$$

But by definition,

$$(f * m_{jj}^{(\rho)})(s) = \int f(t) m_{jj}^{(\rho)}(t^{-1}s)\, d\lambda(t),$$

so we get

$$\sum_{j=1}^{n_\rho} (f * m_{jj}^{(\rho)})(s) = n_\rho \operatorname{tr} \left(\int f(t) M_\rho(t^{-1}s)\, d\lambda(t) \right)$$

and so

$$f(s) = \sum_{\rho \in R} n_\rho \operatorname{tr} \left(\int f(t) M_\rho(t^{-1}s)\, d\lambda(t) \right). \tag{$*_2$}$$

However, we also have

$$M_\rho(\check{f}) = \int f(t) M_\rho(t^{-1})\, d\lambda(t)$$

(where again we integrate term by term), because for every $x \in \mathbb{C}^{n_g}$, by definition the vector $M_\rho(\check{f})(x)$ is the unique vector $\Phi(x) \in \mathbb{C}^{n_\rho}$ such that

$$\langle \Phi(x), y \rangle = \int f(t^{-1}) \langle M_\rho(t)(x), y \rangle\, d\lambda(t) \qquad \text{for all } y \in \mathbb{C}^{n_\rho},$$

and since G is unimodular,

$$\langle \Phi(x), y \rangle = \int f(t^{-1}) \langle M_\rho(t)(x), y \rangle\, d\lambda(t) = \int f(t) \langle M_\rho(t^{-1})(x), y \rangle\, d\lambda(t)$$

so by definition of the weak integral,

$$\Phi(x) = \int f(t) M_\rho(t^{-1})(x)\, d\lambda(t).$$

Recall that we also have

$$M_\rho(t^{-1}s) = M_\rho(t^{-1}) M_\rho(s),$$

Therefore,

$$\begin{aligned}\operatorname{tr}\left(\int f(t)M_\rho(t^{-1}s)\,d\lambda(t)\right) &= \operatorname{tr}\left(\int f(t)M_\rho(t^{-1})M_\rho(s)\,d\lambda(t)\right)\\ &= \operatorname{tr}\left(\left(\int f(t)M_\rho(t^{-1})\,d\lambda(t)\right)M_\rho(s)\right)\\ &= \operatorname{tr}\left(M_\rho(\check{f})M_\rho(s)\right),\end{aligned}$$

and by substituting this result in $(*_2)$ we obtain

$$f(s) = \sum_{\rho\in R} n_\rho \operatorname{tr}\left(M_\rho(\check{f})M_\rho(s)\right) \qquad f\in \mathrm{L}^2(G),\ s\in G. \tag{FI$_1$}$$

The above suggests the following definition for the generalization of the Fourier transform to compact groups.

Definition 13.12. Let G be a compact group. For any function $f \in \mathrm{L}^1(G)$, the *Fourier transform* $\mathcal{F}(f)$ of f is the map with domain R given by

$$\mathcal{F}(f)(\rho) = M_\rho(\check{f}) = \int f(t)M_\rho(t^{-1})\,d\lambda(t) = \int f(t)(M_\rho(t))^*\,d\lambda(t), \qquad \rho\in R.$$

We can view $\mathcal{F}(f)(\rho)$ as being defined as a weak integral, or in view of the discussion at the beginning of this section as the result of integrating term by term the matrix $f(t)(M_\rho(t))^*$.

Observe that $\mathcal{F}(f)(\rho) \in \mathrm{M}_{n_\rho}(\mathbb{C}) \cong \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}^{n_\rho}, \mathbb{C}^{n_\rho})$. The Fourier transform of Definition 13.12 is the natural generalization of the definition of the Fourier transform when G is an abelian compact group (Definition 10.3),

$$\mathcal{F}(f)(\chi) = \int f(s)\overline{\chi(s)}\,d\lambda(s) = \int f(s)\chi(s^{-1})\,d\lambda(s);$$

the character χ is replaced by the irreducible representation M_ρ.

Remark: The Fourier transform of Definition 13.12 is related to the Fourier transform $\mathcal{F}_2$ defined by Hewitt and Ross [48] (Chapter VII, Definition 28.34) as the map

$$\mathcal{F}_2(f)(\rho) = \int f(t)\overline{M_\rho}(t)\,d\lambda(t) = \overline{M_\rho}(f), \qquad f\in\mathrm{L}^1(G),\ \rho\in R.$$

Let $D_\rho\colon \mathbb{C}^{n_\rho}\to\mathbb{C}^{n_\rho}$ be the semilinear map given by

$$D_\rho\left(\sum_{k=1}^{n_\rho}\alpha_k e_k\right) = \sum_{k=1}^{n_\rho}\overline{\alpha_k}\,e_k, \qquad \alpha_k\in\mathbb{C},$$

where $(e_1, \ldots, e_{n_\rho})$ is the canonical basis of $\mathbb{C}^{n_\rho}$. It is immediately verified that

$$\langle D_\rho x, D_\rho y\rangle = \langle y, x\rangle = \overline{\langle x, y\rangle},$$

and

$$D_\rho^2 = \mathrm{id}.$$

Then it is not hard to show (see Hewitt and Ross [48], Chapter IX, Lemma 34.1) that

$$\mathcal{F}(f)(\rho) = D_\rho \circ (\mathcal{F}_2(f)(\rho))^* \circ D_\rho.$$

The definition of the Fourier transform $\mathcal{F}$ given in Definition 13.12 is identical to the definition given by Kirillov; see [55] (Section 2.3) and by Folland [33] (Chapter 5, Section 5.3). It has the advantage that the Fourier cotransform has a simpler formulation, and since for p with $1 \leq p \leq \infty$, the spaces $\mathrm{L}^p(\widehat{G})$ are closed under adjunction and conjugation, all the results proved for the Fourier transform $\mathcal{F}_2$ in Hewitt and Ross [48] also hold for the Fourier transform $\mathcal{F}$.

Equation (FI_1) can also be written as

$$f(s) = \sum_{\rho\in R} n_\rho \,\mathrm{tr}\,(\mathcal{F}(f)(\rho)M_\rho(s)) \qquad f \in \mathrm{L}^2(G),\ s \in G. \tag{FI}$$

The Fourier transform $\mathcal{F}(f)$ is a function with domain R, the set of "equivalence classes" of irreducible representations of G, which plays the analog of $\widehat{G}$, to the space $\prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$, where $\mathrm{M}_{n_\rho}(\mathbb{C})$ is the algebra of $n_\rho \times n_\rho$ complex matrices. Every element $F \in \prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$ is an R-indexed sequence $F = (F(\rho))_{\rho\in R}$ of $n_\rho \times n_\rho$ matrices $F(\rho)$. Sequences in $\prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$ are added, multiplied, and multiplied by a scalar, componentwise. Thus $\prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$ is a (complex) algebra. Given $F \in \prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$, the adjoint F^* of F is defined componentwise by $F^* = (F_\rho^*)_{\rho\in R}$.

Definition 13.13. We define $\widehat{G}$ as $R(G)$ the set of indices of a complete set of unitary irreducible representations of G (see the comment just after Theorem 13.16).

Note the analogy to the situation where $G = \mathbb{T}$ and $\widehat{G} = \widehat{\mathbb{T}} = \mathbb{Z}$, except that, $\mathrm{L}^1(\widehat{\mathbf{T}}) = l^1(\mathbb{Z})$ consists of $\mathbb{Z}$-indexed sequences of complex numbers, but the $F(\rho)$ are matrices. By analogy with the case $G = \mathbb{T}$ and $\widehat{\mathbb{T}} = \mathbb{Z}$, where the numbers $\mathcal{F}(f)(m) = \widehat{f}(m)$ are the Fourier coefficients of $f \in \mathrm{L}^1(\mathbb{T})$, the endomorphisms $\mathcal{F}(f)(\rho) \in \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}^{n_\rho}, \mathbb{C}^{n_\rho})$, represented by matrices in $\mathrm{M}_{n_\rho}(\mathbb{C})$, can be viewed as *generalized Fourier coefficients* of $f \in \mathrm{L}^1(G)$, where G is a compact group.

The equation (FI) is a kind of Fourier inversion formula.

We can define the Fourier cotransform $\overline{\mathcal{F}}$, defined on $\prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$, with input in G, by

$$\overline{\mathcal{F}}(F)(s) = \sum_{\rho\in R} n_\rho \,\mathrm{tr}(F(\rho)M_\rho(s)), \qquad F \in \prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C}),\ s \in G. \tag{FC}$$

Of course, there is an issue of convergence with (FC). The space $\prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$ is just too big, so following Hewitt and Ross [48] (Chapter VII, Section 28.24), we define normed subspaces $\mathrm{L}^p(\widehat{G})$ as follows.

First we need to define some norms on $n \times n$ matrices introduced by von Neumann.

13.5 von Neumann Norms and the Algebras $\mathrm{L}^p(\widehat{G})$

Definition 13.14. Let $A \in \mathrm{M}_n(\mathbb{C})$ be any complex $n\times n$ matrix, and let $(\sigma_1, \ldots, \sigma_n)$ be the sequence of nonnegative square roots of the eigenvalues of A^*A listed in any order (the positive square roots are the singular values of A). For any p, $1 \leq p < \infty$, define the *von Neumann norm* $\|A\|_{\varphi_p}$ of A by

$$\|A\|_{\varphi_p} = \left(\sum_{k=1}^{n} \sigma_k^p\right)^{1/p},$$

and $\|A\|_{\varphi_\infty}$ by

$$\|A\|_{\varphi_\infty} = \max_{1\leq k\leq n} \sigma_k.$$

It is not obvious that the functions defined in Definition 13.14 are matrix norms, but this is proven in Hewitt and Ross, see [48] (Appendix D, Theorem D40).

Since $(\sigma_1^2, \ldots, \sigma_n^2)$ are the eigenvalues of A^*A, we see that

$$\|A\|_{\varphi_2}^2 = \sum_{k=1}^{n} \sigma_k^2 = \mathrm{tr}(A^*A) = \|A\|_{\mathrm{HS}}^2,$$

where $\|A\|_{\mathrm{HS}}$ is a *Hilbert–Schmidt norm*, also known as *Frobenius norm*, of A (see Definition B.6). We also have

$$\|A\|_{\varphi_1} = \sum_{k=1}^{n} \sigma_k,$$

and

$$\|A\|_{\varphi_\infty} = \max_{1\leq k\leq n} \sigma_k = \|A\|_2,$$

where $\|A\|_2$ is the *operator norm* induced by the 2-norm; see Definition B.7 and Proposition B.8.

Next we use the norms of Definition 13.14 to define norms on $\prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$.

Definition 13.15. For any fixed sequence $(a_\rho)_{\rho\in R}$ of reals $a_\rho \geq 1$, for any $F \in \prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$, if $1 \leq p < \infty$, define $\|F\|_p$ by

$$\|F\|_p = \left(\sum_{\rho\in R} a_\rho \|F(\rho)\|_{\varphi_p}^p\right)^{1/p},$$

and for $p = \infty$, let

$$\|F\|_\infty = \sup_{\rho \in R} \|F(\rho)\|_{\varphi_\infty},$$

where $\|F(\rho)\|_{\varphi_p}$ is the von Neumann p-norm of the matrix $F(\rho)$.

Following Hewitt and Ross [48] (Chapter VII, Section 28.24), we make the following definitions.

Definition 13.16. Denote $\prod_{\rho \in R} \mathrm{M}_{n_\rho}(\mathbb{C})$ by $\mathfrak{E}(\widehat{G})$. Pick a fixed sequence $(a_\rho)_{\rho \in R}$ of reals $a_\rho \geq 1$. Let $\mathfrak{E}(\widehat{G})_{0,0}$ be the subspace of $\mathfrak{E}(\widehat{G})$ consisting of all sequences $F = (F(\rho))_{\rho \in R}$ such that the set $\{\rho \in R \mid F(\rho) \neq 0\}$ is finite, and let $\mathfrak{E}(\widehat{G})_0$ be the subspace of $\mathfrak{E}(\widehat{G})$ consisting of all sequences $F = (F(\rho))_{\rho \in R}$ such that the set $\{\rho \in R \mid \|F(\rho)\|_{\varphi_\infty} \geq \epsilon\}$ is finite for all $\epsilon > 0$.

For any p with $1 \leq p \leq \infty$, we define $\mathrm{L}^p(R) = \mathrm{L}^p(\widehat{G})$ as

$$\mathrm{L}^p(\widehat{G}) = \left\{ F \in \prod_{\rho \in R} \mathrm{M}_{n_\rho}(\mathbb{C}) \mid \|F\|_p < \infty \right\} = \left\{ F \in \mathfrak{E}(\widehat{G}) \mid \|F\|_p < \infty \right\}.$$

The following results are shown in Hewitt and Ross [48] (Theorem 28.25 and Theorem 28.26).

Proposition 13.20. *Let G be a compact group. For any fixed sequence $(a_\rho)_{\rho \in R}$ of reals $a_\rho \geq 1$, for any p such that $1 \leq p \leq \infty$, the space $\mathrm{L}^p(\widehat{G})$ is a Banach space. For any $F \in \mathrm{L}^p(\widehat{G})$, we have $F^* \in \mathrm{L}^p(\widehat{G})$ and $\|F^*\|_p = \|F\|_p$. The space $\mathrm{L}^\infty(\widehat{G})$ is a Banach algebra under componentwise multiplication, and $\|FF^*\|_\infty = \|F\|_\infty^2$ for any $F \in \mathrm{L}^\infty(\widehat{G})$.*

The following result is also shown in Hewitt and Ross [48] (Theorem 28.27).

Proposition 13.21. *Let G be a compact group, and let $(a_\rho)_{\rho \in R}$ be any fixed sequence of reals $a_\rho \geq 1$. With the norm $\| \, \|_\infty$, the space $\mathfrak{E}(\widehat{G})_0$ is a closed two-sided ideal of $\mathrm{L}^\infty(\widehat{G})$. For any p such that $1 \leq p < \infty$, the space $\mathfrak{E}(\widehat{G})_{0,0}$ is a dense two-sided ideal of $\mathfrak{E}(\widehat{G})_0$, and a dense two-sided ideal of $\mathrm{L}^p(\widehat{G})$. Both $\mathfrak{E}(\widehat{G})_{0,0}$ and $\mathfrak{E}(\widehat{G})_0$ are closed under adjunction ($F \mapsto F^*$).*

It is also possible to define an inner product on $\mathrm{L}^p(\widehat{G})$ based on the following proposition shown in Hewitt and Ross [48] (Lemma 28.28).

Proposition 13.22. *Let G be a compact group, and let $(a_\rho)_{\rho \in R}$ be any fixed sequence of reals $a_\rho \geq 1$. For any p, $1 \leq p \leq \infty$, if q is defined such that $\frac{1}{p} + \frac{1}{q} = 1$, then for all $E \in \mathrm{L}^p(\widehat{G})$ and all $F \in \mathrm{L}^q(\widehat{G})$, the following facts hold:*

(1) The number

$$\langle E, F \rangle = \sum_{\rho \in R} a_\rho \mathrm{tr}(F_\rho^* E_\rho)$$

is well defined (the series converges absolutely).

(2) *We have*

$$\langle F, E\rangle = \overline{\langle E, F\rangle}.$$

(3) *(Hölder's inequality)*

$$|\langle E, F\rangle| \leq \|E\|_p \|F\|_q .$$

Then we have the following theorem shown in Hewitt and Ross [48] (Theorem 28.30) .

Theorem 13.23. *Let G be a compact group, and let $(a_\rho)_{\rho\in R}$ be any fixed sequence of reals $a_\rho \geq 1$. The space $\mathrm{L}^2(\widehat{G})$ is a Hilbert space with the inner product*

$$\langle E, F\rangle = \sum_{\rho\in R} a_\rho \mathrm{tr}(F_\rho^* E_\rho),$$

and we have

$$\|E\|_2^2 = \langle E, E\rangle.$$

We also have the following result shown in Hewitt and Ross [48] (Theorem 28.32).

Proposition 13.24. *Let G be a compact group, and let $(a_\rho)_{\rho\in R}$ be any fixed sequence of reals $a_\rho \geq 1$.*

(1) *For any p such that $1 \leq p \leq \infty$, if q is such that $\frac{1}{p} + \frac{1}{q} = 1$, for any $E \in \mathrm{L}^p(\widehat{G})$ and $F \in \mathrm{L}^q(\widehat{G})$, we have $EF \in \mathrm{L}^1(\widehat{G})$, and*

$$\|EF\|_1 \leq \|E\|_p \|F\|_q .$$

(2) *For any p, q such that $1 \leq p < q \leq \infty$, we have*

$$\mathrm{L}^p(\widehat{G}) \subseteq \mathrm{L}^q(\widehat{G})$$

and for every $E \in \mathrm{L}^p(\widehat{G})$,

$$\|E\|_q \leq \|E\|_p .$$

(3) *For any p such that $1 \leq p \leq \infty$, for all $E, F \in \mathrm{L}^p(\widehat{G})$, we have $EF \in \mathrm{L}^p(\widehat{G})$, and*

$$\|EF\|_p \leq \|E\|_p \|F\|_p .$$

We now have the following results about the Fourier transform on a compact group, generalizing similar results about the Fourier transform on $\mathbb{T}$. *From now on, we assume that the sequence $(a_\rho)_{\rho\in R}$ of reals $a_\rho \geq 1$ is the sequence of positive integers $(n_\rho)_{\rho\in R}$.*

Theorem 13.25. *Let G be a compact group.*

(1) *If we define the multiplication on* $\mathrm{L}^\infty(\widehat{G})$ *as* $(F_1 \cdot F_2)(\rho) = F_2(\rho)F_1(\rho)$*, then the map* $f \mapsto \mathcal{F}(f)$ *is a non norm-increasing injective involutive algebra homomorphism from* $\mathrm{L}^1(G)$ *into* $\mathrm{L}^\infty(\widehat{G})$*. In particular, for all* $f, g \in \mathrm{L}^1(G)$*, for all* $\rho \in R$*, we have*

$$(\mathcal{F}(f * g))(\rho) = \mathcal{F}(g)(\rho) \circ \mathcal{F}(f)(\rho).$$

(2) *For every* $\rho \in R$*, the map* $f \mapsto \mathcal{F}(f)(\rho)$ *is an algebra representation of* $\mathrm{L}^1(G)$ *in* $\mathbb{C}^{n_\rho}$*.*

Proof sketch. Theorem 13.25 is proven in Hewitt and Ross [48] (Theorem 28.36); see also Folland [33] (Section 5.3, Equations 5.17, 5.18). It is instructive to prove that

$$(\mathcal{F}(f * g))(\rho) = \mathcal{F}(g)(\rho) \circ \mathcal{F}(f)(\rho).$$

By definition as a weak integral, $(\mathcal{F}(f * g)(\rho))(x)$ is the unique vector such that

$$\langle (\mathcal{F}(f * g)(\rho))(x), y \rangle = \int \langle M_\rho^*(s)(x), y\rangle (f * g)(s)\, d\lambda(s) \qquad \text{for all } x, y \in \mathbb{C}^{n_\rho},$$

and using the fact $(\mathcal{F}(g)(\rho))(z)$ is the unique vector such that

$$\langle (\mathcal{F}(g)(\rho))(z), y \rangle = \int \langle M_\rho^*(s)(z), y\rangle g(s)\, d\lambda(s) \qquad \text{for all } y, z \in \mathbb{C}^{n_\rho},$$

and $(\mathcal{F}(f)(\rho))(x)$ is the unique vector such that

$$\langle (\mathcal{F}(f)(\rho))(x), y \rangle = \int \langle M_\rho^*(s)(x), y\rangle f(s)\, d\lambda(s) \qquad \text{for all } x, y \in \mathbb{C}^{n_\rho},$$

we have

$$\begin{aligned}
\int \langle M_\rho^*(s)(x), y\rangle (f * g)(s)\, d\lambda(s) &= \int \langle M_\rho^*(s)(x), y\rangle \int f(t) g(t^{-1}s)\, d\lambda(t)\, d\lambda(s) \\
&= \int \langle M_\rho^*(t^{-1}s) M_\rho^*(t)(x), y\rangle \int f(t) g(t^{-1}s)\, d\lambda(t)\, d\lambda(s) \\
&= \int \left(\int \langle M_\rho^*(t^{-1}s) M_\rho^*(t)(x), y\rangle g(t^{-1}s)\, d\lambda(s) \right) f(t)\, d\lambda(t) \\
&= \int \left(\int \langle M_\rho^*(s) M_\rho^*(t)(x), y\rangle g(s)\, d\lambda(s) \right) f(t)\, d\lambda(t) \\
&= \int \langle (\mathcal{F}(g)(\rho))(M_\rho^*(t)(x)), y\rangle f(t)\, d\lambda(t) \\
&= \int \langle M_\rho^*(t)(x), (\mathcal{F}(g)(\rho))^*(y)\rangle f(t)\, d\lambda(t) \\
&= \langle (\mathcal{F}(f)(\rho))(x), (\mathcal{F}(g)(\rho))^*(y)\rangle \\
&= \langle (\mathcal{F}(g)(\rho))((\mathcal{F}(f)(\rho))(x)), y\rangle,
\end{aligned}$$

which proves that

$$(\mathcal{F}(f * g))(\rho) = \mathcal{F}(g)(\rho) \circ \mathcal{F}(f)(\rho),$$

as claimed. $\square$

Remark: Notice that the order of f and g is switched on the right-hand side. This is the reason why, if we want $\mathcal{F}$ to be a homomorphism, that we have to switch the order of the arguments in the multiplication on $\mathrm{L}^\infty(\widehat{G})$. If we use the Fourier transform $\mathcal{F}_2$ instead of the Fourier transform $\mathcal{F}$, then we get

$$(\mathcal{F}_2(f * g))(\rho) = \mathcal{F}_2(f)(\rho) \circ \mathcal{F}_2(g)(\rho),$$

as in Hewitt and Ross [48].

Definition 13.17. Let G be a compact group. For any $\rho \in R$, let $\mathcal{T}_\rho$ be the space of functions from G to $\mathbb{C}$ spanned by the set of functions

$$s \mapsto \langle M_\rho(s)(x), y\rangle, \qquad x, y \in \mathbb{C}^{n_\rho},$$

called *matrix coefficients*. Let $\mathcal{T}(G)$ be the space of functions spanned by the set

$$\bigcup_{\rho \in R} \mathcal{T}_\rho(G).$$

Since the M_ρ are representations, we have $\mathcal{T}(G) \subseteq \mathcal{C}(G; \mathbb{C})$.

Theorem 13.26. *Let G be a compact group.*

(1) For every $\rho \in R$, we have

$$\{\mathcal{F}(f)(\rho) \mid f \in \mathcal{T}_\rho(G)\} = \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}^{n_\rho}, \mathbb{C}^{n_\rho}).$$

(2) We have

$$\{\mathcal{F}(f) \mid f \in \mathcal{T}(G)\} = \mathfrak{C}_{0,0}(\widehat{G}).$$

Theorem 13.26 is proven in Hewitt and Ross [48] (Theorem 28.39).

Theorem 13.27. *Let G be a compact group. The map $f \mapsto \mathcal{F}(f)$ is a non norm-increasing involutive isomorphism of $\mathrm{L}^1(G)$ onto a dense subalgebra of $\mathfrak{C}_0(\widehat{G}) \subseteq \mathrm{L}^\infty(\widehat{G})$.*

Theorem 13.27 is proven in Hewitt and Ross [48] (Theorem 28.40). It is a version of the Riemann–Lebesgue lemma for compact groups; indeed, since $\widehat{G} = R$ is discrete, by definition of $\mathfrak{C}_0(\widehat{G})$, we can view the vectors in $\mathfrak{C}_0(\widehat{G})$ as functions of $\rho \in R$ that tend to zero at infinity. See Proposition 10.18 in the case of abelian locally compact groups.

Finally, we have the following version of Plancherel's theorem.

Theorem 13.28. *(Plancherel) Let G be a compact group. The map $f \mapsto \mathcal{F}(f)$ is an isometric isomorphism between the Hilbert space $\mathrm{L}^2(G)$ and the Hilbert space $\mathrm{L}^2(\widehat{G})$. If we pick any orthonormal basis $(e_1^\rho, \ldots, e_{n_\rho}^\rho)$ in $\mathbb{C}^{n_\rho}$, then for every $f \in \mathrm{L}^2(G)$, we have*

$$f = \sum_{\rho \in R} n_\rho \sum_{j,k=1}^{n_\rho} \langle (\mathcal{F}(f)(\rho))(e_k^\rho), e_j^\rho\rangle u_{jk}^\rho,$$

where u_{jk}^ρ is the function on G given by

$$u_{jk}^\rho(s) = \langle M_\rho(s)(e_k^\rho), e_j^\rho\rangle, \qquad s \in G,\ 1 \leq j, k \leq n_\rho.$$

The functions u^{ρ}_{jk} are called the *coordinate functions for M_ρ and the basis* $(e^{\rho}_1, \ldots, e^{\rho}_{n_\rho})$. Theorem 13.28 is proven in Hewitt and Ross [48] (Theorem 28.43).

We now return to the Fourier cotransform.

13.6 Fourier Inversion for Compact Groups

Definition 13.18. Let G be a compact group. For any $F \in \mathfrak{E}(\widehat{G}) = \prod_{\rho\in R} \mathrm{M}_{n_\rho}(\mathbb{C})$, the *Fourier cotransform* $\overline{\mathcal{F}}(F)$ of F is the function on G given by

$$\overline{\mathcal{F}}(F)(s) = \sum_{\rho\in R} n_\rho \operatorname{tr}(F(\rho)M_\rho(s)), \qquad s \in G.$$

In the above definition the infinite sum should be viewed as a formal expression. We will give below sufficient conditions that guarantee convergence.

Remark: Since Hewitt and Ross use the Fourier transform $\mathcal{F}_2$, related to the Fourier transform $\mathcal{F}$ by the equation

$$\mathcal{F}(f)(\rho) = D_\rho \circ (\mathcal{F}_2(f)(\rho))^* \circ D_\rho,$$

the definition of the Fourier cotransform (called inverse Fourier transform) given by Hewitt and Ross (Chapter IX, Section 34.47) is

$$\overline{\mathcal{F}}_2(F)(s) = \sum_{\rho\in R} n_\rho \operatorname{tr}(D_\rho F(\rho)^* D_\rho M_\rho(s)), \qquad s \in G.$$

Following Hewitt and Ross, it is natural to make the following definition (see [48], Definition 34.4).

Definition 13.19. Let G be a compact group. The subspace $\mathfrak{R}(G)$ of $\mathrm{L}^1(G)$ is defined by

$$\mathfrak{R}(G) = \{f \in \mathrm{L}^1(G) \mid \|\mathcal{F}(f)\|_1 < \infty\} = \left\{ f \in \mathrm{L}^1(G) \;\middle|\; \sum_{\rho\in R} n_\rho \left\|\mathcal{F}(f)(\rho)\right\|_{\varphi_1} < \infty \right\}.$$

The subspace $\mathfrak{R}(G)$ is called the space of *absolutely convergent Fourier series*. We define $\|f\|_{\varphi_1}$ by

$$\|f\|_{\varphi_1} = \|\mathcal{F}(f)\|_1 .$$

For any function $f \in \mathrm{L}^1(G)$, the formal expression

$$(\overline{\mathcal{F}}(\mathcal{F}(f)))(s) = \sum_{\rho\in R} n_\rho \operatorname{tr}(\mathcal{F}(f)(\rho)M_\rho(s))$$

is called the *Fourier series of f*.

Observe that Definition 13.19 is the generalization of the case $G = \mathbb{T}$ and $\widehat{\mathbb{T}} = \mathbb{Z}$ where for every $f \in \mathrm{L}^1(\mathbb{T})$ we define the Fourier series of f as the map

$$\theta \mapsto \sum_{m\in\mathbb{Z}} \widehat{f}(m)e^{im\theta} = (\overline{\mathcal{F}}(\widehat{f}))(\theta).$$

Here the character $\theta \mapsto e^{im\theta}$ is replaced by the irreducible representation M_ρ, and the trace function is needed to convert the matrix $\mathcal{F}(f)(\rho)M_\rho(s)$ to a number (and the dimensions n_ρ must be accounted for).

The following results are shown in Hewitt and Ross [48] (Theorem 34.5).

Theorem 13.29. *Let G be a compact group.*

(1) If $F \in \mathrm{L}^1(\widehat{G})$, then the map

$$s \mapsto \sum_{\rho\in R} n_\rho \left|\mathrm{tr}(F(\rho)M_\rho(s))\right|$$

is uniformly convergent.

(2) The map

$$s \mapsto (\overline{\mathcal{F}}(F))(s) = \sum_{\rho\in R} n_\rho \,\mathrm{tr}(F(\rho)M_\rho(s))$$

converges uniformly to a continuous function f. Furthermore, we have the Fourier inversion formula

$$(\overline{\mathcal{F}}(\mathcal{F}(f)))(s) = \sum_{\rho\in R} n_\rho \,\mathrm{tr}(\mathcal{F}(f)(\rho)M_\rho(s)), \qquad s \in G,$$

where $(\overline{\mathcal{F}}(\mathcal{F}(f)))(s)$ is the Fourier series of f, so $f \in \mathfrak{R}(G)$.

(3) We have

$$\|f\|_\infty \leq \|f\|_{\varphi_1} = \|\mathcal{F}(f)\|_1\,,$$

where $\|f\|_\infty$ is the sup norm on $\mathcal{C}(G;\mathbb{C})$.

The Fourier series of f is not necessarily convergent, but we have the following results; see Hewitt and Ross [48] (Corollary 34.6 and Corollary 34.7).

Theorem 13.30. *Let G be a compact group.*

(1) For any function $f \in \mathfrak{R}(G)$, the Fourier series of f converges uniformly and

$$f = (\overline{\mathcal{F}}(\mathcal{F}(f)))(s) = \sum_{\rho\in R} n_\rho \,\mathrm{tr}(\mathcal{F}(f)(\rho)M_\rho(s))$$

for almost all $s \in G$. We have

$$\|f\|_\infty \leq \|f\|_{\varphi_1}\,.$$

(2) The map $f \mapsto \mathcal{F}(f)$ is a norm-preserving linear isomorphism from $\mathfrak{R}(G)$ onto $\mathrm{L}^1(\widehat{G})$, so $\mathfrak{R}(G)$ is a Banach space.

For the record, in view of Theorem 13.28 and (FI), we have the following result (see also Hewitt and Ross [48], Chapter IX, Section 34.47(a)).

Theorem 13.31. *(Fourier inversion for $\mathrm{L}^2(G)$) Let G be a compact group. The Fourier cotransform $\overline{\mathcal{F}}(F) \in \mathrm{L}^2(G)$ of any $F \in \mathrm{L}^2(\widehat{G})$ converges as a series in the L^2-norm, and for every $f \in \mathrm{L}^2(G)$, we have*

$$f(s) = (\overline{\mathcal{F}}(\mathcal{F}(f)))(s) = \sum_{\rho \in R} n_\rho \operatorname{tr}\left(\mathcal{F}(f)(\rho) M_\rho(s)\right), \qquad s \in G$$

where the series converges in the L^2-norm.

There is another way to express (FI) in terms of the characters of G. Indeed, we have

$$\operatorname{tr}\left(\int f(t) M_\rho(t^{-1}s)\, d\lambda(t)\right) = \int f(t) \operatorname{tr}\left(M_\rho(t^{-1}s)\right)\, d\lambda(t) = (f * \chi_\rho)(s).$$

Therefore, by $(*_2)$ we have

$$f = \sum_{\rho \in R} n_\rho\, f * \chi_\rho, \qquad f \in \mathrm{L}^2(G). \tag{FI'}$$

See also Hewitt and Ross [48] (Chapter XII, Theorem 27.41). This generalizes Proposition 13.10(4)(b) when f is a central function.

Example 13.6. If G is a finite group, then $\widehat{G} = \{\rho_1, \ldots, \rho_r\}$, where r is the number of conjugacy classes of G. If we give G the counting measure normalized so that G has measure 1, then the Fourier transform of any function $f \in \mathrm{L}^2(G)$ is given by

$$\mathcal{F}(f)(\rho) = \frac{1}{|G|} \sum_{s \in G} f(s) (M_\rho(s))^*,$$

where $M_{\rho_1}, \ldots, M_{\rho_r}$ are the irreducible representations of G (up to equivalence). For every $F \in \mathrm{L}^2(\widehat{G})$, the Fourier cotransform of F is given by

$$\overline{\mathcal{F}}(F)(s) = \sum_{k=1}^{r} n_{\rho_k} \operatorname{tr}(F(\rho_k) M_{\rho_k}(s)), \qquad s \in G,$$

and the Fourier inversion formula is given by

$$f = \sum_{k=1}^{r} n_{\rho_k} \operatorname{tr}((\mathcal{F}(f))(\rho_k) M_{\rho_k}(s)).$$

The fact that $\mathcal{F}$ is an isometry is expressed by the equation

$$\langle f_1, f_2\rangle = \frac{1}{|G|}\sum_{s\in G} f_1(s)\overline{f_2(s)} = \langle \mathcal{F}(f_1), \mathcal{F}(f_2)\rangle = \sum_{k=1}^{r} n_{\rho_k}\mathrm{tr}((\mathcal{F}(f_2))^*\mathcal{F}(f_1)),$$

for all $f_1, f_2 \in \mathrm{L}^2(G)$.

For all $f, g \in \mathrm{L}^2(G)$, the convolution $f * g$ is given by

$$(f * g)(s) = \frac{1}{|G|}\sum_{s_1 s_2 = s} f(s_1)g(s_2) = \frac{1}{|G|}\sum_{t\in G} f(t)g(t^{-1}s),$$

and we can write explicitly the equation

$$(\mathcal{F}(f * g))(\rho) = (\mathcal{F}(g))(\rho) \circ (\mathcal{F}(f))(\rho).$$

We leave it to the diligence of the reader to check that it holds.

Chapter 14

Representations of $\mathbf{SU}(2)$ and Their Matrices

14.1 Irreducible Representations of $\mathbf{SU}(2)$ and $\mathbf{SO}(3)$

In Example 12.7 it was proven that the representations $U_m \colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_m^{\mathbb{C}}(2))$ are irreducible. In Example 12.8 it was proven that the representations $W_\ell \colon \mathbf{SO}(3) \to \mathbf{GL}(\mathcal{P}_{2\ell}^{\mathbb{C}}(2))$ are irreducible. Recall that since $\mathbf{SU}(2)$ is compact and $\mathcal{P}_m^{\mathbb{C}}(2)$ is finite-dimensional there is an invariant inner product on $\mathcal{P}_m^{\mathbb{C}}(2)$ so we may assume that these representations are unitary.

Let us now prove that the representations U_m form a complete set of irreducible unitary representations.

Proposition 14.1. *Every irreducible unitary representation of $\mathbf{SU}(2)$ is equivalent to one of the irreducible unitary representations $U_m \colon \mathbf{SU}(2) \to \mathbf{U}(\mathcal{P}_m^{\mathbb{C}}(2))$. Furthermore, every irreducible unitary representation of $\mathbf{SO}(3)$ is equivalent to one of the irreducible unitary representations $W_m \colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{P}_m^{\mathbb{C}}(2))$.*

Proof. The key point is to figure out what are the characters χ_{U_m} of the irreducible unitary representations U_m. Every unitary matrix $q \in \mathbf{SU}(2)$ is diagonalizable as

$$q = R r_x(\varphi) R^*$$

for some unitary matrix $R \in \mathbf{U}(2)$, where

$$r_x(\varphi) = \begin{pmatrix} e^{i\varphi} & 0 \\ 0 & e^{-i\varphi} \end{pmatrix}.$$

If $\det(R) = e^{i\omega} \neq 1$, we replace R by $e^{-i\omega/2}R$, which is unitary ($(e^{-i\omega/2}R)(e^{-i\omega/2}R)^* = e^{-i\omega/2}Re^{i\omega/2}R^* = I_2$), so that $\det(e^{-i\omega/2}R) = (e^{-i\omega/2})^2\det(R) = e^{-i\omega}e^{i\omega} = +1$, and then

$$(e^{-i\omega/2}R)r_x(\varphi)(e^{-i\omega/2}R)^* = e^{-i\omega/2}Rr_x(\varphi)e^{i\omega/2}R^* = Rr_x(\varphi)R^* = q.$$

Therefore we may assume that $R \in \mathbf{SU}(2)$. Here $-\pi \leq \varphi \leq \pi$, but if $-\pi \leq \varphi < 0$, we can replace R by

$$R \begin{pmatrix} 0 & e^{i\pi/2} \\ e^{i\pi/2} & 0 \end{pmatrix} \in \mathbf{SU}(2),$$

because then

$$\begin{aligned}
\begin{pmatrix} 0 & e^{i\pi/2} \\ e^{i\pi/2} & 0 \end{pmatrix} \begin{pmatrix} e^{i\varphi} & 0 \\ 0 & e^{-i\varphi} \end{pmatrix} \begin{pmatrix} 0 & e^{-i\pi/2} \\ e^{-\pi/2} & 0 \end{pmatrix} &= \begin{pmatrix} 0 & e^{-i(\varphi-\pi/2)} \\ e^{i(\varphi+\pi/2)} & 0 \end{pmatrix} \begin{pmatrix} 0 & e^{-i\pi/2} \\ e^{-i\pi/2} & 0 \end{pmatrix} \\
&= \begin{pmatrix} e^{-i\varphi} & 0 \\ 0 & e^{i\varphi} \end{pmatrix}.
\end{aligned}$$

Thus we may assume that $0 \leq \varphi \leq \pi$. *Therefore we proved that every matrix in* $\mathbf{SU}(2)$ *is conjugate to a unique matrix* $r_x(\varphi)$, *with* $0 \leq \varphi \leq \pi$.

Since the characters are central functions (Proposition 13.10(1)), it suffices to compute the value of the character χ_{U_m} on $r_x(\varphi)$. Also observe that $r_x(\varphi)$ and $r_x(\theta)$ are conjugate iff they have the same eigenvalues iff $\varphi = \pm\theta \pmod{2\pi}$. But then we obtain a bijection between the space of central functions of $\mathrm{L}^2(\mathbf{SU}(2))$ and the space of even continuous 2π-periodic functions from $\mathbb{R}$ to $\mathbb{C}$ given by $f \mapsto S(f)$, with

$$S(f)(\varphi) = f(r_x(\varphi)).$$

We will compute the values $\chi_{U_m}(r_x(\varphi))$ and prove that the characters χ_{U_m} are dense in the the space of central functions of $\mathrm{L}^2(\mathbf{SU}(2))$.

Recall from Example 12.7 that the eigenvalues of $U_m(r_x(\varphi))$ are $(e^{im\varphi}, e^{i(m-2)\varphi}, \ldots, e^{-im\varphi})$. Therefore,

$$\chi_{U_m}(r_x(\varphi)) = \mathrm{tr}(U_m(r_x(\varphi))) = \sum_{k=0}^{m} e^{i(m-2k)\varphi}.$$

We have

$$\begin{aligned}
\sum_{k=0}^{m} e^{i(m-2k)\varphi} &= e^{im\varphi} \sum_{k=0}^{m} (e^{-i2\varphi})^k = e^{im\varphi} \frac{(1-(e^{-i2\varphi})^{m+1})}{1-e^{-i2\varphi}} \\
&= e^{im\varphi} \frac{e^{i\varphi}(1-(e^{-i2\varphi})^{m+1})}{e^{i\varphi}(1-e^{-i2\varphi})} = e^{i(m+1)\varphi} \frac{(1-e^{-i2(m+1)\varphi})}{e^{i\varphi}-e^{-i\varphi}} \\
&= \frac{(e^{i(m+1)\varphi}-e^{-i(m+1)\varphi})}{e^{i\varphi}-e^{-i\varphi}} = \frac{\sin((m+1)\varphi)}{\sin\varphi}.
\end{aligned}$$

We also easily check that

$$\chi_{U_m}(r_x(k\pi)) = e^{imk\pi}(m+1) = (-1)^{mk}(m+1).$$

In summary we obtained the following result.

For every $q \in \mathbf{SU}(2)$, if $r_x(\varphi)$ is the unique diagonal matrix conjugate to q with $0 \leq \varphi \leq \pi$, then $\chi_{U_m}(q)$ is given by

$$\chi_{U_m}(q) = \chi_{U_m}(r_x(\varphi)) = \frac{\sin((m+1)\varphi)}{\sin\varphi}.$$

If we write

$$\kappa_m(\varphi) = \chi_{U_m}(r_x(\varphi)) = \frac{\sin((m+1)\varphi)}{\sin\varphi},$$

then for $m \geq 1$ we get

$$\begin{aligned}\kappa_m(\varphi) = \frac{\sin((m+1)\varphi)}{\sin\varphi} &= \frac{\sin(m\varphi)\cos\varphi + \cos(m\varphi)\sin\varphi}{\sin\varphi} \\ &= \cos(m\varphi) + \kappa_{m-1}(\varphi)\cos\varphi.\end{aligned}$$

Note that $\kappa_0(\varphi) = 1$. The formula for $\kappa_m(\varphi)$ still holds for $\varphi = k\pi$. In summary,

$$\kappa_m(\varphi) = \cos(m\varphi) + \kappa_{m-1}(\varphi)\cos\varphi, \quad m \geq 1, \quad \kappa_0(\varphi) = 1. \tag{κ}$$

The above equation shows that $\kappa_0(\varphi), \kappa_1(\varphi), \ldots, \kappa_m(\varphi)$ generates the same vector space as $1, \cos\varphi, \ldots, \cos m\varphi$.

It is known from Fourier analysis that the space generated by the family of functions $(\cos m\varphi))_{m\geq 0}$ is dense in the space of even 2π-periodic continuous functions from $\mathbb{R}$ to $\mathbb{C}$; for example, see Folland [34, 32]. Consequently the family $(\kappa_m)_{m\geq 0}$ is also dense in the space of even continuous 2π-periodic functions from $\mathbb{R}$ to $\mathbb{C}$. Since the map $f \mapsto S(f)$ is a bijection between the space of central functions of $\mathrm{L}^2(\mathbf{SU}(2))$ and the space of even continuous 2π-periodic functions from $\mathbb{R}$ to $\mathbb{C}$, we conclude that the family of characters $(\chi_{U_m})_{m\geq 0}$ is dense in the the space of central functions of $\mathrm{L}^2(\mathbf{SU}(2))$.

To finish the proof we use Proposition 13.10(4) which says that the characters χ_m of $\mathbf{SU}(2)$ form a Hilbert basis of the space of central functions of $\mathrm{L}^2(\mathbf{SU}(2))$. Since the χ_{U_m} are characters of irreducible unitary representations they are equal to some of the characters χ_ρ of $\mathbf{SU}(2)$, and they are not equivalent since the dimensions $m+1$ of the representing spaces are different. If some character χ_ρ is not equivalent to one of the χ_{U_m}, then by Proposition 13.10(4)(a),

$$\langle \chi_{U_m}, \chi_\rho \rangle = 0 \quad \text{for all} \quad m \geq 0,$$

but since the χ_{U_m} are dense in the space of central functions of $\mathrm{L}^2(\mathbf{SU}(2))$, this implies that $\chi_\rho = 0$, a contradiction.

The second statement follows from the fact that the unitary representations $W_\ell \colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{P}^{\mathbb{C}}_{2\ell}(2))$ given by

$$W_\ell(\rho_q) = U_{2\ell}(q) \quad q \in \mathbf{SU}(2), \quad \ell \geq 0. \qquad \square$$

We now give a more pleasant description of the irreducible representations of $\mathbf{SO}(3)$ in terms of harmonic polynomials.

14.2 Irreducible Representations of $\mathbf{SO}(3)$; Harmonics

Recall that the Laplacian in $\mathbb{R}^n$ is given by

$$\Delta f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}$$

where $f\colon \mathbb{R}^n \to \mathbb{C}$ is twice differentiable. The n-sphere $S^n \subseteq \mathbb{R}^{n+1}$ is given by

$$S^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \cdots + x_{n+1}^2 = 1\}.$$

Definition 14.1. Let $\mathcal{P}_k^{\mathbb{C}}(n+1)$ denote the space of homogeneous polynomials of degree k in $n+1 \geq 2$ variables with complex coefficients, and let $\mathcal{P}_k^{\mathbb{C}}(S^n)$ denote the restrictions of homogeneous polynomials in $\mathcal{P}_k^{\mathbb{C}}(n+1)$ to S^n. Let $\mathcal{H}_k^{\mathbb{C}}(n+1)$ denote the space of *complex harmonic polynomials*, with

$$\mathcal{H}_k^{\mathbb{C}}(n+1) = \{P \in \mathcal{P}_k^{\mathbb{C}}(n+1) \mid \Delta P = 0\};$$

in the above equation, we view P as a function on $\mathbb{R}^{n+1}$. Harmonic polynomials are sometimes called *solid harmonics*. Finally, let $\mathcal{H}_k^{\mathbb{C}}(S^n)$ denote the space of *complex spherical harmonics* as the set of restrictions of harmonic polynomials in $\mathcal{H}_k^{\mathbb{C}}(n+1)$ to S^n.

It not hard to prove that the restriction map from $\mathcal{H}_k^{\mathbb{C}}(n+1)$ to $\mathcal{H}_k^{\mathbb{C}}(S^n)$ is a bijection, and thus a linear isomorphism; see Gallier and Quaintance [39] (Section 7.5). The functions in $\mathcal{H}_k^{\mathbb{C}}(S^n)$, the spherical harmonics, have been studied extensively. They are the eigenspaces of the Laplacian on the sphere S^n; see Gallier and Quaintance [39] (Chapter 7). We will return to these functions later.

The group $\mathbf{SO}(n+1)$ acts on $\mathcal{P}_k^{\mathbb{C}}(n+1)$ by the (left regular) action

$$(\mathbf{R}_Q(P))(x) = P(Q^{-1}x), \quad Q \in \mathbf{SO}(n+1),\ P \in \mathcal{P}_k^{\mathbb{C}}(n+1),\ x \in \mathbb{R}^{n+1}.$$

Note that the above formula shows that $\mathbf{R}$ is also an action of $\mathbf{SO}(n+1)$ on smooth functions on $\mathbb{R}^{n+1}$.

The action $\mathbf{R}$ on $\mathcal{P}_k^{\mathbb{C}}(n+1)$ is reducible for $k \geq 2$. For example, we easily check that the subspace of $\mathcal{P}_2^{\mathbb{C}}(n+1)$ generated by the polynomial $x_1^2 + \cdots + x_{n+1}^2$ is invariant. However this action turns out to be irreducible on $\mathcal{H}_k^{\mathbb{C}}(n+1)$. This will be shown in Section 15.11. But first we need to prove that the action of the Laplacian on smooth functions on $\mathbb{R}^{n+1}$ commutes with the action $\mathbf{R}$. Recall that $\lambda_Q f$ is the function given by $(\lambda_Q f)(x) = f(Q^{-1}x)$.

Proposition 14.2. *The action of the Laplacian on smooth functions on $\mathbb{R}^{n+1}$ commutes with the action $\mathbf{R}$; that is, for every smooth function f on $\mathbb{R}^{n+1}$, for every $Q \in \mathbf{SO}(n+1)$, for all $u \in \mathbb{R}^{n+1}$, we have*

$$\Delta(\lambda_Q f)(u) = (\Delta f)(Q^{-1}u).$$

Proof. For simplicity of notation write $A = Q^{-1} = Q^\top$. The proof makes a heavy use of the chain rule. If we let h be the function given by $h(x) = Ax$ and view f as a function $y \mapsto f(y)$ of the variable y, if we write $g = f \circ h$ so that $g(x) = f(Ax)$, then we need to compute $(\partial g/\partial x_j)(u)$ $(x, u \in \mathbb{R}^{n+1})$, which by the chain rule is given by

$$\frac{\partial g}{\partial x_j}(u) = df_{h(u)}\left(\frac{\partial h}{\partial x_j}(u)\right).$$

Since

$$h(x) = Ax = \left(\sum_{j=1}^{n+1} a_{1j}x_j, \ldots, \sum_{j=1}^{n+1} a_{n+1j}x_j\right),$$

we have

$$\frac{\partial h}{\partial x_j}(u) = (a_{1j}, \ldots, a_{n+1j})$$

(independently of u), and since

$$df_{Au}(w) = \sum_{i=1}^{n+1} w_i \frac{\partial f}{\partial y_i}(Au),$$

we obtain

$$\frac{\partial g}{\partial x_j}(u) = \sum_{i=1}^{n+1} a_{ij}\frac{\partial f}{\partial y_i}(Au).$$

To compute $\frac{\partial^2 g}{\partial x_j^2}(u)$, we view the function $y \mapsto \frac{\partial f}{\partial y_i}(y)$ as the function f, so we obtain

$$\frac{\partial^2 g}{\partial x_j^2}(u) = \sum_{i=1}^{n+1} a_{ij} \sum_{k=1}^{n+1} a_{kj} \frac{\partial^2 f}{\partial y_i \partial y_k}(Au),$$

and thus the Laplacian is given by

$$\Delta g(u) = \sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\sum_{k=1}^{n+1} a_{ij}a_{kj}\frac{\partial^2 f}{\partial y_i \partial y_k}(Au).$$

The right-hand side can be rewritten as

$$\begin{aligned}
\Delta g(u) &= \sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\sum_{k=1}^{n+1} a_{ij}a_{kj}\frac{\partial^2 f}{\partial y_i \partial y_k}(Au) \\
&= \sum_{i=1}^{n+1}\left(\sum_{j=1}^{n+1} a_{ij}^2\right)\frac{\partial^2 f}{\partial y_i^2}(Au) + 2\sum_{i<k}^{n+1}\left(\sum_{j=1}^{n+1} a_{ij}a_{kj}\right)\frac{\partial^2 f}{\partial y_i \partial y_k}(Au),
\end{aligned}$$

and since A is an orthogonal matrix, its rows have unit length and are pairwise orthogonal, which means that

$$\sum_{j=1}^{n+1} a_{ij}^2 = 1, \quad 1 \leq i \leq n+1$$
$$\sum_{j=1}^{n+1} a_{ij}a_{kj} = 0, \quad i < k,$$

so we obtain

$$\Delta g(u) = (\Delta f)(Au),$$

which means that $\Delta(\lambda_{A^{-1}} f)(u) = (\Delta f)(Au)$, as claimed. $\square$

As a corollary of Proposition 14.2, the vector space $\mathcal{H}_k^{\mathbb{C}}(n+1)$ is invariant under $\mathbf{R}$, and so $\mathbf{R} \colon \mathbf{SO}(n+1) \to \mathbf{GL}(\mathcal{H}_k^{\mathbb{C}}(n+1))$ is a representation. Since $\mathbf{SO}(n+1)$ is compact and $\mathcal{H}_k^{\mathbb{C}}(n+1)$ is finite-dimensional, we may assume that $\mathbf{R}$ is unitary.

It is shown in Gallier and Quaintance [39] (Section 7.5) that $\mathcal{H}_k^{\mathbb{C}}(n+1)$ has dimension

$$a_{k,n+1} = \binom{n+k}{k} - \binom{n+k-2}{k-2}$$

if $n \geq 1$, $k \geq 2$, with $a_{0,n+1} = 1$ and $a_{1,n+1} = n$, For $n = 2$, we get $a_{k,3} = 2k+1$. Here is a list of bases of the homogeneous harmonic polynomials of degree k in three variables up to $k = 4$.

$k = 0$	1
$k = 1$	$x,\ y,\ z$
$k = 2$	$x^2 - y^2,\ x^2 - z^2,\ xy,\ xz,\ yz$
$k = 3$	$x^3 - 3xy^2,\ 3x^2y - y^3,\ x^3 - 3xz^2,\ 3x^2z - z^3,$
	$y^3 - 3yz^2,\ 3y^2z - z^3,\ xyz$
$k = 4$	$x^4 - 6x^2y^2 + y^4,\ x^4 - 6x^2z^2 + z^4,\ y^4 - 6y^2z^2 + z^4,$
	$x^3y - xy^3,\ x^3z - xz^3,\ y^3z - yz^3,$
	$3x^2yz - yz^3,\ 3xy^2z - xz^3,\ 3xyz^2 - x^3y.$

To prove that the representations $\mathbf{R} \colon \mathbf{SO}(n+1) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(n+1))$ are irreducible we restrict ourselves to the case where $n = 2$. In order to deal with the case where $n > 2$, we need results from the next chapter. Since these regular representations map to different spaces, for clarity we index them by k, that is, we write $\mathbf{R}_k \colon \mathbf{SO}(n+1) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(n+1))$.

Proposition 14.3. *The representations $\mathbf{R}_k \colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(3))$ are irreducible. In fact, the representations $\mathbf{R}_k \colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(3))$ and $W_k \colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{P}_{2k}^{\mathbb{C}}(2))$ are equivalent.*

Proof. By Peter–Weyl II, the representation $\mathbf{R}_k\colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(3))$ is equivalent to the direct sum of a finite number of irreducible representations $W_{\ell_j}\colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{P}_{2\ell_j}^{\mathbb{C}}(2))$, so that we have an isomorphism

$$\mathcal{H}_k^{\mathbb{C}}(3) \approx \bigoplus_{j=1}^{p} W_{\ell_j}, \quad \ell_j \leq \ell_{j+1}.$$

Since $\dim(\mathcal{H}_k^{\mathbb{C}}(3)) = 2k+1$ and $\dim(W_{\ell_j}) = 2\ell_j + 1$, if we can prove that $\ell_j \geq k$ for some j, then $2k+1 = \sum_{j=1}^{p}(2\ell_j+1)$ implies that $p = 1$ and $k = \ell_1$, and so $\mathbf{R}_k\colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(3))$ is equivalent to $W_k\colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{P}_{2k}^{\mathbb{C}}(2))$.

The key point is to figure out what is the value of the character χ_{W_ℓ} on the rotation of angle φ and axis Ox given by

$$R_x(\varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}.$$

The trick is that $R_x(\varphi)$ is the rotation in $\mathbf{SO}(3)$ corresponding to the (familar) quaternion

$$r_x(\varphi/2) = \begin{pmatrix} e^{i\varphi/2} & 0 \\ 0 & e^{-i\varphi/2} \end{pmatrix}.$$

This fact is easily verified by direct computation. But remember that W_ℓ is given by

$$W_\ell(\rho_q) = U_{2\ell}(q) \quad q \in \mathbf{SU}(2), \quad \ell \geq 0,$$

which proves that

$$\chi_{W_\ell}(\rho_q) = \mathrm{tr}(W_\ell(\rho_q)) = \mathrm{tr}(U_{2\ell}(q)) = \chi_{U_{2\ell}}(q).$$

If we apply the above equation to $q = r_x(\varphi/2)$ and $R_x(\varphi)$, we obtain

$$\chi_{W_\ell}(R_x(\varphi)) = \chi_{U_{2\ell}}(r_x(\varphi/2)) = \sum_{j=0}^{2\ell} e^{i(2\ell-2j)\varphi/2} = \sum_{j=0}^{2\ell} e^{i(\ell-j)\varphi}.$$

Since $\mathbf{R}_k\colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(3))$ is equivalent to a finite direct sum of p irreducible representations $W_{\ell_j}\colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{P}_{2\ell_j}^{\mathbb{C}}(2))$, by Proposition 13.18, the value of the character $\chi_{\mathbf{R}_k}$ on $R_x(\varphi)$ is the sum of the values of the characters $\chi_{W_{\ell_j}}$ on $R_x(\varphi)$, and by the above equation, it is an integral combination of terms of the form $e^{ij\varphi}$, with $|j| \leq \ell_p$. Consequently, if we find an eigenvector of $\mathbf{R}_k(R_x(\varphi))$ for the eigenvalue $e^{-ik\varphi}$, the representation $W_k\colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{P}_{2k}^{\mathbb{C}}(2))$ must occur. Consider $P(x) = P(x_1, x_2, x_3) = (x_2 + ix_3)^k$. We immediately check that $\Delta P = 0$, and since

$$R_x(\varphi)^{-1} = R_x(-\varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix},$$

we obtain

$$\begin{aligned}(\mathbf{R}_k(R_x(\varphi)))(P) &= P(R_x(-\varphi))\\ &= P((\cos\varphi)x_2 + (\sin\varphi)x_3, -(\sin\varphi)x_2 + (\cos\theta)x_3)\\ &= ((\cos\varphi)x_2 + (\sin\varphi)x_3 + i(-(\sin\varphi)x_2 + (\cos\theta)x_3))^k\\ &= (\cos\varphi - i\sin\varphi)x_2 + i(\cos\varphi - i\sin\varphi)x_3)^k\\ &= e^{-ik\varphi}(x_2 + ix_3)^k = e^{-ik\varphi}P(x),\end{aligned}$$

as desired. □

Proposition 14.3 also shows that the representations $\mathbf{R}_k\colon \mathbf{SO}(3) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(3))$ form a complete set of irreducible representations of $\mathbf{SO}(3)$.

14.3 Factorization of the Unit Quaternions Using Euler Angles

In order to obtain formulae for the matrix elements of the representations of $\mathbf{SU}(2)$ in terms of special functions, the Jacobi polynomials, it is necessary to understand how to express the unit quaternions in terms of Euler angles. The key fact is that there are three types of unit quaternions, $r_x(\varphi), r_y(\theta), r_z(\psi)$ that define rotations around the x-axis, y-axis, and z-axis, respectively, namely

$$r_x(\varphi/2) = \begin{pmatrix} e^{\frac{i\varphi}{2}} & 0\\ 0 & e^{-\frac{i\varphi}{2}}\end{pmatrix}, \quad r_y(\psi/2) = \begin{pmatrix} \cos\frac{\psi}{2} & -\sin\frac{\psi}{2}\\ \sin\frac{\psi}{2} & \cos\frac{\psi}{2}\end{pmatrix}, \quad r_z(\theta/2) = \begin{pmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2}\\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2}\end{pmatrix}.$$

We immediately check that the rotations corresponding to $r_x(\varphi/2)$, $r_y(\psi/2)$, $r_z(\theta/2)$ under the homomorphism $\rho\colon \mathbf{SU}(2) \to \mathbf{SO}(3)$(see Theorem 12.9) are given by the matrices

$$R_x(\varphi) = \begin{pmatrix} 1 & 0 & 0\\ 0 & \cos\varphi & -\sin\varphi\\ 0 & \sin\varphi & \cos\varphi\end{pmatrix}, \quad R_y(\psi) = \begin{pmatrix} \cos\psi & 0 & -\sin\psi\\ 0 & 1 & 0\\ \sin\psi & 0 & \cos\psi\end{pmatrix},$$
$$R_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}.$$

So $R_x(\varphi)$ is a rotation by φ around the x-axis (with the plane orthogonal to the x-axis oriented by (e_2, e_3, e_1)), $R_y(\psi)$ is a rotation by ψ around the $-y$-axis (with the plane orthogonal to the $-y$-axis oriented by $(e_1, e_3, -e_2)$, or equivalently a rotation by $-\psi$ around the y-axis with the plane orthogonal to the y-axis oriented by (e_3, e_1, e_2)), and $R_z(\theta)$ is a rotation by θ around the z-axis (with the plane orthogonal to the z-axis oriented by (e_1, e_2, e_3)).

Remark: Beware that a number of authors switch the roles of x and z, in particular Vilenkin [101]. As a consequence, the orientation of the plane normal to the y-axis is flipped. In this case, $R_x(\varphi)$ and $R_z(\varphi)$ are swapped, but $R_y(\psi)$ becomes $R_y(-\psi)$, which is a rotation by ψ around the y-axis (with the plane orthogonal to the y-axis oriented by (e_3, e_1, e_2)). Vilenkin denotes our matrices r_x, r_y, r_z as $\omega_3, \omega_2, \omega_1$.

Analogously to the factorization of rotation matrices in terms of the Euler angles, we will prove that every unit quaternion q can be written in the form

$$q = u(\varphi, \theta, \psi) = r_x(\varphi/2) r_z(\theta/2) r_x(\psi/2).$$

Multiplying out the above matrices we get

$$\begin{aligned} u(\varphi, \theta, \psi) &= \begin{pmatrix} e^{\frac{i\varphi}{2}} & 0 \\ 0 & e^{-\frac{i\varphi}{2}} \end{pmatrix} \begin{pmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2} \\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \begin{pmatrix} e^{\frac{i\psi}{2}} & 0 \\ 0 & e^{-\frac{i\psi}{2}} \end{pmatrix} \\ &= \begin{pmatrix} \cos\frac{\theta}{2}\, e^{\frac{i(\varphi+\psi)}{2}} & i\sin\frac{\theta}{2}\, e^{\frac{i(\varphi-\psi)}{2}} \\ i\sin\frac{\theta}{2}\, e^{-\frac{i(\varphi-\psi)}{2}} & \cos\frac{\theta}{2}\, e^{-\frac{i(\varphi+\psi)}{2}} \end{pmatrix}. \end{aligned}$$

The reader can reconfirm by inspection that $u(\varphi, \theta, \psi)^{-1} = u(\varphi, \theta, \psi)^*$.

Proposition 14.4. *Every unit quaternion*

$$q = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}, \quad \alpha, \beta \in \mathbb{C},\ |\alpha|^2 + |\beta|^2 = 1$$

can be expressed as

$$q = u(\varphi, \theta, \psi) = \begin{pmatrix} e^{\frac{i\varphi}{2}} & 0 \\ 0 & e^{-\frac{i\varphi}{2}} \end{pmatrix} \begin{pmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2} \\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \begin{pmatrix} e^{\frac{i\psi}{2}} & 0 \\ 0 & e^{-\frac{i\psi}{2}} \end{pmatrix}.$$

If $\beta = 0$, we can pick $\theta = 0$ and φ and ψ such that

$$\alpha = e^{i\frac{(\varphi+\psi)}{2}},$$

and in particular, $\psi = 0$. If $\alpha = 0$, we can pick $\theta = \pi$ and φ and ψ such that

$$\beta = e^{i\frac{(\varphi-\psi+\pi)}{2}},$$

and in particular, $\psi = \pi$. If $\alpha\beta \neq 0$ and if we require that

$$0 \leq \varphi < 2\pi, \quad 0 < \theta < \pi, \quad -2\pi \leq \psi < 2\pi,$$

then φ and ψ are unique. In this case,

$$\cos\theta = 2|\alpha|^2 - 1, \quad e^{i\varphi} = -\frac{\alpha\beta i}{|\alpha||\beta|}, \quad e^{\frac{i\psi}{2}} = \frac{\alpha}{|\alpha|}\, e^{-\frac{i\varphi}{2}}.$$

Proof. Since $|\alpha|^2+|\beta|^2=1$, we can write $\alpha = re^{i\omega}$ and $\beta = \sqrt{1-r^2}e^{i\sigma}$, with $0\leq r\leq 1$ and where ω and σ are defined modulo 2π. We will see shortly that it is convenient to assume that $0\leq \omega < 2\pi$ and $\frac{\pi}{2}\leq \sigma < \frac{5\pi}{2}$. The equation $q = u(\varphi,\theta,\psi)$ is equivalent to the two equations

$$\begin{aligned} re^{i\omega} &= \cos\frac{\theta}{2}\, e^{\frac{i(\varphi+\psi)}{2}} \\ \sqrt{1-r^2}e^{i\sigma} &= i\sin\frac{\theta}{2}\, e^{\frac{i(\varphi-\psi)}{2}} = \sin\frac{\theta}{2}\, e^{\frac{i(\varphi-\psi+\pi)}{2}}, \end{aligned}$$

since $i = e^{\frac{i\pi}{2}}$. If $r=1$, we pick $\theta = 0$ and then $e^{i\omega} = e^{\frac{i(\varphi+\psi)}{2}}$, so we choose φ,ψ so that $2\omega = \varphi+\psi$. If $r=0$, we pick $\theta=\pi$ and then $e^{i\sigma} = e^{\frac{i(\varphi-\psi+\pi)}{2}}$, so we choose φ,ψ so that $2\sigma = \varphi-\psi+\pi$. If $0<r<1$, namely $\alpha\beta\neq 0$, then there is a unique θ such that $0<\theta<\pi$ and $r = \cos\frac{\theta}{2}$, $\sqrt{1-r^2} = \sin\frac{\theta}{2}$. The angles φ and ψ must satisfy the equations

$$\begin{aligned} \omega + k_1 2\pi &= \frac{(\varphi+\psi)}{2} \\ \sigma + k_2 2\pi &= \frac{(\varphi-\psi+\pi)}{2}, \end{aligned}$$

with $k_1,k_2\in\mathbb{Z}$, and these are equivalent to the equations

$$\begin{aligned} \varphi &= \omega+\sigma-\frac{\pi}{2}+(k_1+k_2)2\pi \\ \psi &= \omega-\sigma+\frac{\pi}{2}+(k_1-k_2)2\pi, \end{aligned}$$

with $k_1,k_2\in\mathbb{Z}$. These equations always have solutions, but we would like to show that if we require that $0\leq\varphi<2\pi$ and $-2\pi\leq\psi<2\pi$, then φ and ψ are unique.

First, since $-\sigma+\frac{\pi}{2} = -(\sigma-\frac{\pi}{2})$, we let $\delta = \sigma-\frac{\pi}{2}$ so that the above equations become

$$\begin{aligned} \varphi &= \omega+\delta+(k_1+k_2)2\pi \\ \psi &= \omega-\delta+(k_1-k_2)2\pi, \end{aligned}$$

with $k_1,k_2\in\mathbb{Z}$, and we may assume that $0\leq\omega<2\pi$, $0\leq\delta<2\pi$. Since $0\leq\omega,\delta<2\pi$, we have $0\leq\omega+\delta<4\pi$.

If $\omega+\delta<2\pi$, since $\omega,\delta\geq 0$, we have $-2\pi<\omega-\delta<2\pi$, so we must pick $k_1=0$ and $k_2=0$ to make sure that $0\leq\varphi<2\pi$ and $-2\pi\leq\psi<2\pi$.

Let us now assume that $2\pi\leq\omega+\delta<4\pi$. Since $0\leq\omega<2\pi$, $0\leq\delta<2\pi$, we have $-2\pi<\omega-\delta<2\pi$.

Case 1. $\omega-\delta\geq 0$. Since $2\pi\leq\omega+\delta<4\pi$, by subtracting 2π we get $0\leq\omega+\delta-2\pi<2\pi$. This can be achieved by setting $k_1=-1$, $k_2=0$. Then since $\omega-\delta\geq 0$, we have

$$\omega-\delta-2\pi = \omega-\delta-2\pi \geq -2\pi.$$

Consequently,

$$\begin{aligned}\varphi &= \omega + \delta - 2\pi \\ \psi &= \omega - \delta - 2\pi\end{aligned}$$

satisfy the required conditions $0 \leq \varphi < 2\pi$ and $-2\pi \leq \psi < 2\pi$.

Case 2. $\omega - \delta < 0$. Since $2\pi \leq \omega + \delta < 4\pi$, by subtracting 2π we get $0 \leq \omega + \delta - 2\pi < 2\pi$. This can be achieved by setting $k_1 = 0$, $k_2 = -1$. Then since $\omega - \delta < 0$, we have

$$\omega - \delta + 2\pi = \omega - \delta + 2\pi < 2\pi.$$

Consequently,

$$\begin{aligned}\varphi &= \omega + \delta - 2\pi \\ \psi &= \omega - \delta + 2\pi\end{aligned}$$

satisfy the required conditions $0 \leq \varphi < 2\pi$ and $-2\pi \leq \psi < 2\pi$.

The last part is immediately verified. $\square$

An interesting corollary of Proposition 14.4 is the fact that every rotation matrix $Q \in \mathbf{SO}(3)$ can be written in the terms of the Euler angles as a product

$$Q = R_x(\varphi) R_z(\theta) R_x(\psi),$$

namely

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix}.$$

But in this case, we may assume that $0 \leq \psi < 2\pi$. This is because both q and $-q$ define the same rotation ρ_q, but since $e^{i\pi} = e^{-i\pi} = -1$, we have $-r_x(\psi/2) = r_x(\frac{\psi + 2\pi}{2})$, so if $-2\pi \leq \psi < 0$, then $0 \leq \psi + 2\pi < 2\pi$ and $Q = R_x(\varphi) R_z(\theta) R_x(\psi + 2\pi)$.

One might wonder what happens if we make the bold move of replacing the *real* angle parameters φ, θ, ψ by *arbitrary* complex numbers? This certainly makes sense since the complex power series $z \mapsto e^z, z \mapsto \cos z, \mapsto \sin z$ are perfectly well-defined. We see immediately that $\det(u(\varphi, \theta, \psi)) = 1$, so these complex matrices belong to $\mathbf{SL}(2, \mathbb{C})$. Remarkably, *every* matrix $A \in \mathbf{SL}(2, \mathbb{C})$ can be expressed as $A = u(\varphi, \theta, \psi)$ for some choice of complex numbers φ, θ, ψ. We also have uniqueness of the representation if $\varphi, \theta, \psi \in \mathbb{C}$ satisfy the conditions

$$0 < \Re(\theta) < \pi, \quad 0 \leq \Re(\varphi) < 2\pi, \quad -2\pi \leq \Re(\psi) < 2\pi.$$

See Vilenkin [101] (Chapter III, Section 1.4). In some sense, the above fact illustrates the fact that $\mathbf{SL}(2, \mathbb{C})$ is the complexification of $\mathbf{SU}(2)$.

14.4 Dehomogenized Representations of $\mathbf{SL}(2,\mathbb{C})$ and $\mathbf{SU}(2)$

In Example 12.7 we defined the irreducible representations $U_m\colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_m^{\mathbb{C}}(2))$ of $\mathbf{SU}(2)$ whose representing spaces are the vector spaces $\mathcal{P}_m^{\mathbb{C}}(2)$ of homogeneous polynomials in two variables. We also said that it is customary, especially in the physics literature, to index homogeneous polynomials in terms of $\ell = m/2$, which is an integer when m is even but a half integer when m is odd. In this context, in terms of $\ell = m/2$, a homogeneous polynomial is written as

$$P(z_1, z_2) = \sum_{k=-\ell}^{\ell} c_k z_1^{\ell-k} z_2^{\ell+k},$$

where it is assumed that $\ell + k = j$ where j takes the *integral* values $j = 0, 1, \ldots, 2\ell = m$, so that $\ell - k = 2\ell - (\ell + k) = 2\ell - j$ takes the values $2\ell, 2\ell - 1, \ldots, 0$. Note that $k = j - \ell = j - m/2$ with $j = 0, 1, \ldots, 2\ell = m$, so k is an integer only if m is even. If m is odd, say $m = 2h + 1$, then $\ell = h + \frac{1}{2}$ and k takes the $2\ell + 1 = m + 1$ values

$$-h - \frac{1}{2},\ -(h-1) - \frac{1}{2}, \ldots, -\frac{1}{2},\ \frac{1}{2},\ 1 + \frac{1}{2}, \ldots, h + \frac{1}{2},$$

and so $k \neq 0$. If m is even, say $m = 2h$, then $\ell = h$ and k takes the $2\ell + 1 = m + 1$ values

$$-h,\ -(h-1), \ldots, -1,\ 0,\ 1, \ldots, h-1, h.$$

For example, if $\ell = \frac{3}{2}$, then k takes the four values

$$-\frac{3}{2},\ -\frac{1}{2},\ \frac{1}{2},\ \frac{3}{2},$$

and if $\ell = 2$, then k takes the five values

$$-2,\ -1,\ 0,\ 1,\ 2.$$

The representing space is then $\mathcal{P}_{2\ell}^{\mathbb{C}}(2)$ and it has dimension $2\ell + 1$. Using the standard technique of "dehomogenizing" and "homogenizing" we can use the space of complex polynomials of degree $2\ell + 1$ in *one* variable z instead of the space $\mathcal{P}_{2\ell}^{\mathbb{C}}(2)$ of homogeneous polynomials in two variables z_1, z_2. Given a homogeneous polynomial $P(z_1, z_2)$ of degree $m = 2\ell$, by dehomogenizing we obtain the polynomial $Q(z)$ of degree $m = 2\ell$ given by

$$Q(z) = P(z, 1). \tag{dehomog}$$

So given

$$P(z_1, z_2) = \sum_{k=-\ell}^{\ell} c_k z_1^{\ell-k} z_2^{\ell+k},$$

we obtain

$$Q(z) = \sum_{k=-\ell}^{\ell} c_k z^{\ell-k}. \tag{Q}$$

Observe that due to our indexing scheme, the coefficients of Q have "funny" indices. For example, for $\ell = 2$, so that $m = 2\ell = 4$,

$$Q(z) = c_{-2}z^4 + c_{-1}z^3 + c_0 z^2 + c_1 z + c_2,$$

and when $\ell = 5/2$, so that $m = 2\ell = 5$, we have

$$Q(z) = c_{-5/2}z^5 + c_{-3/2}z^4 + c_{-1/2}z^3 + c_{1/2}z^2 + c_{3/2}z + c_{5/2}.$$

Conversely, given a polynomial $Q(z)$ of degree $m = 2\ell$, by homogenizing we obtain the homogeneous polynomial $P(z_1, z_2)$ of degree $m = 2\ell$ given by

$$P(z_1, z_2) = z_2^{2\ell} Q\left(\frac{z_1}{z_2}\right). \tag{homog}$$

Definition 14.2. Following Vilenkin, we denote the space of polynomials of degree $m = 2\ell$ with complex coefficients in one variable by $\mathcal{P}_\ell^{\mathbb{C}}$.

Note that the "funny" index ℓ is a half-integer when m is odd. We can convert our representations $U_m\colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_m^{\mathbb{C}}(2))$ to representations in the spaces $\mathcal{P}_\ell^{\mathbb{C}}$. Actually, until we use the fact that $\mathbf{SU}(2)$ is compact, we consider representations of $\mathbf{SL}(2,\mathbb{C})$.

Definition 14.3. Given any matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad a, b, c, d \in \mathbb{C},\ ad - bc = 1,$$

in $\mathbf{SL}(2,\mathbb{C})$, for any polynomial $Q \in \mathcal{P}_\ell^{\mathbb{C}}$, define $T_\ell(A)(Q(z))$ by

$$T_\ell(A)(Q(z)) = (bz + d)^{2\ell} Q\left(\frac{az + c}{bz + d}\right). \tag{T_ℓ}$$

It is immediately verified that the above formula yields a representation $T_\ell\colon \mathbf{SL}(2,\mathbb{C}) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$ which yields a representation $T_\ell\colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$ when restricted to the subgroup $\mathbf{SU}(2)$ of $\mathbf{SL}(2,\mathbb{C})$.

Note that the above formula for $T_\ell(A)(Q(z))$ is *not* what we would obtain directly from the representation U_ℓ. We are using Vilenkin's formula to facilitate comparison with his exposition; see Vilenkin [101] (Chapter III, Section 2.1) and Kosmann-Schwarzbach [59]. With our version we define the representations T_ℓ as

$$T_\ell(A)(Q(z)) = (-cz + a)^{2\ell} Q\left(\frac{dz - b}{-cz + a}\right).$$

In its homogeneous form, Vilenkin's version of the representation U_ℓ is

$$U_\ell^v(A)(Q(z_1, z_2)) = Q(az_1 + cz_2, bz_1 + dz_2).$$

Observe that

$$\begin{pmatrix} az_1 + cz_2 \\ bz_1 + dz_2 \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = A^\top \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

but in our case

$$\begin{pmatrix} dz_1 - bz_2 \\ -cz_1 + az_2 \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = A^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

We immediately check that if

$$Y = \begin{pmatrix} b & d \\ -a & -c \end{pmatrix},$$

then

$$YA^\top = A^{-1}Y,$$

and $\det(Y) = ad - bc = 1$. Then Y defines a linear isomorphism of $\mathcal{P}_{2\ell}^{\mathbb{C}}(2)$ given by $Q(z_1, z_2) \mapsto Q(bz_1 + dz_2, -az_1 - cz_2)$, and this map is an equivalence between the representations U_ℓ and U_ℓ^v (we leave the details as an exercise). We also leave it as an exercise (using the dehomogenization and the homogenization maps, which are linear isomorphisms) to check that the representation $U_{2\ell} \colon \mathbf{SL}(2, \mathbb{C}) \to \mathbf{GL}(\mathcal{P}_{2\ell}^{\mathbb{C}}(2))$ is equivalent to the representation $T_\ell \colon \mathbf{SL}(2, \mathbb{C}) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$ and similarly the representation $U_{2\ell} \colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_{2\ell}^{\mathbb{C}}(2))$ is equivalent to the representation $T_\ell \colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$. In particular, the representations $T_\ell \colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$ form a complete set of irreducible representations of $\mathbf{SU}(2)$.

14.5 The Lie Algebra Representation Associated with T_ℓ

We will need to define an $\mathbf{SU}(2)$-invariant hermitian inner product on each space $\mathcal{P}_\ell^{\mathbb{C}}$, and for this it is useful to figure out what is the derivative of the representation $T_\ell \colon \mathbf{SL}(2, \mathbb{C}) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$ at the identity. This yields a representation $\mathfrak{t}_\ell \colon \mathfrak{sl}(2, \mathbb{C}) \to \mathrm{Hom}(\mathcal{P}_\ell^{\mathbb{C}}, \mathcal{P}_\ell^{\mathbb{C}})$, which is a representation of Lie algebras! Following Kosmann-Schwarzbach [59] (Problem 9), we use the standard technique of "passing a curve" though the identity whose tangent vector for $t = 0$ is a vector in the tangent space. So for any tangent vector $X \in \mathfrak{sl}(2, \mathbb{C})$,

$$X = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad \alpha + \delta = 0,$$

we consider the curve

$$C(t) = e^{tX} = \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix}$$

through I_2, such that $C'(0) = X$, and by the chain rule we have

$$\begin{aligned}(\mathfrak{t}_\ell(X))(Q(z)) = (d(T_\ell)_I(X))(Q(z)) &= \frac{d}{dt}\Big(T_\ell(C(t))(Q(z))\Big)\Big|_{t=0}\\ &= \frac{d}{dt}\left((b(t)z+d(t))^{2\ell}Q\left(\frac{a(t)z+c(t)}{b(t)z+d(t)}\right)\right)\Bigg|_{t=0}.\end{aligned}$$

We have

$$\frac{d}{dt}\left(\frac{a(t)z+c(t)}{b(t)z+d(t)}\right) = \frac{(a'(t)z+c'(t))(b(t)z+d(t)) - (a(t)z+c(t))(b'(t)z+d'(t))}{(b(t)z+d(t))^2},$$

and since $a(0) = d(0) = 1$, $b(0) = c(0) = 0$, $a'(0) = \alpha$, $b'(0) = \beta$, $c'(0) = \gamma$, $d'(0) = \delta$,

$$\frac{d}{dt}\left(\frac{a(t)z+c(t)}{b(t)z+d(t)}\right)\Bigg|_{t=0} = \alpha z + \gamma - z(\beta z + \delta) = -\beta z^2 + (\alpha - \delta)z + \gamma,$$

so we obtain

$$(\mathfrak{t}_\ell(X))(Q(z)) = 2\ell(\beta z + \delta)Q(z) + (-\beta z^2 + (\alpha-\delta)z + \gamma)\frac{d}{dz}(Q(z)),$$

which can be written as

$$\mathfrak{t}_\ell(X) = 2\ell(\beta z + \delta) + (-\beta z^2 + (\alpha-\delta)z + \gamma)\frac{d}{dz},$$

viewed as a differential operator on polynomials $Q(z)$ in z. In summary we obtained the following result.

Proposition 14.5. *For any representation $T_\ell\colon \mathbf{SL}(2,\mathbb{C}) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$, the derivative $\mathfrak{t}_\ell = d(T_\ell)_I$ of T_ℓ at the identity is the representation $\mathfrak{t}_\ell\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(\mathcal{P}_\ell^{\mathbb{C}}, \mathcal{P}_\ell^{\mathbb{C}})$ given by*

$$\mathfrak{t}_\ell(X) = 2\ell(\beta z + \delta) + (-\beta z^2 + (\alpha-\delta)z + \gamma)\frac{d}{dz} \tag{$\mathfrak{t}_\ell$}$$

viewed as a differential operator on polynomials $Q(z)$ in z, for any

$$X = \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} \in \mathfrak{sl}(2,\mathbb{C}),$$

Now $\mathfrak{su}(2)$ is the *real* vector space consisting of skew-hermitian matrices with zero trace, which are of the form

$$X = \begin{pmatrix} iu_1 & u_2 + iu_3\\ -u_2 + iu_3 & -iu_1\end{pmatrix}, \quad u_1, u_2, u_3 \in \mathbb{R},$$

and a basis (of course, over $\mathbb{R}$) is given by the matrices

$$\begin{pmatrix} i & 0\\ 0 & -i\end{pmatrix}, \quad \begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix}, \quad \begin{pmatrix} 0 & i\\ i & 0\end{pmatrix}.$$

The exponential map $\exp\colon \mathfrak{su}(2) \to \mathbf{SU}(2)$ is very nicely expressed. For any $X \in \mathfrak{su}(2)$ given by

$$X = \begin{pmatrix} iu_1 & u_2 + iu_3 \\ -u_2 + iu_3 & -iu_1 \end{pmatrix},$$

if we write $\theta = \sqrt{u_1^2 + u_2^2 + u_3^2}$, then

$$e^X = \cos\theta I + \frac{\sin\theta}{\theta} X, \quad \theta \neq 0,$$

and $e^0 = I$. It is not hard to prove that the map $\exp\colon \mathfrak{su}(2) \to \mathbf{SU}(2)$ is surjective. See Gallier and Quaintance [40] (Section 15.5).

The Lie algebra $\mathfrak{sl}(2, \mathbb{C})$ is the complex vector space consisting of all complex 2×2 matrices with zero trace, and since the above three matrices are linearly independent over $\mathbb{C}$, they also form a basis of $\mathfrak{sl}(2, \mathbb{C})$. However, for the sake of consistency with other sources, especially Kosmann-Schwarzbach and Vilenkin, it is preferable to use the basis denoted (ξ_1, ξ_2, ξ_3) in Kosmann-Schwarzbach [59] (Chapter 5, Section 1).

Definition 14.4. The *basis* (ξ_1, ξ_2, ξ_3) of $\mathfrak{sl}(2, \mathbb{C})$ (over $\mathbb{C}$), which is also a basis of $\mathfrak{su}(2)$ (over $\mathbb{R}$), is given by

$$\xi_3 = \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \xi_2 = \frac{1}{2}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \xi_1 = \frac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

It is denoted (a_1, a_2, a_3) in Vilenkin [101] (Chapter III, Section 1.3).

The basis (ξ_1, ξ_2, ξ_3) has the advantage that

$$e^{\varphi\xi_3} = r_x(\varphi/2), \quad e^{\theta\xi_2} = r_y(\theta/2), \quad e^{\psi\xi_1} = r_z(\psi/2).$$

If we pick the following basis for $\mathfrak{so}(3)$,

$$E_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad E_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix},$$

then we easily check that

$$e^{\varphi E_3} = R_x(\varphi), \quad e^{\theta E_2} = R_y(\theta), \quad e^{\psi E_1} = R_z(\psi).$$

Remark: The swap between ξ_1 and ξ_3 has to do with fact that Vilenkin and Kosmann-Schwarzbach swap x and z.

It is useful to obtain formulae for the action of $\mathfrak{t}_\ell$ on the basis (ξ_1, ξ_2, ξ_3) of $\mathfrak{sl}(2, \mathbb{C})$. Using Formula $(\mathfrak{t}_\ell)$ we obtain

$$\mathfrak{t}_\ell(\xi_1) = i\ell z + \frac{i}{2}(1 - z^2)\frac{d}{dz} \tag{$\mathfrak{t}1$}$$
$$\mathfrak{t}_\ell(\xi_2) = -\ell z + \frac{1}{2}(1 + z^2)\frac{d}{dz} \tag{$\mathfrak{t}2$}$$
$$\mathfrak{t}_\ell(\xi_3) = i\left(z\frac{d}{dz} - \ell\right). \tag{$\mathfrak{t}3$}$$

It is instructive to see what is the action of the above operators on the basis of $\mathcal{P}_\ell^{\mathbb{C}}$ consisting of the $2\ell + 1$ polynomials $z^{\ell-k}$, $k = -\ell, -\ell+1, \ldots, +\ell$. We obtain

$$\mathfrak{t}_\ell(\xi_1)z^{\ell-k} = \frac{i}{2}(\ell - k)z^{\ell-k-1} + \frac{i}{2}(\ell + k)z^{\ell-k+1} \tag{$\mathfrak{t}4$}$$
$$\mathfrak{t}_\ell(\xi_2)z^{\ell-k} = \frac{1}{2}(\ell - k)z^{\ell-k-1} - \frac{1}{2}(\ell + k)z^{\ell-k+1} \tag{$\mathfrak{t}5$}$$
$$\mathfrak{t}_\ell(\xi_3)z^{\ell-k} = -ikz^{\ell-k}. \tag{$\mathfrak{t}6$}$$

These formulae can be made more revealing by introducing the linear maps H_+, H_-, H_3 on $\mathfrak{sl}(2, \mathbb{C})$ given by

$$H_+ = i\mathfrak{t}_\ell(\xi_1) - \mathfrak{t}_\ell(\xi_2) = -\frac{d}{dz} \tag{$\mathfrak{t}7$}$$
$$H_- = i\mathfrak{t}_\ell(\xi_1) + \mathfrak{t}_\ell(\xi_2) = -2\ell z + z^2\frac{d}{dz} \tag{$\mathfrak{t}8$}$$
$$H_3 = i\mathfrak{t}_\ell(\xi_3) = \ell - z\frac{d}{dz}. \tag{$\mathfrak{t}9$}$$

Remark: Kosmann-Schwarzbach defines J_+, J_-, J_3 as

$$J_+ = i\xi_1 - \xi_2 = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}, \quad J_- = i\xi_1 + \xi_2 = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}, \quad J_3 = i\xi_3 = \frac{1}{2}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

and so $H_+ = \mathfrak{t}_\ell(J_+), H_- = \mathfrak{t}_\ell(J_-), H_3 = \mathfrak{t}_\ell(J_3)$. In quantum physics, the linear operator H_3 on $\mathfrak{sl}(2, \mathbb{C})$ is an observable. Another notation found in the literature, for example Dieudonné [21] (Chapter XXI, Section 9) is $X_+ = -J_-, X_- = -J_+, H = -2J_3$, that is,

$$X_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad X_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In Serre [92] (Chapter IV), X_+ is denoted X and X_- is denoted Y.

Using the formulae $(\mathfrak{t}7), (\mathfrak{t}8), (\mathfrak{t}9)$, we obtain

$$H_+z^{\ell-k} = -(\ell - k)z^{\ell-k-1} \tag{H_+}$$
$$H_-z^{\ell-k} = -(\ell + k)z^{\ell-k+1} \tag{H_-}$$
$$H_3z^{\ell-k} = kz^{\ell-k}. \tag{H_3}$$

In all of the above formulae, recall that $k = -\ell, -\ell+1, \ldots, +\ell$.

The above formulae show the following interesting facts:

(1) The polynomial $z^{\ell-k}$ is an eigenvector of H_3 for the eigenvalue k.

(2) The linear map H_+ send $z^{\ell-k}$ to an eigenvector of H_3 for the eigenvalue $k+1$. In particular, when $k = \ell$, $H_+(1)$ is the zero polynomial.

(3) The linear map H_- send $z^{\ell-k}$ to an eigenvector of H_3 for the eigenvalue $k-1$. In particular, when $k = -\ell$, $H_-(z^{2\ell})$ is the zero polynomial.

The above facts can be used to prove that the representation $\mathfrak{t}_\ell$ of $\mathfrak{sl}(2,\mathbb{C})$ is irreducible; see Section 14.6. Then it can be shown that the representation T_ℓ of $\mathbf{SL}(2,\mathbb{C})$ and its subgroup $\mathbf{SU}(2)$ is also irreducible.

Remark: Another interesting linear operator on $\mathfrak{sl}(2,\mathbb{C})$ is the operator traditionally denoted J^2 given by

$$J^2 = \mathfrak{t}_\ell(i\xi_1)^2 + \mathfrak{t}_\ell(i\xi_2)^2 + \mathfrak{t}_\ell(i\xi_3)^2 = -(\mathfrak{t}_\ell(\xi_1)^2 + \mathfrak{t}_\ell(\xi_2)^2 + \mathfrak{t}_\ell(\xi_3)^2).$$

It is easy to see that

$$J^2 = H_+H_- + H_3(H_3 - I) = H_-H_+ + H_3(H_3 + I).$$

Using the formulae above, we can check that

$$J^2(z^{\ell-k}) = \ell(\ell+1)z^{\ell-k}.$$

Thus $\ell(\ell+1)$ is a common eigenvalue for all basis vectors $z^{\ell-k}$. In some sense the operator J^2 behave like a Laplacian. It is called the *Casimir operator* of the representation $\mathfrak{t}_\ell$. In quantum physics it an observable of the angular momentum.

14.6 Irreducible Lie Algebra Representations of $\mathfrak{sl}(2,\mathbb{C})$ and $\mathfrak{su}(2)$

This section assumes some background of Lie algebras and Lie groups. Elementary presentations are found in Carter, Segal and Macdonald [17], Hall [42], and Gallier and Quaintance [38]. More advanced treatments are given in Dieudonné [21], Duistermaat and Kolk [29], Fulton and Harris [36], Hall [42], Helgason [47], Humphreys [51], Knapp [57, 56], Samelson [82], Serre [92, 91], and Varadarajan [98].

In this section we determine all the irreducible Lie algebra representations of $\mathfrak{sl}(2,\mathbb{C})$. The Lie algebra $\mathfrak{sl}(2,\mathbb{C})$ is a *simple* (complex) Lie algebra, which means that it is not abelian and that its only ideals are $\{0\}$ and $\mathfrak{sl}(2,\mathbb{C})$ itself. One of the most beautiful result of Lie

theory is that the complex simple(!) Lie algebras fall into four infinite families plus five exceptional simple Lie algebras. Furthermore, the irreducible representations of the simple Lie algebras can be completely determined. These results are presented in Fulton and Harris [36] and Knapp [57] among other sources. The determination of the irreducible Lie algebra representations of $\mathfrak{sl}(2,\mathbb{C})$ is a "miniature" case.

As a basis of $\mathfrak{sl}(2,\mathbb{C})$, it is convenient to use the basis (X,Y,H) of Section 14.5, namely

$$X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We immediately find the equations

$$[X,Y] = H, \quad [H,X] = 2X, \quad [H,Y] = -2Y.$$

(The Lie bracket $[A,B]$ of two square matrices A and B is defined as $[A,B] = AB - BA$.)

Since we never actually defined Lie algebra representations we recall the definition below.

Definition 14.5. Let K denote the field $K = \mathbb{R}$ or $K = \mathbb{C}$ and let $\mathfrak{g}$ be a Lie algebra. If $\mathfrak{g}$ is a real Lie algebra, then a *Lie algebra representation* of $\mathfrak{g}$ in a K-vector space V is a $\mathbb{R}$-linear map $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V,V)$, which means that $\rho(\lambda X) = \lambda\rho(X)$ for all $\lambda \in \mathbb{R}$ and all $X \in \mathfrak{g}$. If $\mathfrak{g}$ is a complex Lie algebra, then a *Lie algebra representation* of $\mathfrak{g}$ in a $\mathbb{C}$-vector space V is a $\mathbb{C}$-linear map $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V,V)$, which means that $\rho(\lambda X) = \lambda\rho(X)$ for all $\lambda \in \mathbb{C}$ and all $X \in \mathfrak{g}$. In both cases, ρ also has the property

$$\rho([X,Y])(v) = \rho(X)(\rho(Y)(v)) - \rho(Y)(\rho(X)(v)), \quad X,Y \in \mathfrak{g},\ v \in V. \tag{$[-,-]$}$$

When no confusion arises, $\rho(X)(v)$ is abbreviated as $X \cdot v$. With this convention the above equation is written

$$[X,Y]\cdot v = X\cdot(Y\cdot v) - Y\cdot(X\cdot v), \quad X,Y \in \mathfrak{g},\ v \in V. \tag{$[-,-]$}$$

It should be noted that if $\mathfrak{g}$ is a real Lie algebra and if V is a complex vector space, then the linear maps $\rho(X)\colon V \to V$ are $\mathbb{C}$-linear.

Definition 14.6. A representation $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V,V)$ is *irreducible* if $V \neq \{0\}$ and if the only subspaces W of V invariant under $\rho(X)$ for all $X \in \mathfrak{g}$ are $W = \{0\}$, and $W = V$. Note that if V is a *complex* space, then W is also a *complex* subspace of V.

The notion of map of Lie algebra representations is essentially the same as in the case of groups (see Definition 12.3).

Definition 14.7. Given any two representations $\rho_1\colon \mathfrak{g} \to \mathrm{Hom}(V_1,V_1)$ and $\rho_2\colon \mathfrak{g} \to \mathrm{Hom}(V_2,V_2)$ of a Lie algebra $\mathfrak{g}$, a *map (or morphism) of representations* $\varphi\colon \rho_1 \to \rho_2$ is a linear map

$\varphi\colon V_1 \to V_2$ which is *equivariant*, which means that the following diagram commutes for every $X \in \mathfrak{g}$:

$$\begin{tikzcd} V_1 \arrow[r, "\rho_1(X)"] \arrow[d, "\varphi"'] & V_1 \arrow[d, "\varphi"] \\ V_2 \arrow[r, "\rho_2(X)"'] & V_2, \end{tikzcd}$$

i.e.

$$\varphi \circ \rho_1(X) = \rho_2(X) \circ \varphi, \qquad X \in \mathfrak{g}.$$

The space of all maps between two representations as above is denoted $\mathrm{Hom}_{\mathfrak{g}}(\rho_1, \rho_2)$. Two representations $\rho_1\colon \mathfrak{g} \to \mathrm{Hom}(V_1, V_1)$ and $\rho_2\colon \mathfrak{g} \to \mathrm{Hom}(V_2, V_2)$ are *equivalent* iff $\varphi\colon V_1 \to V_2$ is an invertible linear map.

It should be noted that the map $\varphi\colon V_1 \to V_2$ is $\mathbb{R}$-linear if both V_1 and V_2 are real vector spaces (in which case $\mathfrak{g}$ is a real Lie algebra), and $\mathbb{C}$-linear if both V_1 and V_2 are complex vector spaces (in which case $\mathfrak{g}$ is a real or a complex Lie algebra).

As in Section 17.3, given a real Lie algebra $\mathfrak{g}$ we can construct its complexification $\mathfrak{g}_{\mathbb{C}}$, which is the complex Lie algebra whose carrier is the complex vector space $\mathfrak{g} \oplus i\mathfrak{g}$ as a direct sum of real subspaces (technically, $(\mathfrak{g}_{\mathbb{C}})|_{\mathbb{R}} = \mathfrak{g} \oplus i\mathfrak{g}$, see the beginning of Section 17.3) with the Lie bracket given by

$$[u + iv, x + iy]_{\mathbb{C}} = [u, x] - [v, y] + i([u, y] + [v, x]).$$

Then for any representation $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V, V)$ with V a complex vector space, we obtain the complex representation $\rho_{\mathbb{C}}\colon \mathfrak{g}_{\mathbb{C}} \to \mathrm{Hom}(V, V)$ given by

$$\rho_{\mathbb{C}}(X + iY) = \rho(X) + i\rho(Y), \quad X, Y \in \mathfrak{g}.$$

Since $\rho(X)\colon V \to V$ and $\rho(Y)\colon V \to V$ are $\mathbb{C}$-linear maps of the complex vector space V, $i\rho(Y)\colon V \to V$ is also a $\mathbb{C}$-linear map, and so $\rho_{\mathbb{C}}(X + iY)$ makes sense. Observe that the restriction of $\rho_{\mathbb{C}}$ to $\mathfrak{g}$ is the original representation $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V, V)$.

We have the following useful result which shows that for a real Lie algebra $\mathfrak{g}$ and a complex vector space V, the study of the representations $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V, V)$ is equivalent to the study of the complex representations $\rho_{\mathbb{C}}\colon \mathfrak{g}_{\mathbb{C}} \to \mathrm{Hom}(V, V)$.

Proposition 14.6. *Let $\mathfrak{g}$ be a real Lie algebra and let V be a complex vector space V. There is a bijection between the set of representations $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V, V)$ of $\mathfrak{g}$ in V and the set of representations $\rho'\colon \mathfrak{g}_{\mathbb{C}} \to \mathrm{Hom}(V, V)$ of $\mathfrak{g}_{\mathbb{C}}$ in V given by the map $\rho \mapsto \rho_{\mathbb{C}}$, whose inverse is the restriction of ρ' to $\mathfrak{g}$. The representation $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V, V)$ is irreducible iff the representation $\rho_{\mathbb{C}}\colon \mathfrak{g}_{\mathbb{C}} \to \mathrm{Hom}(V, V)$ is irreducible.*

Proof. We already explained the reason for the bijection. Suppose that $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V,V)$ is irreducible, and let W be any subspace of V invariant under $\rho_{\mathbb{C}}(X+iY)$ for all $X,Y \in \mathfrak{g}$. Then by setting $Y=0$, the subspace W is invariant under $\rho(X)$ for all $X \in \mathfrak{g}$, and since $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V,V)$ is irreducible, we must have $W=\{0\}$ or $W=V$, so $\rho_{\mathbb{C}}$ is also irreducible.

Let us now assume that $\rho_{\mathbb{C}}\colon \mathfrak{g}_{\mathbb{C}} \to \mathrm{Hom}(V,V)$ is irreducible and let W be any subspace of V invariant under $\rho(X)$ for all $X \in \mathfrak{g}$. Since W is a complex subspace, we have $\rho_{\mathbb{C}}(X+iY) = \rho(X)+i\rho(Y) \in W$ for all $X,Y \in \mathfrak{g}$, and since $\rho_{\mathbb{C}}\colon \mathfrak{g}_{\mathbb{C}} \to \mathrm{Hom}(V,V)$ is irreducible, must have $W=\{0\}$ or $W=V$, so ρ is also irreducible. $\square$

Proposition 14.6 applies to the real Lie algebra $\mathfrak{su}(2)$. Indeed, the matrices

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

are a basis (over $\mathbb{R}$) of $\mathfrak{su}(2)$ and also a basis (over $\mathbb{C}$) of $\mathfrak{sl}(2,\mathbb{C})$, and it can be shown that

$$\mathfrak{sl}(2,\mathbb{C}) = \mathfrak{su}(2) \oplus i\mathfrak{su}(2)$$

as a direct sum of real subspaces; see Example 17.1 for details. Therefore $\mathfrak{sl}(2,\mathbb{C})$ is the complexification of $\mathfrak{su}(2)$, and by Proposition 14.6, the irreducible representations of $\mathfrak{su}(2)$ in a complex vector space V are in bijection with the irreducible representations of $\mathfrak{sl}(2,\mathbb{C})$ in V.

We now consider the complex Lie algebra $\mathfrak{sl}(2,\mathbb{C})$. Let $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ be any representation of $\mathfrak{sl}(2,\mathbb{C})$ with V of finite dimension $m+1$. Since $\rho(H)$ is a linear map over a complex vector space of finite dimension $m+1$, it has $m+1$ complex eigenvalues (counted with their multiplicities). For every eigenvalue λ of $\rho(H)$, let V^{λ} be the corresponding eigenspace. In the context of Lie algebras, λ is called a *weight*. It turns our that $\rho(H)$ is diagonalizable, but we will not need this fact to characterize the irreducible representations of $\mathfrak{sl}(2,\mathbb{C})$.

The first important property is this.

Proposition 14.7. *For any complex representation $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ with V finite-dimensional, for any eigenvalue λ of $\rho(H)$ and any vector $v \in V^{\lambda}$, we have*

$$H\cdot(X\cdot v) = (\lambda+2)X\cdot v, \quad H\cdot(Y\cdot v) = (\lambda-2)Y\cdot v. \tag{V1}$$

Consequently $X\colon V^{\lambda} \to V^{\lambda+2}$ and $Y\colon V^{\lambda} \to V^{\lambda-2}$.

Proof. Since $HX - XH = [H,X] = 2X$ and v is an eigenvector of $\rho(H)$ for λ, we get

$$H\cdot(X\cdot v) = [H,X]\cdot v + X\cdot(H\cdot v) = 2X\cdot v + X\cdot \lambda v = (\lambda+2)X\cdot v.$$

Similarly, since $HY - YH = [H,Y] = -2Y$,

$$H\cdot(Y\cdot v) = [H,Y]\cdot v + Y\cdot(H\cdot v) = -2Y\cdot v + Y\cdot \lambda v = (\lambda-2)Y\cdot v,$$

as claimed. $\square$

Now let $z \neq 0$ be some vector $z \in V^{\lambda}$, for some eigenvalue λ of $\rho(H)$. Consider the sequence

$$z,\ X \cdot z,\ X^2 \cdot z, \ldots, X^n \cdot z, \ldots.$$

By Proposition 14.7, $X^n \cdot z \in V^{\lambda+2n}$. The nonzero vectors of the form $X^n \cdot z$ correspond to distinct eigenvalues $\lambda + 2n$ of $\rho(H)$ so they are linearly independent. But V is finite-dimensional, so there is a smallest n such that $X^{n+1} \cdot z = 0$, and if we let $x = X^n \cdot z$, then $x \neq 0$, $X \cdot x = 0$, and $Hx = (\lambda + 2n)x$. This suggests the following definition.

Definition 14.8. Let $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ be complex representation of $\mathfrak{g}$ with V finite-dimensional. A *nonzero* vector $e \in V$ is *primitive of weight* $\lambda \in \mathbb{C}$ if

$$Xe = 0, \quad He = \lambda e. \tag{V2}$$

The argument just before Definition 14.8 proved the following result.

Proposition 14.8. *Given any complex representation $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ with V finite-dimensional, there is some primitive element $e \in V$ for some weight λ.*

A priori, λ is a complex number, but in fact we will prove that it is a nonnegative integer. The next proposition is the key result.

Proposition 14.9. *Given any complex representation $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ with V finite-dimensional, consider any primitive element $e \in V$ of weight λ. Define the sequence $(e_n)_{n \geq -1}$ defined as follows:*

$$e_n = (1/n!)Y^n \cdot e, \quad n \geq 0, \tag{V3}$$

with $e_{-1} = 0$). Then the following properties hold:

$$\begin{aligned} H \cdot e_n &= (\lambda - 2n)e_n & \text{(V4)}\\ Y \cdot e_n &= (n+1)e_{n+1} & \text{(V5)}\\ X \cdot e_n &= (\lambda - n + 1)e_{n-1}. & \text{(V6)} \end{aligned}$$

There is a smallest $m \geq 0$ such that $e_{m+1} = 0$ and $(e_0, \ldots, e_m)$ are linearly independent. (we also have $e_n = 0$ for all $n \geq m+1$). The weight λ is a nonnegative integer, namely $\lambda = m$, and $e_n \in V^{m-2n}$ for $n = 0, \ldots, m$. See the diagram below.

$$\begin{tikzcd}
V^{-m} \arrow[loop below, "H"] \arrow[r, shift left, "X"] & V^{-(m-2)} \arrow[loop below, "H"] \arrow[l, shift left, "Y"] \arrow[r, shift left, "X"] & \cdots \arrow[l, shift left, "Y"] \arrow[r, shift left, "X"] & V^{m-2} \arrow[loop below, "H"] \arrow[l, shift left, "Y"] \arrow[r, shift left, "X"] & V^{m} \arrow[loop below, "H"] \arrow[l, shift left, "Y"]
\end{tikzcd}$$

Proof. Equation (V5) follows by the definition of e_n since

$$Y \cdot e_n = (1/n!)Y \cdot Y^n \cdot e = (n+1)(1/((n+1)!)Y^{n+1} \cdot e = (n+1)e_{n+1}.$$

Equation (V4) is proven by induction. For the base case $n = 0$, since $e_0 = e$ is a primitive element of weight λ, we have $H \cdot e_0 = H \cdot e = \lambda e = \lambda e_0$.

For the induction step, since by the induction hypothesis, $H \cdot e_n = (\lambda - 2n)e_n$, from the second equation of Proposition 14.7 with $\lambda - 2n$ instead of λ, we get

$$H \cdot (Y \cdot e_n) = (\lambda - 2n - 2) \cdot (Y \cdot e_n),$$

and by (V5), we obtain

$$H \cdot e_{n+1} = (\lambda - 2(n+1)) \cdot e_{n+1}.$$

Equation (V6) is proven by induction. The base case is trivial since we set $e_{-1} = 0$. For the induction step, since $XY - YX = [X, Y] = H$ and $ne_n = Y \cdot e_{n-1}$, we have

$$\begin{aligned} nX \cdot e_n &= XY \cdot e_{n-1} \\ &= [X, Y] \cdot e_{n-1} + YX \cdot e_{n-1} \\ &= H \cdot e_{n-1} + (\lambda - n + 2)Y \cdot e_{n-2} \\ &= (\lambda - 2n + 2)e_{n-1} + (\lambda - n + 2)(n-1)e_{n-1} \\ &= n(\lambda - n + 1)e_{n-1}, \end{aligned}$$

finishing the induction step.

By (V4), the nonzero e_n's correspond to distinct eigenvalues $\lambda - 2n$, so they are linearly independent, and since V is finite-dimensional, there is some smallest $m \geq 0$ such that $e_{m+1} = 0$. If we apply (V6) with $n = m + 1$, we get

$$0 = X \cdot 0 = X \cdot e_{m+1} = (\lambda - m)e_m$$

with $e_m \neq 0$, so $\lambda = m$. $\square$

We deduce the following theorem.

Theorem 14.10. *Given any complex representation $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V, V)$ with V finite-dimensional, for any primitive element $e \in V$ of weight $m \in \mathbb{N}$, the subspace W of V with basis $(e_0, \ldots, e_m)$ as in Proposition 14.9 is invariant under ρ and the restriction $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(W, W)$ of ρ to W is irreducible.*

Proof. Equations (V4), (V5), (V6) show that W is invariant under ρ. By Equation (V4), the $m+1$ eigenvalues of the restriction of $\rho(H)$ to W are $m, m-2, m-4, \ldots, -(m-2), -m$ and have multiplicity 1 (since W has dimension $m+1$). Suppose W' is a nonzero subspace of W invariant under ρ. Since $(e_0, \ldots, e_m)$ is a basis of W, one of the e_i must belong to W'. Since W' is invariant under ρ, we can apply (V6) to e_i several times and since $m - j + 1 \neq 0$ if $0 \leq j \leq i \leq m$, we see that $e_i, e_{i-1}, \ldots, e_0 = e$ belongs to W'. By applying (V5) to e_i we see that $e_i, e_{i+1}, \ldots, e_m$ all belong to W', so $W' = W$, that is, W is irreducible. $\square$

The nonnegative integer m is called the *highest weight* of the irreducible representation ρ in W.

Remark: For *any* representation $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ with V finite-dimensional, $\rho(H)$ is diagonalizable so we have a direct sum

$$V = \bigoplus_{\lambda} V^{\lambda}.$$

This is because $\mathfrak{sl}(2,\mathbb{C})$ is a semisimple Lie algebra (in fact, a simple Lie algebra) and $\mathrm{ad}(H)$ is diagonalizable since it has the three distinct eigenvalues $2, 0, -2$ (recall that $\mathrm{ad}(H)(Z) = [H,Z]$, and that $[H,X] = 2X, [H,H] = 0, [H,Y] = -2Y$). Then for any complex representation $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ with V finite-dimensional, $\rho(H)$ is diagonalizable. This is a special case of results about semisimple Lie algebras found in Fulton and Harris [36] (Appendix C, Section C.2) and Serre [91] (Part I, Chapter VI, Theorem 5.7).

We can now characterize all the irreducible (complex) representations of $\mathfrak{sl}(2,\mathbb{C})$.

Definition 14.9. Let $m \geq 0$ be any natural number, and let W_m be a complex vector space of dimension $m+1$ with basis $(e_0, \ldots, e_m)$. Define the endormorphisms $X^{W_m}, Y^{W_m}, Z^{W_m}$ of W_m as follows (by convention, $e_{-1} = e_{m+1} = 0$).

$$H^{W_m} e_n = (m-2n)e_n \tag{V7}$$
$$Y^{W_m} e_n = (n+1)e_{n+1} \tag{V8}$$
$$X^{W_m} e_n = (m-n+1)e_{n-1}. \tag{V9}$$

We define the homomorphism $\rho_m\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(W_m, W_m)$ by

$$\rho_m(H) = H^{W_m}, \quad \rho_m(X) = X^{W_m}, \quad \rho_m(Y) = Y^{W_m}. \tag{V10}$$

Theorem 14.11. *The homomorphism $\rho_m\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(W_m, W_m)$ is an irreducible representation of $\mathfrak{sl}(2,\mathbb{C})$. Every irreducible complex representation $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ with $\dim(V) = m+1$ is equivalent to the representation $\rho_m\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(W_m, W_m)$.*

Proof. It is easy to check that the formulae defining $X^{W_m}, Y^{W_m}, H^{W_m}$ imply that

$$\begin{aligned} H^{W_m}X^{W_m}(e_n) - X^{W_m}H^{W_m}(e_n) &= 2X^{W_m}(e_n) \\ H^{W_m}Y^{W_m}(e_n) - Y^{W_m}H^{W_m}(e_n) &= -2Y^{W_m}(e_n) \\ X^{W_m}Y^{W_m}(e_n) - Y^{W_m}X^{W_m}(e_n) &= H^{W_m}(e_n), \end{aligned}$$

so ρ_m is a representation. Observe that by construction $e = e_0$ is a primitive element of weight m and that the $Y^n e$ span W_m. Theorem 14.10 implies that ρ_m is irreducible.

Let $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ be any irreducible representation of $\mathfrak{sl}(2,\mathbb{C})$ of dimension $m+1$. By Proposition 14.8 and Proposition 14.9, V contains some primitive element e' of weight m', for some natural number m'. By Proposition 14.9, e' generates a subspace W of V invariant under ρ that has dimension $m'+1$. Since ρ is irreducible, we must have $W = V$. It follows that V has $(e'_0, e'_1, \ldots, e'_m)$ as a basis, with $e'_n = (1/n!)Y^n \cdot e'$. Define the linear

isomorphism $\varphi\colon W_m \to V$ by $\varphi(e_n) = e'_n$, for $n = 0, \ldots, m$. Since by Proposition 14.9 the e'_n satisfy Equations (V4), (V5), (V6), and by construction the e_n satisfy Equations (V7), (V8), (V9), it is immediately verified that φ is an equivalence between the representations $\rho_m\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(W_m, W_m)$ and $\rho\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V, V)$. $\square$

Since $\mathfrak{sl}(2,\mathbb{C})$ is the complexification of $\mathfrak{su}(2)$, by Proposition 14.6 and Theorem 14.11, we obtain the following result.

Theorem 14.12. *The irreducible representation $\rho_m\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(W_m, W_m)$ induces by restriction an irreducible representation $\rho_m\colon \mathfrak{su}(2) \to \mathrm{Hom}(W_m, W_m)$. Every irreducible representation $\rho\colon \mathfrak{su}(2) \to \mathrm{Hom}(V, V)$ with V a complex vector space of dimension $m+1$ is equivalent to the irreducible representation $\rho_m\colon \mathfrak{su}(2) \to \mathrm{Hom}(W_m, W_m)$.*

For $m = 0$, the space W_0, is one-dimensional space isomorphic to $\mathbb{C}$, in which case H, X, Y are the zero map on W_0; ρ_0 is the trivial representation.

For $m = 1$, the space $W_1 \simeq \mathbb{C}^2$ has the basis (e_0, e_1), and $H \cdot e_0 = e_0$, $H \cdot e_1 = -e_1$, $X \cdot e_0 = 0$, $X \cdot e_1 = e_0$, $Y \cdot e_0 = e_1$, $Y \cdot e_1 = 0$, so $W_1 = W_1^{-1} \oplus W_1^1$ where W_1^{-1} is the eigenspace spanned by e_1 associated with the eigenvalue -1, W_1^1 is the eigenspace spanned by e_0 associated with the eigenvalue 1, and ρ_1 is the standard representation on $\mathbb{C}^2$.

Remark: It turns out that ρ_m is equivalent to the representation induced by ρ_1 on the symmetric tensor power $\mathrm{Sym}^m W_1$, but given a Lie algebra representation $\rho\colon \mathfrak{g} \to \mathrm{Hom}(V, V)$, one has to define the representation $\rho^{m\odot}\colon \mathfrak{g} \to \mathrm{Hom}(\mathrm{Sym}^m V, \mathrm{Sym}^m V)$. This can be done as follows. First given two representations $\rho_1\colon \mathfrak{g} \to \mathrm{Hom}(V, V)$ and $\rho_2\colon \mathfrak{g} \to \mathrm{Hom}(W, W)$, we define the tensor product representation $\rho_1 \otimes \rho_2\colon \mathfrak{g} \to \mathrm{Hom}(V \otimes W, V \otimes W)$ by

$$[(\rho_1 \otimes \rho_2)(X)](v \otimes w) = [\rho_1(X)](v) \otimes w + v \otimes [\rho_2(X)](w), \quad X \in \mathfrak{g},\ v \in V,\ w \in W.$$

Taking inspiration from the above equation, since $\mathrm{Sym}^m W_1$ is generated by the m-fold powers $v_1 \odot \cdots \odot v_m$ with $v_1, \ldots, v_m \in V$, we define $\rho^{m\odot}$ recursively by

$$[(\rho^{m\odot})(X)](v_1 \odot \cdots \odot v_m) = [\rho(X)](v_1) \odot v_2 \cdots \odot v_m + v_1 \odot [\rho^{(m-1)}(X)](v_2 \odot \cdots \odot v_m)$$

for $m \geq 2$, with $\rho^{\odot} = \rho$. Since W_1 has the basis (e_0, e_1), it is a fact of linear algebra that $\mathrm{Sym}^m W_1$ has the basis

$$(e_0^m, \ldots, e_0^{m-n} e_1^n, \ldots, e_1^m), \quad 0 \leq n \leq m,$$

where for notational simplicity we suppressed the symbol $\odot$, so we can find $[\rho_1^{m\odot}(H)](e_0^{m-n} e_1^n)$. We can show by induction that

$$\rho_1^{m\odot}(H)(e_0^{m-n} e_1^n) = (m-n) e_0^{m-n-1} e_1^n \rho_1(H) \cdot e_0 + n e_0^{m-n} e_1^{n-1} \rho_1(H) \cdot e_1,$$

and since $\rho_1(H) \cdot e_0 = e_0$ and $\rho_1(H) \cdot e_1 = -e_1$, we get

$$\rho_1^{m\odot}(H)(e_0^{m-n} e_1^n) = (m-2n) e_0^{m-n} e_1^n.$$

Thus the eigenvalues of $\rho_1^{m\odot}(H)$ are the $m+1$ integers $m, m-2, \ldots, -(m-2), -m$, and this implies that $\rho_1^{m\odot}$ is equivalent to ρ_m. Since $\mathrm{Sym}^m W_1$ is isomorphic to the space of homogeneous polynomials of degree m in two variables, we have an "a posteriori" explanation of the fact that the spaces of the irreducible representations of $\mathbf{SL}(2,\mathbb{C})$ (and $\mathbf{SU}(2)$) are these spaces of homogeneous polynomials. See Fulton and Harris [36] (Chapter 11, Section 11.1).

As in the case of the representations of compact groups, we have the following result but its proof is far from immediate.

Theorem 14.13. *Every representation $\rho_m\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(V,V)$ of $\mathfrak{sl}(2,\mathbb{C})$ with V of finite dimension splits as a direct sum of irreducible representation $\rho_m\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathrm{Hom}(W_m,W_m)$. The number of irreducible factors isomorphic to ρ_m is the sum of the multiplicities of 0 and 1 as eigenvalues of $\rho(H)$.*

Theorem 14.13 is known as *complete reducibility* and is usually attributed to H. Weyl. A fascinating account of the history of its proof, starting in the mid 1890's with proofs of E. Cartan and G. Fano, can be found in Borel [5].

For a proof of Theorem 14.13, see Fulton and Harris [36] (Appendix C, Section C.2), Serre [91] (Part I, Chapter VI, Section 3), Humphreys [51] (Chapter II, Section 6.3) and Samelson [82] (Chapter 1, Section 1.12). See also Fulton and Harris [36] (Chapter 9, Section 3) for a sketch of a proof using "Weyl's unitary trick."

Weyl's unitary trick (actually called "unitarian trick" by Weyl himself) is discussed in Serre [92] (Chapter IV, Theorem 6) in the special case of $\mathbf{SU}(2), \mathbf{SL}(2,\mathbb{C})$, $\mathfrak{su}(2)$, and $\mathfrak{sl}(2,\mathbb{C})$.

Let G be a complex Lie group and let $\mathfrak{g}$ be its (complex) Lie algebra. The trick works for the following reasons:

(1) The complex Lie algebra $\mathfrak{sl}(2,\mathbb{C})$ is the complexification of the real Lie algebra $\mathfrak{su}(2)$. It follows by (an easy adaptation of) Proposition 14.6 that there is a bijection d between the set $\mathrm{Hom}_{\mathbb{C}}(\mathfrak{sl}(2,\mathbb{C}),\mathfrak{g})$ of $\mathbb{C}$-homomorphisms of the Lie algebras $\mathfrak{sl}(2,\mathbb{C})$ and $\mathfrak{g}$ and the set $\mathrm{Hom}_{\mathbb{R}}(\mathfrak{su}(2),\mathfrak{g})$ of $\mathbb{R}$-homomorphisms of the Lie algebras $\mathfrak{su}(2)$ and $\mathfrak{g}$.

(2) The Lie groups $\mathbf{SU}(2)$ and $\mathbf{SL}(2,\mathbb{C})$ are connected and simply-connected.

(3) It follows from (2) (see Gallier and Quaintance [38] (Theorem 19.20), Fulton and Harris [36] (Chapter 8, Section 3), Warner [102] (Chapter 3, Theorem 3.27) that there is a bijection b between the set $\mathrm{Hom}_{\mathbb{C}}(\mathbf{SL}(2,\mathbb{C}),G)$ of $\mathbb{C}$-homomorphisms (holomorphic maps) of the Lie groups $\mathbf{SL}(2,\mathbb{C})$ and G and the set $\mathrm{Hom}_{\mathbb{C}}(\mathfrak{sl}(2,\mathbb{C}),\mathfrak{g})$ of $\mathbb{C}$-homomorphisms of the Lie algebras $\mathfrak{sl}(2,\mathbb{C})$ and $\mathfrak{g}$, and a bijection c between the set $\mathrm{Hom}_{\mathbb{R}}(\mathbf{SU}(2),G)$ of $\mathbb{R}$-homomorphisms of the Lie groups $\mathbf{SU}(2)$ and G and the set $\mathrm{Hom}_{\mathbb{R}}(\mathfrak{su}(2),\mathfrak{g})$ of $\mathbb{R}$-homomorphisms of the Lie algebras $\mathfrak{su}(2)$ and $\mathfrak{g}$.

As a consequence we obtain the following result.

Theorem 14.14. *(Weyl's Unitarian Trick) Let G be a complex Lie group and let $\mathfrak{g}$ be its (complex) Lie algebra. The following diagram commutes and all maps in it are bijections.*

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbb{C}}(\mathbf{SL}(2,\mathbb{C}), G) \arrow[r, "a"] \arrow[d, "b"'] & \mathrm{Hom}_{\mathbb{R}}(\mathbf{SU}(2), G) \arrow[d, "c"] \\
\mathrm{Hom}_{\mathbb{C}}(\mathfrak{sl}(2,\mathbb{C}), \mathfrak{g}) \arrow[r, "d"'] & \mathrm{Hom}_{\mathbb{R}}(\mathfrak{su}(2), \mathfrak{g}).
\end{tikzcd}$$

The only nonobvious map is a, which is the composition $c^{-1} \circ d \circ b$.

If we apply Theorem 14.14 to $G = \mathbf{GL}(V)$ and $\mathfrak{g} = \mathrm{Hom}(V, V)$ where V is a complex vector space, since $\mathbf{SU}(2)$ is a compact Lie group, by Peter–Weyl II we obtain complete reducibility, the fact that the representations of $\mathfrak{sl}(2,\mathbb{C})$, $\mathfrak{su}(2)$ and $\mathbf{SL}(2,\mathbb{C})$ (and of course $\mathbf{SU}(2)$) split as direct sums of irreducible representations whose representing spaces are all described by Definition 14.9. Using the isomorphism (c), we also rediscover the structure of the irreducible representations of $\mathbf{SU}(2)$.

14.7 $\mathbf{SU}(2)$-Invariant Hermitian Inner Product on $\mathcal{P}_\ell^{\mathbb{C}}$

We now restrict our attention to the representations T_ℓ of $\mathbf{SU}(2)$. Our goal is to find explicitly an $\mathbf{SU}(2)$-invariant hermitian inner product on $\mathcal{P}_\ell^{\mathbb{C}}$. Because $\mathbf{SU}(2)$ is compact, such an inner product must exist. If such an invariant hermitian inner product $\langle -, - \rangle$ exists, in particular it must be invariant for the matrices $r_x(\varphi/2)$, $r_y(\theta/2)$ and $r_z(\psi/2)$, so we assert such invariance and deduce consequences by taking derivatives.

First we need to figure out what is $T_\ell(r_x(\varphi/2))(z^{\ell-k})$. Since

$$r_x(\varphi/2) = \begin{pmatrix} e^{i\frac{\varphi}{2}} & 0 \\ 0 & e^{-i\frac{\varphi}{2}} \end{pmatrix},$$

with $a = e^{i\frac{\varphi}{2}}$, $b = c = 0$, and $d = e^{-i\frac{\varphi}{2}}$, the formula

$$T_\ell(A)(Q(z)) = (bz + d)^{2\ell} Q\left(\frac{az + c}{bz + d}\right).$$

yields

$$T_\ell(r_x(\varphi/2))(z^{\ell-k}) = e^{-i\ell\varphi}\left(\frac{e^{i\frac{\varphi}{2}} z}{e^{-i\frac{\varphi}{2}}}\right)^{\ell-k} = e^{-i\ell\varphi} e^{i(\ell-k)\varphi} z^{\ell-k} = e^{-ik\varphi} z^{\ell-k},$$

that is,

$$T_\ell(r_x(\varphi/2))(z^{\ell-k}) = e^{-ik\varphi} z^{\ell-k}.$$

The above equation is important enough to be recorded as a proposition.

Proposition 14.15. *Each polynomial $z^{\ell-k}$ is an eigenvector of $T_\ell(r_x(\varphi/2))$ for the eigenvalue $e^{-ik\varphi}$, that is,*

$$T_\ell(r_x(\varphi/2))(z^{\ell-k}) = e^{-ik\varphi} z^{\ell-k}. \tag{$*_1$}$$

Thus in the basis $(z^{\ell-k})_{-\ell\leq k\leq \ell}$, the matrix of $T_\ell(r_x(\varphi/2))$ is the diagonal matrix

$$\begin{pmatrix} e^{i\ell\varphi} & & & & \\ & e^{i(\ell-1)\varphi} & & & \\ & & \ddots & & \\ & & & e^{-i(\ell-1)\varphi} & \\ & & & & e^{-i\ell\varphi} \end{pmatrix}.$$

The invariance of the inner product for $T_\ell(r_x(\varphi/2))$ is stated as

$$\langle T_\ell(r_x(\varphi/2))(z^{\ell-j}),\, T_\ell(r_x(\varphi/2))(z^{\ell-k})\rangle = \langle z^{\ell-j},\, z^{\ell-k}\rangle \tag{$*_2$}$$

for all j, k with $-\ell \leq j, k \leq \ell$, and since

$$T_\ell(r_x(\varphi/2))(z^{\ell-j}) = e^{-ij\varphi} z^{\ell-j} \quad \text{and} \quad T_\ell(r_x(\varphi/2))(z^{\ell-k}) = e^{-ik\varphi} z^{\ell-k}$$

(remembering that the hermitian inner product is semilinear on the second argument!), we obtain

$$\langle T_\ell(r_x(\varphi/2))(z^{\ell-j}),\, T_\ell(r_x(\varphi/2))(z^{\ell-k})\rangle = e^{-i(j-k)\varphi}\langle z^{\ell-j}, z^{\ell-k}\rangle. \tag{$*_3$}$$

Equations $(*_2)$ and $(*_3)$ yield

$$e^{-i(j-k)\varphi}\langle z^{\ell-j}, z^{\ell-k}\rangle = \langle z^{\ell-j}, z^{\ell-k}\rangle,$$

and these equations show that

$$\langle z^{\ell-j},\, z^{\ell-k}\rangle = 0, \quad \text{for all } j \neq k. \tag{$*_4$}$$

Next we need to compute $\langle z^{\ell-k}, z^{\ell-k}\rangle$ to find the normalization factors. Here we assert invariance of the inner product for $T_\ell(r_y(\theta/2))$ for $z^{\ell-k}$ and $z^{\ell-k+1}$, which is stated as

$$\langle T_\ell(r_y(\theta/2))(z^{\ell-k}),\, T_\ell(r_y(\theta/2))(z^{\ell-k+1})\rangle = \langle z^{\ell-k},\, z^{\ell-k+1}\rangle \tag{$*_5$}$$

for all k with $-\ell k \leq \ell$. The trick is to differentiate the above equation at $\theta = 0$. Since $r_y(\theta/2) = e^{\theta\xi_2}$, we obtain

$$\langle \mathfrak{t}_\ell(\xi_2)(z^{\ell-k}),\, z^{\ell-k+1}\rangle + \langle z^{\ell-k},\, \mathfrak{t}_\ell(\xi_2)(z^{\ell-k+1})\rangle = 0. \tag{$*_6$}$$

Using Equation ($\mathfrak{t}$5), we obtain

$$-(\ell+k)\langle z^{\ell-k+1},\, z^{\ell-k+1}\rangle + (\ell-k+1)\langle z^{\ell-k},\, z^{\ell-k}\rangle = 0. \tag{$*_7$}$$

By changing k to $k+1$, we obtain

$$(\ell + k + 1)\langle z^{\ell-k},\, z^{\ell-k}\rangle = (\ell - k)\langle z^{\ell-k-1},\, z^{\ell-k-1}\rangle, \tag{$*_8$}$$

and this recurrence equation yields

$$\langle z^{\ell-k},\, z^{\ell-k}\rangle = \frac{(\ell-k)!}{(\ell+k+1)\cdots(2\ell)}\langle 1,\, 1\rangle = \frac{(\ell-k)!(\ell+k)!}{(2\ell)!}\langle 1,\, 1\rangle.$$

It is natural to pick

$$\langle 1, 1\rangle = (2\ell)!,$$

and so we obtain

$$\langle z^{\ell-k},\, z^{\ell-k}\rangle = (\ell-k)!(\ell+k)!, \quad -\ell \leq k \leq \ell. \tag{$*_9$}$$

Equations ($*_4$) and ($*_9$) shows that the $2\ell + 1$ polynomials

$$\frac{z^{\ell-k}}{\sqrt{(\ell-k)!(\ell+k)!}},$$

form an orthonormal basis of $\mathcal{P}_\ell^{\mathbb{C}}$ for an invariant hermitian inner product on $\mathbf{SU}(2)$ which is uniquely determined by setting $\langle 1, 1\rangle = (2\ell)!$. This is an important result that we record below.

Proposition 14.16. *In Vilenkin's notation, the polynomials*

$$\psi_k(z) = \frac{z^{\ell-k}}{\sqrt{(\ell-k)!(\ell+k)!}}, \quad -\ell \leq k \leq \ell \tag{$*_{10}$}$$

form an orthonormal basis of $\mathcal{P}_\ell^{\mathbb{C}}$ for a unique invariant hermitian inner product on $\mathbf{SU}(2)$. The ψ_k are the unit-length eigenvectors of the linear map $T_\ell(r_x(\varphi/2))$.

Also note that the formulae (H_+), (H_-), (H_3) become

$$H_+\psi_k(z) = -\sqrt{(\ell-k)(\ell+k+1)}\,\psi_{k+1}(z) \tag{H'_+}$$

$$H_-\psi_k(z) = -\sqrt{(\ell+k)(\ell-k+1)}\,\psi_{k-1}(z) \tag{H'_-}$$

$$H_3\psi_k(z) = k\,\psi_k(z). \tag{H'_3}$$

Actually, it is remarkable that if we define a hermitian inner product on $\mathcal{P}_\ell^{\mathbb{C}}$ by requiring that the polynomials ψ_k form an orthonormal basis, then this inner product is $\mathbf{SU}(2)$ invariant. The proof of this fact relies on two standard facts of Lie group theory about the relationship between a representation and its derivative.

First recall that if $f\colon G \to H$ is a homomorphism of Lie groups, then the derivative df_e of f at the identity element e of G is a Lie algebra homorphism $df_e\colon \mathfrak{g} \to \mathfrak{h}$; see Gallier

and Quaintance [38] (Chapter 19). In particular, if $H = \mathbf{GL}(E)$, where E is a finite-dimensional (complex) vector space, then f is a representation, and since the Lie algebra of the group $\mathbf{GL}(E)$ is $\mathfrak{gl}(E) = \mathrm{Hom}(E, E)$, the space of all linear maps from E to itself, $df_e \colon \mathfrak{g} \to \mathrm{Hom}(E, E)$ is what is called a *Lie algebra representation*.

If E has a hermitian inner product $\langle -, - \rangle$ and if $H = \mathbf{U}(E)$, the group of unitary linear maps with respect to the hermitian inner product $\langle -, - \rangle$, we claim that the Lie algebra $\mathfrak{u}(E)$ of $\mathbf{U}(E)$ consists of the linear maps $Z \colon E \to E$ such that

$$\langle Z(u), v \rangle + \langle u, Z(v) \rangle = 0, \quad \text{for all } u, v \in E, \tag{skew$_1$}$$

or equivalently

$$Z^* = -Z, \tag{skew$_2$}$$

where Z^* is the adjoint of Z with respect to the hermitian inner product $\langle -, - \rangle$, which is the unique linear map Z^* defined by the property that

$$\langle Z(u), v \rangle = \langle u, Z^*(v) \rangle, \quad \text{for all } u, v \in E.$$

Linear maps $Z \colon E \to E$ satisfying property (skew$_1$) (equivalently (skew$_2$)) are called *skew-hermitian* with respect to the hermitian inner product $\langle -, - \rangle$. First, since $\mathfrak{u}(E)$ is the tangent space to $\mathbf{U}(E)$ at id, by definition $Z = C'(0)$ for any smooth curve $C \colon (-\epsilon, \epsilon) \to \mathbf{U}(E)$ such that $C(0) = \mathrm{id}$, and since each $C(t)$ is unitary, we have

$$\langle C(t)(u), C(t)(v) \rangle = \langle u, v \rangle$$

for all $t \in (-\epsilon, \epsilon)$ and all $u, v \in E$, so by differentiating at $t = 0$ we get

$$\langle C'(0)(u), C(0)(v) \rangle + \langle C(0)(u), C'(0)(v) \rangle = 0,$$

which, since $C'(0) = Z$ and $C(0) = \mathrm{id}$, yields

$$\langle Z(u), v \rangle + \langle u, Z(v) \rangle = 0,$$

which is Equation (skew$_1$). To show that all skew-hermitian linear maps belong to $\mathfrak{u}(E)$, we use standard properties of the exponential map, namely that if Z is skew-hermitian, then $(e^Z)^* = e^{(Z^*)} = e^{-Z}$, and so e^Z is unitary (all with respect to the hermitian inner product $\langle -, - \rangle$ on E). Since E is finite-dimensional, we can pick an orthonormal basis of E with respect to $\langle -, - \rangle$, and work with matrices. As a corollary we have the following result.

Proposition 14.17. *Define a hermitian inner product $\langle -, - \rangle$ on $\mathcal{P}_\ell^{\mathbb{C}}$ by requiring that the polynomials ψ_k form an orthonormal basis. Then for any unitary representation $T_\ell \colon \mathbf{SU}(2) \to \mathbf{U}(\mathcal{P}_\ell^{\mathbb{C}})$ (with respect to $\langle -, - \rangle$) we obtain the Lie algebra representation $\mathfrak{t}_\ell \colon \mathfrak{su}(2) \to \mathfrak{u}(\mathcal{P}_\ell^{\mathbb{C}})$, where $\mathfrak{t}_\ell = d(T_\ell)_I$. Thus*

$$\mathfrak{t}_\ell(X)^* = -\mathfrak{t}_\ell(X), \quad X \in \mathfrak{su}(2), \tag{$*_{15}$}$$

namely $\mathfrak{t}_\ell(X)$ is skew-hermitian with respect to the hermitian inner product $\langle -, - \rangle$.

The converse is true.

Proposition 14.18. *For any representation $T_\ell\colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$, let $\mathfrak{t}_\ell = d(T_\ell)_I$, so that $\mathfrak{t}_\ell\colon \mathfrak{su}(2) \to \mathrm{Hom}(\mathcal{P}_\ell^{\mathbb{C}}, \mathcal{P}_\ell^{\mathbb{C}})$ is the corresponding Lie algebra representation. If for every $X \in \mathfrak{su}(2)$ the linear map $\mathfrak{t}_\ell(X)\colon \mathcal{P}_\ell^{\mathbb{C}} \to \mathcal{P}_\ell^{\mathbb{C}}$ is skew-hermitian with respect to the hermitian inner product $\langle -, - \rangle$ on $\mathcal{P}_\ell^{\mathbb{C}}$ making the basis (ψ_k) orthonormal, then $T_\ell(A)$ is unitary with respect to $\langle -, - \rangle$ for all $A \in \mathbf{SU}(2)$; in other words, T_ℓ is a unitary representation $T_\ell\colon \mathbf{SU}(2) \to \mathbf{U}(\mathcal{P}_\ell^{\mathbb{C}})$.*

Proof. This result is actually true for any representation $U\colon G \to \mathbf{GL}(E)$ where G is a connected Lie group and E is finite-dimensional and equipped with a hermitian inner product, but since the exponential map $\exp\colon \mathfrak{su}(2) \to \mathbf{SU}(2)$ is surjective (Gallier and Quaintance [40], Section 15.5) we can give a simpler proof. Since every $q \in \mathbf{SU}(2)$ can be written as $q = e^X$ for some $X \in \mathfrak{su}(2)$, consider the function $F\colon \mathbb{R} \to \mathbb{C}$ given by

$$F(t) = \langle T_\ell(e^{tX})(\psi_j), T_\ell(e^{tX})(\psi_k)\rangle, \tag{$*_{16}$}$$

which has the property that $F(0) = \langle \psi_j, \psi_k\rangle$. We prove that F is constant by showing that its derivative is zero for all t. If so, since $\mathbf{SU}(2)$ is connected, F must be constant, and since its value at $t = 0$ is $\langle \psi_j, \psi_k\rangle$, for $t = 1$ we obtain

$$\langle T_\ell(q)(\psi_j), T_\ell(q)(\psi_k)\rangle = \langle \psi_j, \psi_k\rangle,$$

which proves that $T_\ell(q)$ is unitary with respect to $\langle -, -\rangle$. Because the map $t \mapsto h(t) = T_\ell(e^{tX})$ is a one-parameter group and $h'(0) = d(T_\ell)_I(X) = \mathfrak{t}_\ell(X)$, by Lie group theory,

$$T_\ell(e^{tX}) = e^{t\mathfrak{t}_\ell(X)};$$

see Gallier and Quaintance [38] (Proposition 4.13). By the chain rule

$$d(T_\ell(e^{tX}))_s = d(e^{t\mathfrak{t}_\ell(X)})_s = \mathfrak{t}_\ell(X) \circ e^{s\mathfrak{t}_\ell(X)} = \mathfrak{t}_\ell(X) \circ T_\ell(e^{sX}).$$

If we take the derivative of Equation ($*_{16}$) (at any $t = s$) we get

$$F'(s) = \langle \mathfrak{t}_\ell(X)(T_\ell(e^{sX})(\psi_j)), T_\ell(e^{sX})(\psi_k)\rangle + \langle T_\ell(e^{sX})(\psi_j), \mathfrak{t}_\ell(X)(T_\ell(e^{sX})(\psi_k))\rangle. \tag{$*_{17}$}$$

Since $\mathfrak{t}_\ell(X)$ is skew-hermitian by hypothesis, we conclude that $F'(s) = 0$ for all s. $\square$

According to Proposition 14.18, to prove that the hermitian inner product $\langle -, -\rangle$ on $\mathcal{P}_\ell^{\mathbb{C}}$ making the basis (ψ_k) orthonormal is $\mathbf{SU}(2)$-invariant, it suffices to prove that the linear maps $\mathfrak{t}_\ell(X)$ are skew-hermitian with respect to $\langle -, -\rangle$ for all $X \in \mathfrak{su}(2)$. Since (ξ_1, ξ_2, ξ_3) is a basis of $\mathfrak{su}(2)$, we need to prove that $\mathfrak{t}_\ell(\xi_i)$ is skew-hermitian for $i = 1, 2, 3$.

Proposition 14.19. *The linear maps $\mathfrak{t}_\ell(\xi_i)$ $(1 \leq i \leq 3)$ are skew-hermitian for the hermitian inner product $\langle -, -\rangle$ on $\mathcal{P}_\ell^{\mathbb{C}}$ making the basis (ψ_k) orthonormal.*

Proof. First we prove that $\mathfrak{t}_\ell(\xi_1)$ is skew-hermitian using Equation ($\mathfrak{t}4$),

$$\mathfrak{t}_\ell(\xi_1) z^{\ell-k} = \frac{i}{2}(\ell-k) z^{\ell-k-1} + \frac{i}{2}(\ell+k) z^{\ell-k+1},$$

which is expressed in terms of the basis $(z^{\ell-k})$, and thus needs some adjustment. We divide both sides by $\sqrt{(\ell-k)!(\ell+k)!}$, which yields

$$\mathfrak{t}_\ell(\xi_1)\frac{z^{\ell-k}}{\sqrt{(\ell-k)!(\ell+k)!}} = \frac{i}{2}(\ell-k)\frac{z^{\ell-k-1}}{\sqrt{(\ell-k)!(\ell+k)!}} + \frac{i}{2}(\ell+k)\frac{z^{\ell-k+1}}{\sqrt{(\ell-k)!(\ell+k)!}}.$$

Since

$$\psi_{k+1}(z) = \frac{z^{\ell-k-1}}{\sqrt{(\ell-k-1)!(\ell+k+1)!}}, \quad \psi_{k-1}(z) = \frac{z^{\ell-k+1}}{\sqrt{(\ell-k+1)!(\ell+k-1)!}},$$

we need to compute

$$\begin{aligned}\frac{(\ell-k)\sqrt{(\ell-k-1)!(\ell+k+1)!}}{\sqrt{(\ell-k)!(\ell+k)!}} &= \sqrt{(\ell-k)(\ell+k+1)}\frac{\sqrt{(\ell-k-1)!(\ell-k)(\ell+k)!}}{\sqrt{(\ell-k)!(\ell+k)!}}\\ &= \sqrt{(\ell-k)(\ell+k+1)},\end{aligned}$$

and

$$\begin{aligned}\frac{(\ell+k)\sqrt{(\ell-k+1)!(\ell+k-1)!}}{\sqrt{(\ell-k)!(\ell+k)!}} &= \sqrt{(\ell+k)(\ell-k+1)}\frac{\sqrt{(\ell-k)!(\ell+k-1)!(\ell+k)}}{\sqrt{(\ell-k)!(\ell+k)!}}\\ &= \sqrt{(\ell+k)(\ell-k+1)}.\end{aligned}$$

Consequently we obtain the equation

$$\mathfrak{t}_\ell(\xi_1)\psi_k(z) = \frac{i}{2}\sqrt{(\ell-k)(\ell+k+1)}\psi_{k+1}(z) + \frac{i}{2}\sqrt{(\ell+k)(\ell-k+1)}\psi_{k-1}(z). \qquad (*_{18})$$

Since (ψ_k) is an orthonormal basis, the (j,k) entry of the matrix $\mathbf{t}^{(1)}$ representing $\mathfrak{t}_\ell(\xi_1)$ is

$$\begin{aligned}\mathbf{t}^{(1)}_{jk} = \langle \mathfrak{t}_\ell(\xi_1)(\psi_k(z)), \psi_j(z)\rangle &= \frac{i}{2}\sqrt{(\ell-k)(\ell+k+1)}\langle \psi_{k+1}(z), \psi_j(z)\rangle\\ &\quad + \frac{i}{2}\sqrt{(\ell+k)(\ell-k+1)}\langle \psi_{k-1}(z), \psi_j(z)\rangle,\end{aligned}$$

and so the only nonzero entries are

$$\begin{aligned}\mathbf{t}^{(1)}_{k+1k} &= \frac{i}{2}\sqrt{(\ell-k)(\ell+k+1)}\\ \mathbf{t}^{(1)}_{k-1k} &= \frac{i}{2}\sqrt{(\ell+k)(\ell-k+1)},\end{aligned}$$

and by changing k to $k+1$

$$\mathbf{t}^{(1)}_{kk+1} = \frac{i}{2}\sqrt{(\ell+k+1)(\ell-k)},$$

and finally

$$\mathbf{t}^{(1)}_{kk+1} = \mathbf{t}^{(1)}_{k+1k} = \frac{i}{2}\sqrt{(\ell-k)(\ell+k+1)}.$$

It follows that $\mathbf{t}^{(1)}$ is a pure imaginary matrix such that $-\overline{\mathbf{t}^{(1)}_{kk+1}} = \mathbf{t}^{(1)}_{kk+1} = \mathbf{t}^{(1)}_{k+1k}$, which proves that $\mathfrak{t}_\ell(\xi_1)$ is skew-hermitian.

To prove that $\mathfrak{t}_\ell(\xi_2)$ is skew-hermitian we use Equation ($\mathfrak{t}$5),

$$\mathfrak{t}_\ell(\xi_2)z^{\ell-k} = \frac{1}{2}(\ell-k)z^{\ell-k-1} - \frac{1}{2}(\ell+k)z^{\ell-k+1},$$

which only differs by the absence of i and the fact that the sign in front of the second term is -1 instead of 1. This time we find that

$$\mathfrak{t}_\ell(\xi_2)\psi_k(z) = \frac{1}{2}\sqrt{(\ell-k)(\ell+k+1)}\psi_{k+1}(z) - \frac{1}{2}\sqrt{(\ell+k)(\ell-k+1)}\psi_{k-1}(z). \qquad (*_{19})$$

The (j,k) entry of the matrix $\mathbf{t}^{(2)}$ representing $\mathfrak{t}_\ell(\xi_2)$ is nonzero iff

$$\begin{aligned}\mathbf{t}^{(2)}_{kk+1} &= \frac{1}{2}\sqrt{(\ell-k)(\ell+k+1)}\\ \mathbf{t}^{(2)}_{k+1k} &= -\frac{1}{2}\sqrt{(\ell+k+1)(\ell-k)},\end{aligned}$$

It follows that $\mathbf{t}^{(2)}$ is a real matrix such that $-\overline{\mathbf{t}^{(2)}_{kk+1}} = \mathbf{t}^{(2)}_{k+1k}$, which proves that $\mathfrak{t}_\ell(\xi_2)$ is skew-hermitian.

To prove that $\mathfrak{t}_\ell(\xi_3)$ is skew-hermitian we use Equation ($\mathfrak{t}$6),

$$\mathfrak{t}_\ell(\xi_3)z^{\ell-k} = -ikz^{\ell-k}.$$

Since $z^{\ell-k}$ is an eigenvector, this is simpler. By dividing both sides by $\sqrt{(\ell-k)!(\ell+k)!}$ we obtain

$$\mathfrak{t}_\ell(\xi_3)\psi_k(z) = -ik\psi_k(z). \qquad (*_{20})$$

It follows that $\mathbf{t}^{(3)}$ is a pure imaginary diagonal matrix with diagonal elements

$$\mathbf{t}^{(3)}_{kk} = -ik,$$

and so $\mathfrak{t}_\ell(\xi_3)$ is skew-hermitian. Having verified that the three linear maps $\mathfrak{t}_\ell(\xi_i)$ are skew-hermitian, we conclude as we said earlier that the hermitian inner product defined by requiring that the (ψ_k) form an orthonormal basis is $\mathbf{SU}(2)$-invariant. $\square$

In summary we proved the following result.

Proposition 14.20. *The hermitian inner product on $\mathcal{P}_\ell^{\mathbb{C}}$ making the basis (ψ_k) orthonormal is* $\mathbf{SU}(2)$*-invariant.*

Note that this inner product is *not* invariant with respect to $\mathbf{SL}(2,\mathbb{C})$, because as before the linear maps $\mathfrak{t}_\ell(X)$ are skew-hermitian for $X \in \mathfrak{su}(2)$, but are *hermitian* for $X \in i\mathfrak{su}(2)$ (recall that $\mathfrak{sl}(2,\mathbb{C}) = \mathfrak{su}(2) \oplus i\mathfrak{su}(2)$, as a *real* vector space).

14.8 Matrices of the Irreducible Representations of $\mathbf{SL}(2,\mathbb{C})$

We now use the orthonormal basis (ψ_k) to find various expressions for the matrix entries of the matrix $t^{(\ell)}(A)$ representing $T_\ell(A)$ in this basis. We this section we consider an arbitrary matrix

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad \alpha,\beta,\gamma,\delta \in \mathbb{C},\ \alpha\delta - \beta\gamma = 1$$

in $\mathbf{SL}(2,\mathbb{C})$. The special case of $\mathbf{SU}(2)$ is considered in later sections. In this latter case these matrices are unitary. We use $\alpha,\beta,\gamma,\delta$ instead of a,b,c,d to make it easier to follow Vilenkin's exposition. Since the ψ_k form an orthonormal basis, we have

$$t^{(\ell)}_{jk}(A) = \langle T_\ell(A)(\psi_k),\, \psi_j\rangle = \frac{\langle T_\ell(A)(z^{\ell-k}),\, z^{\ell-j}\rangle}{\sqrt{(\ell-j)!(\ell+j)!(\ell-k)!(\ell+k)!}}. \tag{$*_{21}$}$$

By (T_ℓ) we have

$$T_\ell(A)(z^{\ell-k}) = (\beta z+\delta)^{2\ell}\left(\frac{\alpha z+\gamma}{\beta z+\delta}\right)^{\ell-k} = (\alpha z+\gamma)^{\ell-k}(\beta z+\delta)^{\ell+k},$$

so we obtain

$$t^{(\ell)}_{jk}(A) = \frac{\langle (\alpha z+\gamma)^{\ell-k}(\beta z+\delta)^{\ell+k},\, z^{\ell-j}\rangle}{\sqrt{(\ell-j)!(\ell+j)!(\ell-k)!(\ell+k)!}}. \tag{$*_{22}$}$$

The expression on the right-hand side can be "doctored on" in various ways.

The first brute-force method is to use the binomial formula together with the orthogonality of $z^{\ell-j}$ and $z^{\ell-k}$ for $j \neq k$ and the formulae

$$\langle z^{\ell-k},\, z^{\ell-k}\rangle = (\ell-k)!(\ell+k)!, \quad -\ell \leq k \leq \ell.$$

We get

$$t^{(\ell)}_{jk}(A) = \sqrt{\frac{(\ell-j)!(\ell+j)!}{(\ell-k)!(\ell+k)!}}\sum_{h=M}^{N}\binom{\ell-k}{\ell-j-h}\binom{\ell+k}{h}\alpha^{\ell-j-h}\beta^{h}\gamma^{j+h-k}\delta^{\ell+k-h} \tag{$*_{23}$}$$

with $M = \max(0, k-j)$, $N = \min(\ell - j, \ell + k)$, which can be somewhat simplified as

$$\begin{aligned} t^{(\ell)}_{jk}(A) &= \sqrt{(\ell-j)!(\ell+j)!(\ell-k)!(\ell+k)!} \\ &\quad \times \sum_{h=M}^{N} (h!(\ell-j-h)!(\ell+k-h)!(j-k+h)!)^{-1}\, \alpha^{\ell-j-h}\beta^{h}\gamma^{j+h-k}\delta^{\ell+k-h}, \qquad (*_{24}) \end{aligned}$$

also with $M = \max(0, k-j)$, $N = \min(\ell - j, \ell + k)$. It is understood that if any of $\alpha, \beta, \gamma, \delta$ is zero, then the corresponding exponent must be zero. Of course, since $\alpha\delta - \beta\gamma = 1$, at most two of these coefficients must be nonzero.

Using the factorization of A as the product of an upper triangular matrix and a lower triangular matrix, Vilenkin obtains simpler formulae; see Vilenkin [101] (Chapter III, Section 3.2). Suppose $\delta \neq 0$. Then we check immediately that if $\alpha\delta - \beta\gamma = 1$, then $\alpha = \delta^{-1} + (\beta\gamma)/\delta$, and so

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \delta^{-1} & \beta \\ 0 & \delta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \gamma/\delta & 1 \end{pmatrix}.$$

If we denote the first of the two matrices on the right-hand side by B and the second matrix by C, we have $A = BC$, and since T_ℓ is a representation,

$$T_\ell(A) = T_\ell(B)T_\ell(C),$$

which in terms of the matrices $t^{(\ell)}(A), t^{(\ell)}(B), t^{(\ell)}(C)$ means that

$$t^{(\ell)}(A) = t^{(\ell)}(B)t^{(\ell)}(C).$$

Therefore, if we compute the matrices $t^{(\ell)}(B)$ and $t^{(\ell)}(C)$, then $t^{(\ell)}_{jk}(A)$ will be given by

$$t^{(\ell)}_{jk}(A) = \sum_{h=-\ell}^{\ell} t^{(\ell)}_{jh}(B)t^{(\ell)}_{hk}(C).$$

To compute $t^{(\ell)}(B)$, we set $\gamma = 0$ and $\alpha = \delta^{-1}$ in Formula $(*_{24})$. The only nonzero term is obtained for $h = k - j$, and since h must be a nonnegative integer, we must have $j \leq k$. We obtain the formula

$$t^{(\ell)}_{jk}(B) = \begin{cases} 0 & \text{if } j > k \\ \sqrt{\dfrac{(\ell-j)!(\ell+k)!}{(\ell+j)!(\ell-k)!}}\, \dfrac{\beta^{k-j}\delta^{j+k}}{(k-j)!} & \text{if } j \leq k. \end{cases}$$

To compute $t^{(\ell)}(C)$, we set $\beta = 0$, $\alpha = \delta = 1$, and substitute γ/δ for γ in Formula $(*_{24})$. The only nonzero term is obtained for $h = 0$, and for $(j-k)!$ to make sense we must have $j \geq k$. We obtain the formula

$$t^{(\ell)}_{jk}(C) = \begin{cases} 0 & \text{if } j < k \\ \sqrt{\dfrac{(\ell+j)!(\ell-k)!}{(\ell-j)!(\ell+k)!}}\, \dfrac{\gamma^{j-k}\delta^{k-j}}{(j-k)!} & \text{if } j \geq k. \end{cases}$$

The $\beta\delta$ term in $t^{(\ell)}_{jh}(B)$ is $\beta^{h-j}\delta^{j+h}$ and the $\gamma\delta$ term in $t^{(\ell)}_{hk}(B)$ is $\gamma^{h-k}\delta^{k-h}$, so the $\beta\gamma\delta$ term in $t^{(\ell)}_{jh}(B)t^{(\ell)}_{hk}(C)$ is

$$\beta^{h-j}\delta^{j+h}\gamma^{h-k}\delta^{k-h} = \beta^{h-j}\gamma^{h-k}\delta^{j+k}.$$

Since $t^{(\ell)}_{jh}(B) = 0$ if $j > h$ and $t^{(\ell)}_{hj}(C) = 0$ if $h < k$, the only nonzero terms occur for $h \geq \max(j,k)$. In summary we proved the following result.

Proposition 14.21. *With respect to the orthonormal basis (ψ_k) of $\mathcal{P}^{\mathbb{C}}_\ell$, the entries in the matrix $t^{(\ell)}(A)$ are given by the formulae below.*

(1) If $\delta \neq 0$, then

$$t^{(\ell)}_{jk}(A) = \sqrt{\frac{(\ell-j)!(\ell-k)!}{(\ell+j)!(\ell+k)!}} \times \sum_{h=\max(j,k)}^{\ell} \frac{(\ell+h)!}{(\ell-h)!(h-j)!(h-k)!}\,\beta^{h-j}\gamma^{h-k}\delta^{j+k}. \quad (*_{25})$$

In particular, if $\beta = \gamma = 0$, then $\alpha\delta = 1$, $A = \begin{pmatrix}\alpha & 0\\ 0 & 1/\alpha\end{pmatrix}$, and $t^{(\ell)}_{jk}(A)$ is the diagonal matrix with

$$t^{(\ell)}_{kk}(A) = \alpha^{-2k} = \delta^{2k}.$$

(2) If $\delta = 0$ and $\alpha \neq 0$, then

$$t^{(\ell)}_{jk}(A) = \begin{cases} 0 & \text{if } j+k > 0\\ \sqrt{\dfrac{(\ell-j)!(\ell-k)!}{(\ell+j)!(\ell+k)!}}\,\dfrac{(-1)^{\ell+j}\beta^{k-j}}{(-(j+k))!\,\alpha^{j+k}} & \text{if } j+k \leq 0. \end{cases} \quad (*_{26})$$

(3) If $\alpha = \delta = 0$, then we obtain an anti-diagonal matrix

$$t^{(\ell)}_{jk}(A) = \begin{cases} 0 & \text{if } j+k \neq 0\\ (-1)^{\ell-j}\gamma^{2j} & \text{if } k = -j. \end{cases} \quad (*_{27})$$

In particular, if $A = r_z(\pi/2) = \begin{pmatrix}0 & i\\ i & 0\end{pmatrix}$, then $t^{(\ell)}_{jk}(A) = 0$ if $j \neq k$ and $t^{(\ell)}_{j\,-j}(A) = i^{2\ell}$.

In Proposition 14.21, we should remember that if ℓ is a half-integer, then in $(*_{25})$ h is also a half-integer. Of course, if ℓ is a half-integer, then so are j, k.

Observe that

$$\begin{aligned}\frac{(\ell+h)!}{(\ell-h)!(h-j)!(h-k)!} &= \frac{(\ell+h)!}{(\ell-h)!(2h)!}\,\frac{(2h)!}{(h-k)!(h+k)!}\,\frac{(h+k)!}{(h-j)!(j+k)!}\,(j+k)!\\ &= \binom{\ell+h}{2h}\binom{2h}{h-k}\binom{h+k}{h-j}(j+k)!,\end{aligned}$$

with $\max(j,k) \leq h \leq \ell$. In particular, if $j = -k$, then

$$\frac{(\ell+h)!}{(\ell-h)!(h+k)!(h-k)!} = \binom{\ell+h}{2h}\binom{2h}{h-k}. \tag{†}$$

Another strategy is to use Taylor's formula. Recall that for polynomial $P(z)$ of degree m we have

$$P(z) = \sum_{j=0}^{m} \frac{P^{(j)}(0)}{j!} z^j,$$

where $P^{(k)}(0)$ is the value of the kth derivative of P at $z = 0$. Now by definition the kth column of the matrix $t^{(\ell)}(A)$ consists of the coordinates $t^{(\ell)}_{jk}(A)$ of

$$T_\ell(A)(\psi_k(z)) = \sum_{j=-\ell}^{\ell} t^{(\ell)}_{jk}(A)\psi_j(z) = \sum_{j=-\ell}^{\ell} \frac{t^{(\ell)}_{jk}(A)}{\sqrt{(\ell-j)!(\ell+j)!}} z^{\ell-j},$$

and since

$$T_\ell(A)(\psi_k(z)) = \frac{(\alpha z+\gamma)^{\ell-k}(\beta z+\delta)^{\ell+k}}{\sqrt{(\ell-k)!(\ell+k)!}}, \tag{$*_{28}$}$$

we deduce that $t^{(\ell)}_{jk}(A)/\sqrt{(\ell-j)!(\ell+j)!}$ is the coefficient of $z^{\ell-j}$ in the expansion of ($*_{28}$) in powers of z. Using Taylor's formula we obtain

$$t^{(\ell)}_{jk}(A) = \sqrt{\frac{(\ell+j)!}{(\ell-k)!(\ell+k)!(\ell-j)!}}\,\frac{d^{\ell-j}}{z^{\ell-j}}[(\alpha z+\gamma)^{\ell-k}(\beta z+\delta)^{\ell+k}]_{z=0}.$$

If $\alpha = 0$ or $\beta = 0$, the above formula simplifies. If $\alpha\beta \neq 0$, we substitute $y = \alpha(\beta z + \delta)$, then from $\alpha\delta - \beta\gamma = 1$ we get $\alpha z + \gamma = (y-1)/\beta$, so we obtain

$$t^{(\ell)}_{jk}(A) = \sqrt{\frac{(\ell+j)!}{(\ell-k)!(\ell+k)!(\ell-j)!}}\,\frac{\beta^{k-j}}{\alpha^{k+j}}\frac{d^{\ell-j}}{dy^{\ell-j}}[y^{\ell+k}(y-1)^{\ell-k}]_{y=\alpha\delta}.$$

Finally we let $z = y - 1$, and since $\alpha\delta - 1 = \beta\gamma$, we obtain

$$t^{(\ell)}_{jk}(A) = \sqrt{\frac{(\ell+j)!}{(\ell-k)!(\ell+k)!(\ell-j)!}}\,\frac{\beta^{k-j}}{\alpha^{k+j}}\frac{d^{\ell-j}}{dz^{\ell-j}}[z^{\ell-k}(z+1)^{\ell+k}]_{z=\beta\gamma}.$$

In summary we have the following result.

Proposition 14.22. *With respect to the orthonormal basis (ψ_k) of $\mathcal{P}^{\mathbb{C}}_\ell$, the entries in the matrix $t^{(\ell)}(A)$ are given by the formulae below.*

$$t^{(\ell)}_{jk}(A) = \sqrt{\frac{(\ell+j)!}{(\ell-k)!(\ell+k)!(\ell-j)!}}\,\frac{d^{\ell-j}}{z^{\ell-j}}[(\alpha z+\gamma)^{\ell-k}(\beta z+\delta)^{\ell+k}]_{z=0}. \tag{$*_{29}$}$$

If $\alpha\beta \neq 0$, then

$$t^{(\ell)}_{jk}(A) = \sqrt{\frac{(\ell+j)!}{(\ell-k)!(\ell+k)!(\ell-j)!}}\,\frac{\beta^{k-j}}{\alpha^{k+j}}\frac{d^{\ell-j}}{dz^{\ell-j}}[z^{\ell-k}(z+1)^{\ell+k}]_{z=\beta\gamma}. \tag{$*_{30}$}$$

14.9 Euler Angles Matrix Representations of T_ℓ

The "best" formula is obtained by using the Euler angles. We now restrict ourselves to $\mathbf{SU}(2)$, although it possible to handle the more general case; see Vilenkin [101] (Chapter III, Sections 3.3–3.9).

By Proposition 14.4 every matrix $q \in \mathbf{SU}(2)$ can be expressed as

$$q = u(\varphi, \theta, \psi) = r_x(\varphi/2) r_z(\theta/2) r_x(\psi/2) = \begin{pmatrix} e^{\frac{i\varphi}{2}} & 0 \\ 0 & e^{-\frac{i\varphi}{2}} \end{pmatrix} \begin{pmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2} \\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \begin{pmatrix} e^{\frac{i\psi}{2}} & 0 \\ 0 & e^{-\frac{i\psi}{2}} \end{pmatrix}$$

with

$$0 \le \varphi < 2\pi, \quad 0 \le \theta \le \pi, \quad -2\pi \le \psi < 2\pi.$$

Furthermore, if $\alpha\beta \neq 0$ and if we require that that $0 < \theta < \pi$, then φ, θ, ψ are unique. Since T_ℓ is a representation we have

$$T_\ell(q) = T_\ell(r_x(\varphi/2)) T_\ell(r_z(\theta/2)) T_\ell(r_x(\psi/2)).$$

We also proved that the polynomials in the basis $(\psi_k(z))$ are eigenvectors of $T_\ell(r_x(\varphi/2))$ and $T_\ell(r_x(\psi/2))$, namely (by $(*_1)$)

$$\begin{aligned} T_\ell(r_x(\varphi/2))\psi_k(z) &= e^{-ik\varphi}\psi_k(z) \\ T_\ell(r_x(\psi/2))\psi_k(z) &= e^{-ik\psi}\psi_k(z). \end{aligned}$$

Thus $T_\ell(r_x(\varphi/2))$ is represented by the diagonal matrix $t^{(\ell)}(r_x(\varphi/2))$ with $t^{(\ell)}_{kk}(r_x(\varphi/2)) = e^{-ik\varphi}$, and $T_\ell(r_x(\psi/2))$ is represented by the diagonal matrix $t^{(\ell)}(r_x(\psi/2))$ with $t^{(\ell)}_{kk}(r_x(\psi/2)) = e^{-ik\psi}$. Since

$$T_\ell(q) = T_\ell(r_x(\varphi/2)) T_\ell(r_z(\theta/2)) T_\ell(r_x(\psi/2)),$$

we have

$$t^{(\ell)}(q) = t^{(\ell)}(r_x(\varphi/2)) t^{(\ell)}(r_z(\theta/2)) t^{(\ell)}(r_x(\psi/2)),$$

and since $t^{(\ell)}(r_x(\varphi/2))$ and $t^{(\ell)}(r_x(\psi/2))$ are diagonal matrices, the (j,k) entry of the matrix $t^{(\ell)}(q)$ is

$$\begin{aligned} t^{(\ell)}_{jk}(q) &= t^{(\ell)}_{jj}(r_x(\varphi/2)) t^{(\ell)}_{jk}(r_z(\theta/2)) t^{(\ell)}_{kk}(r_x(\psi/2)) \\ &= e^{-ij\varphi}\, t^{(\ell)}_{jk}(r_z(\theta/2))\, e^{-ik\psi} = e^{-i(j\varphi + k\psi)}\, t^{(\ell)}\ell_{jk}(r_z(\theta/2)), \end{aligned}$$

that is,

$$t^{(\ell)}_{jk}(q) = e^{-i(j\varphi + k\psi)}\, t^{(\ell)}_{jk}(r_z(\theta/2)).$$

We record this important result below.

Proposition 14.23. *For any matrix $q \in \mathbf{SU}(2)$ expressed in terms of the Euler angles as $q = u(\varphi, \theta, \psi) = r_x(\varphi/2) r_z(\theta/2) r_x(\psi/2)$, with respect to the orthonormal basis (ψ_k) of $\mathcal{P}^{\mathbb{C}}_\ell$, we have*

$$t^{(\ell)}_{jk}(q) = e^{-i(j\varphi + k\psi)}\, t^{(\ell)}_{jk}(r_z(\theta/2)). \tag{$*_{31}$}$$

Thus we are left with finding an explicit expression for the matrix $t^{(\ell)}(r_z(\theta/2))$,

Definition 14.10. Define the matrix $t^{(\ell)}(\theta)$ as $t^{(\ell)}(\theta) = t^{(\ell)}(r_z(\theta/2))$, with

$$r_z(\theta/2) = \begin{pmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2} \\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}.$$

If $\theta = \pi$, then $r_z(\pi/2) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$, and by $(*_{27})$ we know that $t^{(\ell)}(\pi)$ is the anti-diagonal matrix with $t^{(\ell)}_{jk}(\pi) = 0$ if $j \neq k$ and $t^{(\ell)}_{j\,-j}(\pi) = i^{2\ell}$.

If $\theta = 0$, then $r_z(0)$ is the identity matrix I_2, and $t^{(\ell)}(0)$ is the identity matrix $I_{2\ell+1}$. If $0 \leq \theta < \pi$, then we can find the matrix $t^{(\ell)}(\theta)$ using Equation $(*_{25})$ in we which we set $\alpha = \delta = \cos\frac{\theta}{2} \neq 0$ (since $0 \leq \theta < \pi$), and $\beta = \gamma = i\sin\frac{\theta}{2}$. We obtain the following formula.

Proposition 14.24. *The elements of the matrix $t^{(\ell)}(\theta) = t^{(\ell)}(r_z(\theta/2))$ are given by the formula*

$$\begin{aligned} t^{(\ell)}_{jk}(\theta) &= i^{-(j+k)}\sqrt{\frac{(\ell-j)!(\ell-k)!}{(\ell+j)!(\ell+k)!}}\left(\cos\frac{\theta}{2}\right)^{j+k} \\ &\quad\times \sum_{h=\max(j,k)}^{\ell} \frac{(\ell+h)!\, i^{2h}}{(\ell-h)!(h-j)!(h-k)!}\left(\sin\frac{\theta}{2}\right)^{2h-(j+k)}. \end{aligned} \qquad (*_{32})$$

If ℓ is a half-integer, then h is also a half-integer. For $\theta = 0$, we must have $h = j = k$, and $t^{(\ell)}(0)$ is the identity matrix $I_{2\ell+1}$, as we already know.

If we assume that $0 < \theta < \pi$, then we obtain the following formula given in Vilenkin:

$$\begin{aligned} t^{(\ell)}_{jk}(\theta) &= i^{-(j+k)}\sqrt{\frac{(\ell-j)!(\ell-k)!}{(\ell+j)!(\ell+k)!}}\left(\cot\frac{\theta}{2}\right)^{j+k} \\ &\quad\times \sum_{h=\max(j,k)}^{\ell} \frac{(\ell+h)!\, i^{2h}}{(\ell-h)!(h-j)!(h-k)!}\left(\sin\frac{\theta}{2}\right)^{2h}. \end{aligned} \qquad (*_{33})$$

If we recall from (†) that if $j = -k$ then

$$\frac{(\ell+h)!}{(\ell-h)!(h+k)!(h-k)!} = \binom{\ell+h}{2h}\binom{2h}{h-k},$$

we obtain

$$t^{(\ell)}_{k-k}(\theta) = t^{(\ell)}_{-kk}(\theta) = \sum_{h=\max(-k,k)}^{\ell} \binom{\ell+h}{2h}\binom{2h}{h-k} i^{2h}\left(\sin\frac{\theta}{2}\right)^{2h}. \qquad (*_{34})$$

Even though this equation was derived assuming that $\theta < \pi$, it is still correct for $\theta = \pi$, namely the following equation holds

$$\sum_{h=\max(-k,k)}^{\ell} \binom{\ell+h}{2h}\binom{2h}{h-k} i^{2h} = i^{2\ell},$$

or equivalently, since we may assume that $k \geq 0$,

$$\sum_{h=k}^{\ell} (-1)^{\ell-h} \binom{\ell+h}{2h}\binom{2h}{h-k} = 1. \tag{††}$$

This can be proven using an identity due to Euler. As a first step we can prove that

$$\sum_{h=k}^{\ell} (-1)^{h-\ell} \binom{\ell+h}{2h}\binom{2h}{h-k} = \sum_{h=k}^{\ell} (-1)^{\ell-h} \binom{\ell+h}{\ell-k}\binom{\ell-k}{\ell-h}.$$

Next there are two cases depending on ℓ being an integer or a half integer. The second case reduces to the first by writing $\ell = ll + 1/2, k = kk + 1/2, h = hh + 1/2$ where ll, kk, hh are integers. The details are left as an exercise. If ℓ is an integer, then by changing the index of summation we have

$$\begin{aligned}\sum_{h=k}^{\ell} (-1)^{\ell-h} \binom{\ell+h}{\ell-k}\binom{\ell-k}{\ell-h} &= \sum_{h=0}^{\ell-k} (-1)^{\ell-h-k} \binom{\ell-k}{h}\binom{\ell+h+k}{\ell-k} \\ &= (-1)^N \sum_{h=0}^{N} (-1)^h \binom{N}{h}\binom{N+2k+h}{N},\end{aligned}$$

with $N = \ell - k$. At this stage we use an identity known as *Euler's finite difference formula*, namely

$$\sum_{h=0}^{n} (-1)^h \binom{n}{h}\binom{x+hy}{n} = (-1)^n y^n.$$

Remarkably the result is independent of x. Finally we let $n = N$, $x = N + 2k$ and $y = 1$ to match Euler's formula. We leave the details as an exercise.

14.10 Representations of $\mathbf{SL}(2, \mathbb{C})$ and $\mathbf{SU}(2)$ Using Finite Fourier Series

There is one more method for computing the matrix elements $t^{(\ell)}_{jk}(A)$ (where $A \in \mathbf{SL}(2, \mathbb{C})$) based on integration. The idea is to use another representing space for the representation T_ℓ, namely the vector space (of dimension $2\ell + 1$) of finite Fourier series

$$\Phi(e^{i\varphi}) = \sum_{k=-\ell}^{\ell} c_k e^{-ik\varphi},$$

with $c_k \in \mathbb{C}$. Observe that if $Q(z)$ is the polynomial of degree 2ℓ given by

$$Q(z) = \sum_{k=-\ell}^{\ell} c_k z^{\ell-k}$$

so that the powers appears in the order $z^{2\ell}, z^{2\ell-1}, \ldots, z, 1$, the Fourier series $\Phi(e^{i\varphi})$ with the same coefficients is given by

$$\Phi(e^{i\varphi}) = e^{-i\ell\varphi} Q(e^{i\varphi}).$$

Denote the space of Fourier series of dimension $2\ell+1$ as $\mathfrak{F}_\ell$. We would like to define a representation of $\mathbf{SL}(2,\mathbb{C})$ in $\mathfrak{F}_\ell$. By analogy with what we did when we defined the representation T_ℓ in the space $\mathcal{P}_\ell^{\mathbb{C}}$ from the representation U_ℓ in the space $\mathcal{P}_{2\ell}^{\mathbb{C}}(2)$, observe that

$$\begin{aligned}
e^{-i\ell\varphi}(ae^{i\varphi}+c)^\ell(be^{i\varphi}+d)^\ell \Phi\left(\frac{ae^{i\varphi}+c}{be^{i\varphi}+d}\right) &= e^{-i\ell\varphi}(ae^{i\varphi}+c)^\ell(be^{i\varphi}+d)^\ell \sum_{k=-\ell}^{\ell} c_k \left(\frac{ae^{i\varphi}+c}{be^{i\varphi}+d}\right)^{-k} \\
&= e^{-i\ell\varphi} \sum_{k=-\ell}^{\ell} c_k (ae^{i\varphi}+c)^{\ell-k}(be^{i\varphi}+d)^{\ell+k} \\
&= e^{-i\ell\varphi} S(e^{i\varphi}),
\end{aligned}$$

where $S(z)$ is the polynomial of degree 2ℓ given by

$$S(z) = \sum_{k=-\ell}^{\ell} c_k (az+c)^{\ell-k}(bz+d)^{\ell+k},$$

and so $e^{-i\ell\varphi}S(e^{i\varphi})$ is indeed a Fourier series in the space $\mathfrak{F}_\ell$. Consequently we make the following definition.

Definition 14.11. The map $\mathcal{T}_\ell \colon \mathbf{SL}(2,\mathbb{C}) \to \mathbf{GL}(\mathfrak{F}_\ell)$ is defined by

$$\mathcal{T}_\ell(A)(\Phi(e^{i\varphi})) = e^{-i\ell\varphi}(ae^{i\varphi}+c)^\ell(be^{i\varphi}+d)^\ell \Phi\left(\frac{ae^{i\varphi}+c}{be^{i\varphi}+d}\right) \tag{$\mathcal{T}_\ell$}$$

for every matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}(2,\mathbb{C})$ and every Fourier series $\Phi(e^{i\varphi}) \in \mathfrak{F}_\ell$.

It is easily verified that $\mathcal{T}_\ell \colon \mathbf{SL}(2,\mathbb{C}) \to \mathbf{GL}(\mathfrak{F}_\ell)$ is a representation. We also check immediately that the linear map defined on basis vectors by $z^{\ell-k} \mapsto e^{-ik\varphi}$ is an isomorphism between the vector spaces $\mathcal{P}_\ell^{\mathbb{C}}$ and $\mathfrak{F}_\ell$, and we make it an isometry by declaring that the inner product on $\mathfrak{F}_\ell$ is defined such that

$$\langle e^{-im\varphi}, e^{-in\varphi} \rangle = \begin{cases} 0 & \text{if } m \neq n \\ (\ell-m)!(\ell+m)! & \text{if } m = n. \end{cases}$$

Observe the very useful facts that

$$\langle e^{-im\varphi}, e^{-in\varphi}\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{-im\varphi}e^{in\varphi}d\varphi = 0, \quad m \neq n,$$

and

$$\langle e^{-im\varphi}, e^{-im\varphi}\rangle = \frac{(\ell-m)!\ell+m)!}{2\pi}\int_0^{2\pi} e^{-im\varphi}e^{im\varphi}\, d\varphi.$$

Therefore for any Fourier series $\Phi(e^{i\varphi}) \in \mathfrak{F}_\ell$, we have

$$\langle \Phi(e^{i\varphi}), e^{-im\varphi}\rangle = \frac{(\ell-m)!(\ell+m)!}{2\pi}\int_0^{2\pi} \Phi(e^{i\varphi})e^{im\varphi}\, d\varphi. \qquad (*_{35})$$

To show that the representation $\mathcal{T}_\ell\colon \mathbf{SL}(2,\mathbb{C}) \to \mathbf{GL}(\mathfrak{F}_\ell)$ is equivalent to the representation $T_\ell\colon \mathbf{SL}(2,\mathbb{C}) \to \mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$ we proceed as follows. Let $F\colon \mathcal{P}_\ell^{\mathbb{C}} \to \mathfrak{F}_\ell$ be the linear map given by

$$F((Q(z)) = e^{-i\ell\varphi}Q(e^{i\varphi}) = \Phi(e^{i\varphi}),$$

which on the basis $(z^{\ell-k})$ is given by $F(z^{\ell-k}) = e^{-ik\varphi}$. These equations show that F is an isomorphism. Moreover it is a unitary map because it preserves the hermitian inner product (we defined the hermitian product on $\mathfrak{F}_\ell$ to achieve this).

We need to prove that

$$F \circ T_\ell(A) = \mathcal{T}_\ell(A) \circ F \quad \text{for all } A \in \mathbf{SL}(2,\mathbb{C}).$$

For any $Q \in \mathcal{P}_\ell^{\mathbb{C}}$, if we write

$$Q(z) = \sum_{k=-\ell}^{\ell} c_k z^{\ell-k},$$

then we have

$$\begin{aligned} T_\ell(A)(Q(z)) &= (bz+d)^{2\ell}Q\left(\frac{az+c}{bz+d}\right) \\ &= \sum_{k=-\ell}^{\ell} c_k(bz+d)^{2\ell}\left(\frac{az+c}{bz+d}\right)^{\ell-k} \\ &= \sum_{k=-\ell}^{\ell} c_k(az+c)^{\ell-k}(bz+d)^{\ell+k} = S(z). \end{aligned}$$

Using the above equation we have

$$F(T_\ell(A)(Q(z))) = e^{-i\ell\varphi}S(e^{i\varphi}) = e^{-i\ell\varphi}\sum_{k=-\ell}^{\ell} c_k(ae^{i\varphi}+c)^{\ell-k}(be^{i\varphi}+d)^{\ell+k}.$$

We also proved earlier that

$$\begin{aligned}\mathcal{T}_\ell(A)(\Phi(e^{i\varphi})) &= e^{-i\ell\varphi}(ae^{i\varphi}+c)^\ell(be^{i\varphi}+d)^\ell\Phi\left(\frac{ae^{i\varphi}+c}{be^{i\varphi}+d}\right)\\ &= e^{-i\ell\varphi}S(e^{i\varphi}),\end{aligned}$$

with

$$S(z)=\sum_{k=-\ell}^{\ell} c_k(az+c)^{\ell-k}(bz+d)^{\ell+k}.$$

But by definition $\Phi(e^{i\varphi})=F(Q(z))$ for Q as above, so we proved that

$$F\circ T_\ell(A)=\mathcal{T}_\ell(A)\circ F\quad \text{for all } A\in\mathbf{SL}(2,\mathbb{C}),$$

which shows that $\mathcal{T}_\ell\colon \mathbf{SL}(2,\mathbb{C})\to\mathbf{GL}(\mathfrak{F}_\ell)$ is a representation equivalent to the representation $T_\ell\colon \mathbf{SL}(2,\mathbb{C})\to\mathbf{GL}(\mathcal{P}_\ell^{\mathbb{C}})$. Because the linear map $F\colon\mathcal{P}_\ell^{\mathbb{C}}\to\mathfrak{F}_\ell$ is unitary, we claim that the matrix of $T_\ell(A)$ in the basis $(\psi_k(z))$ is identical to the matrix of $\mathcal{T}_\ell(A)$ in the basis $(\psi'_k(\varphi))$ with $\psi'_k(\varphi)=\dfrac{e^{-ik\varphi}}{\sqrt{(\ell-k)!(\ell+k)!}}$. Recall that $\psi_k(z)=\dfrac{z^{\ell-k}}{\sqrt{(\ell-k)!(\ell+k)!}}$. This is because $\mathcal{T}_\ell(A)=F\circ T_\ell(A)\circ F^{-1}$, $F(\psi_k(z))=\psi'_k(\varphi)$, and the (j,k)-entry $t_{jk}^{(\ell)'}(A)$ of the matrix of $\mathcal{T}_\ell(A)$ in the basis $(\psi'_k(\varphi))$ is given by

$$t_{jk}^{(\ell)'}(A)=\langle\mathcal{T}_\ell(A)(\psi'_k(\varphi)),\,\psi'_j(\varphi)\rangle,$$

which is rewritten as

$$t_{jk}^{(\ell)'}(A)=\langle(F\circ T_\ell(A)\circ F^{-1})(\psi'_k(\varphi)),\,F(\psi_j(z))\rangle,$$

and then as

$$t_{jk}^{(\ell)'}(A)=\langle F(T_\ell(A)(\psi_k(z))),\,F(\psi_j(z))\rangle.$$

Since F is unitary, we obtain

$$t_{jk}^{(\ell)'}(A)=\langle F(T_\ell(A)(\psi_k(z))),\,F(\psi_j(z))\rangle=\langle T_\ell(A)(\psi_k(z)),\,\psi_j(z)\rangle=t_{jk}^{(\ell)}(A),$$

establishing our claim.

We now compute $\mathcal{T}_\ell(A)(\Phi(e^{i\varphi}))$ with $\Phi(e^{i\varphi})=(e^{i\varphi})^{-k}$ using Formula $(\mathcal{T}_\ell)$. We get

$$\begin{aligned}\mathcal{T}_\ell(A)(e^{-ik\varphi}) &= e^{-i\ell\varphi}(ae^{i\varphi}+c)^\ell(be^{i\varphi}+d)^\ell\left(\frac{ae^{i\varphi}+c}{be^{i\varphi}+d}\right)^{-k}\\ &= (ae^{i\varphi}+c)^{\ell-k}(be^{i\varphi}+d)^{\ell+k}\,e^{-i\ell\varphi}.\end{aligned}$$

As a consequence the matrix elements are given by

$$\begin{aligned}t_{jk}^{(\ell)}(A)=\langle\mathcal{T}_\ell(A)(\psi'_k(\varphi)),\,\psi'_j(\varphi)\rangle &= \frac{\langle\mathcal{T}_\ell(A)(e^{-ik\varphi}),\,e^{-ij\varphi}\rangle}{\sqrt{(\ell-j)!(\ell+j)!(\ell-k)!(\ell+k)!}}\\ &= \frac{\langle(ae^{i\varphi}+c)^{\ell-k}(be^{i\varphi}+d)^{\ell+k}\,e^{-i\ell\varphi},\,e^{-ij\varphi}\rangle}{\sqrt{(\ell-j)!(\ell+j)!(\ell-k)!(\ell+k)!}}.\end{aligned}$$

Using $(*_{35})$ we obtain the following result.

Proposition 14.25. *The matrix elements $t^{(\ell)}_{jk}(A)$ are given by the following formula:*

$$t^{(\ell)}_{jk}(A) = \frac{1}{2\pi}\sqrt{\frac{(\ell-j)!(\ell+j)!}{(\ell-k)!(\ell+k)!}}\int_0^{2\pi}(ae^{i\varphi}+c)^{\ell-k}(be^{i\varphi}+d)^{\ell+k}\, e^{i(j-\ell)\varphi}\, d\varphi. \tag{$*_{36}$}$$

By applying the above formula to the matrix

$$r_z(\theta/2) = \begin{pmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2} \\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$$

we obtain

$$t^{(\ell)}_{jk}(\theta) = \frac{1}{2\pi}\sqrt{\frac{(\ell-j)!(\ell+j)!}{(\ell-k)!(\ell+k)!}}\int_0^{2\pi}\left(\cos\frac{\theta}{2}e^{i\varphi}+i\sin\frac{\theta}{2}\right)^{\ell-k}\left(i\sin\frac{\theta}{2}e^{i\varphi}+\cos\frac{\theta}{2}\right)^{\ell+k} e^{i(j-\ell)\varphi}\, d\varphi$$

and since $e^{-i\ell\varphi} = e^{-\frac{i(\ell+k)\varphi}{2}}e^{-\frac{i(\ell-k)\varphi}{2}}$, the above formula is also written as stated below.

Proposition 14.26. *The matrix elements $t^{(\ell)}_{jk}(\theta)$ are given by the following formula:*

$$\begin{aligned} t^{(\ell)}_{jk}(\theta) &= \frac{1}{2\pi}\sqrt{\frac{(\ell-j)!(\ell+j)!}{(\ell-k)!(\ell+k)!}} \\ &\quad\times\int_0^{2\pi}\left(\cos\frac{\theta}{2}e^{\frac{i\varphi}{2}}+i\sin\frac{\theta}{2}e^{-\frac{i\varphi}{2}}\right)^{\ell-k}\left(i\sin\frac{\theta}{2}e^{\frac{i\varphi}{2}}+\cos\frac{\theta}{2}e^{-\frac{i\varphi}{2}}\right)^{\ell+k} e^{ij\varphi}\, d\varphi. \end{aligned} \tag{$*_{37}$}$$

For small values of ℓ, this equation is quite practical. For example, here is a list of the matrices $t^{\ell}(\theta)$ for $\ell = 0, 1/2, 1, 3/2$ as in Vilenkin [101] (Chapter III, Section 3.7).

$$t^{(0)}(\theta) = (1), \quad t^{(1/2)}(\theta) = r_z(\theta/2) = \begin{pmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2} \\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix},$$

$$t^{(1)}(\theta) = \begin{pmatrix} \cos^2\frac{\theta}{2} & \frac{i}{\sqrt{2}}\sin\frac{\theta}{2} & -\sin^2\frac{\theta}{2} \\ \frac{i}{\sqrt{2}}\sin\frac{\theta}{2} & \cos\theta & \frac{i}{\sqrt{2}}\sin\frac{\theta}{2} \\ -\sin^2\frac{\theta}{2} & \frac{i}{\sqrt{2}}\sin\frac{\theta}{2} & \cos^2\frac{\theta}{2} \end{pmatrix},$$

and

$$t^{(3/2)}(\theta) =$$
$$\begin{pmatrix} \cos^3\frac{\theta}{2} & i\sqrt{3}\sin\frac{\theta}{2}\cos^2\frac{\theta}{2} & -\sqrt{3}\sin^2\frac{\theta}{2}\cos\frac{\theta}{2} & -i\sin^3\frac{\theta}{2} \\ i\sqrt{3}\sin\frac{\theta}{2}\cos^2\frac{\theta}{2} & \cos^3\frac{\theta}{2}-2\cos\frac{\theta}{2}\sin^2\frac{\theta}{2} & 2i\cos^2\frac{\theta}{2}\sin\frac{\theta}{2}-i\sin^3\frac{\theta}{2} & -\sqrt{3}\sin^2\frac{\theta}{2}\cos\frac{\theta}{2} \\ -\sqrt{3}\sin^2\frac{\theta}{2}\cos\frac{\theta}{2} & 2i\cos^2\frac{\theta}{2}\sin\frac{\theta}{2}-i\sin^3\frac{\theta}{2} & \cos^3\frac{\theta}{2}-2\cos\frac{\theta}{2}\sin^2\frac{\theta}{2} & i\sqrt{3}\sin\frac{\theta}{2}\cos^2\frac{\theta}{2} \\ -i\sin^3\frac{\theta}{2} & -\sqrt{3}\sin^2\frac{\theta}{2}\cos\frac{\theta}{2} & i\sqrt{3}\sin\frac{\theta}{2}\cos^2\frac{\theta}{2} & \cos^3\frac{\theta}{2} \end{pmatrix}.$$

Wigner's d-matrices are obtained by using Formula ($*_{24}$) with $\alpha = \delta = \cos\frac{\theta}{2}$ and $\beta = \gamma = i\sin\frac{\theta}{2}$, we obtain

$$t^{(\ell)}_{jk}(\theta) = i^{j-k}\sqrt{(\ell-j)!(\ell+j)!(\ell-k)!(\ell+k)!}$$
$$\times \sum_{h=M}^{N}(-1)^h(h!(\ell-j-h)!(\ell+k-h)!(j-k+h)!)^{-1}\left(\cos\frac{\theta}{2}\right)^{2\ell+k-j-2h}\left(\sin\frac{\theta}{2}\right)^{2h+j-k} \qquad (*_{38})$$

with $M = \max(0, k-j)$, $N = \min(\ell-j, \ell+k)$. Formula ($*_{38}$) is the expression for the *Wigner d-matrices*, except that when working with quantum mechanics physicists prefer these functions to be real and the sign to be $(-1)^{h+j-k} = (-1)^h(i^{2(j-k)})$. Since $i^{2(j-k)} = (i^{j-k})^2$, Wigner's d-matrices are the matrices $d^\ell(\theta)$ given by

$$d^\ell_{jk}(\theta) = i^{j-k}t^{(\ell)}_{jk}(\theta).$$

Remark: *Wigner's D-matrices* are the matrices D^ℓ given by

$$D^\ell_{jk}(\varphi,\theta,\psi) = e^{-i(j\varphi+k\psi)}d^\ell_{jk}(\theta),$$

but beware that notation varies wildly, so φ,θ,ψ are also denoted α,β,γ, and j,k and ℓ might may appear under other names. Also note that

$$D^\ell = i^{j-k}\,t^{(\ell)}(A) = i^{j-k}\,t^{(\ell)}(u(\varphi,\theta,\psi)).$$

14.11 Matrix Elements of $T_\ell(q)$ and Jacobi Polynomials

In this section we assume again that $q \in \mathbf{SU}(2)$ is given in terms of the Euler angles as $q = u(\varphi,\theta,\psi) = r_x(\varphi/2)r_z(\theta/2)r_x(\psi/2)$. Since $\cos\theta = 2\cos^2\frac{\theta}{2} - 1$ and $\sin^2\frac{\theta}{2} + \cos^2\frac{\theta}{2} = 1$, for $0 \le \theta \le \pi$, we have $0 \le \cos\frac{\theta}{2} \le 1$ and $0 \le \sin\frac{\theta}{2} \le 1$, so

$$\cos\frac{\theta}{2} = \sqrt{\frac{1+\cos\theta}{2}} \qquad \sin\frac{\theta}{2} = \sqrt{\frac{1-\cos\theta}{2}} \qquad \cot\frac{\theta}{2} = \sqrt{\frac{1+\cos\theta}{1-\cos\theta}}, \qquad (*_{39})$$

with $\theta > 0$ for the third formula. Thus we see that $t^{(\ell)}_{jk}(\theta)$ is a function of $\cos\theta$ for $0 \le \theta < \pi$. Therefore there is a function $P^\ell_{jk}(z)$ such that

$$t^{(\ell)}_{jk}(\theta) = P^\ell_{jk}(\cos\theta), \quad 0 \le \theta < \pi,$$

and ($*_{31}$) is also written as

$$t^{(\ell)}_{jk}(q) = e^{-i(j\varphi+k\psi)}\,P^\ell_{jk}(\cos\theta).$$

By Equation ($*_{32}$) and the above trigonometric identities we obtain the following result.

Proposition 14.27. *The polynomial $P^\ell_{jk}(z)$ $(-1 < z \leq 1)$ given by*

$$P^\ell_{jk}(z) = i^{-(j+k)}\sqrt{\frac{(\ell-j)!(\ell-k)!}{(\ell+j)!(\ell+k)!}}\,\left(\frac{1+z}{2}\right)^{\frac{j+k}{2}} \times \sum_{h=\max(j,k)}^{\ell} \frac{(\ell+h)!\,i^{2h}}{(\ell-h)!(h-j)!(h-k)!}\,\left(\frac{1-z}{2}\right)^{\frac{2h-(j+k)}{2}} \tag{$*_{40}$}$$

has the property that

$$t^{(\ell)}_{jk}(\theta) = P^\ell_{jk}(\cos\theta), \quad 0 \leq \theta < \pi, \tag{$*_{41}$}$$

and

$$t^{(\ell)}_{jk}(q) = e^{-i(j\varphi+k\psi)}\,P^\ell_{jk}(\cos\theta). \tag{$*_{42}$}$$

If ℓ is a half-integer, then h is also a half-integer. It is understood that if $z = 1$, then $P^\ell_{jk}(1) = 1$ iff $j = k$, and $P^\ell_{jk}(1) = 0$ otherwise.

If $0 < \theta < \pi$, since $\alpha = \delta = \cos\frac{\theta}{2}$ and $\beta = \gamma = i\sin\frac{\theta}{2}$ are all nonzero, we obtain another formula from Equation $(*_{30})$ recalled below:

$$t^{(\ell)}_{jk}(q) = \sqrt{\frac{(\ell+j)!}{(\ell-k)!(\ell+k)!(\ell-j)!}}\,\frac{\beta^{k-j}}{\alpha^{k+j}}\,\frac{d^{\ell-j}}{dz^{\ell-j}}[z^{\ell-k}(z+1)^{\ell+k}]_{z=\beta\gamma}.$$

We perform the change of variable $z = (y-1)/2$, so $y = 2z+1$ and since $\beta\gamma = -\sin^2\frac{\theta}{2}$, the condition $z = \beta\gamma$ becomes $y = -2\sin^2\frac{\theta}{2} + 1 = \cos\theta$, and

$$\frac{d^{\ell-j}}{dy^{\ell-j}} = 2^{\ell-j}\,\frac{d^{\ell-j}}{dz^{\ell-j}}.$$

We also have $z = -\frac{1-y}{2}$, $z+1 = \frac{1+y}{2}$,

$$z^{\ell-k}(z+1)^{\ell+k} = (-1)^{\ell-k}2^{-2\ell}(1-y)^{\ell-k}(1+y)^{\ell+k}.$$

Using the trigonometric identities in Equations $(*_{39})$, we obtain

$$\begin{aligned}\frac{\beta^{k-j}}{\alpha^{k+j}} &= \frac{\left(i\sin\frac{\theta}{2}\right)^{k-j}}{\left(\cos\frac{\theta}{2}\right)^{k+j}} = i^{k-j}\left(\frac{1-\cos\theta}{2}\right)^{\frac{k-j}{2}}\left(\frac{1+\cos\theta}{2}\right)^{-\frac{(k+j)}{2}}\\ &= i^{k-j}\,2^{j}(1+\cos\theta)^{-\frac{(k+j)}{2}}(1-\cos\theta)^{\frac{k-j}{2}}.\end{aligned}$$

and with $z = \cos\theta$ we obtain the following result.

Proposition 14.28. *If $0 < \theta < \pi$, so that $-1 < z < 1$, then we have*

$$P^\ell_{jk}(z) = \frac{(-1)^{\ell-k}i^{k-j}}{2^\ell}\sqrt{\frac{(\ell+j)!}{(\ell-k)!(\ell+k)!(\ell-j)!}} \times (1+z)^{-\frac{(j+k)}{2}}(1-z)^{\frac{k-j}{2}}\,\frac{d^{\ell-j}}{dy^{\ell-j}}[(1-y)^{\ell-k}(1+y)^{\ell+k}]_{y=z}. \tag{$*_{43}$}$$

The polynomials $P^\ell_{jk}(z)$ enjoy some symmetry relations. For example, Formula $(*_{40})$ shows that

$$P^\ell_{jk}(z) = P^\ell_{kj}(z), \quad -1 < z \leq 1.$$

Since $r_z(\theta/2)$ and $r_z(\pi/2) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$ commute, it can be shown that

$$P^\ell_{jk}(z) = P^\ell_{-j\,-k}(z), \quad -1 < z \leq 1.$$

We leave the proof as an exercise. By Formula $(*_{43})$ we see immediately that

$$P^\ell_{jk}(-z) = i^{2(\ell-j-k)} P^\ell_{j\,-k}(z), \quad -1 < z < 1.$$

It is also immediately verified that

$$r_z(\theta/2)^{-1} = r_z(-\theta/2) = r_x(\pi/2) r_z(\theta/2) r_x(-\pi/2),$$

where

$$r_x(\pi/2) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad r_x(-\pi/2) = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix},$$

so by $(*_{42})$ (with $\varphi = \pi, \psi = -\pi$) we obtain

$$t^{(\ell)}_{jk}(-\theta) = (-1)^{k-j} P^\ell_{jk}(\cos\theta).$$

Formula $(*_{43})$ also reveals a relationship with the Jacobi polynomials.

Definition 14.12. The *Jacobi polynomials* $P^{\lambda,\mu}_h(z)$, with $\lambda, \mu \in \mathbb{R}$, $h \in \mathbb{N}$, are defined by the formula

$$P^{\lambda,\nu}_h(z) = \frac{(-1)^h}{2^h h!}(1-z)^{-\lambda}(1+z)^{-\mu}\,\frac{d^h}{dz^h}[(1-z)^{\lambda+h}(1+z)^{\mu+h}]. \tag{Ja}$$

To show that the $P^\ell_{jk}(z)$ are related to the Jacobi polynomials, taking a cue from Vilenkin we compute

$$D = 2^j i^{k-j}\sqrt{\frac{(\ell-k)!(\ell+k)!}{(\ell-j)!(\ell+j)!}}\,(1-z)^{\frac{k-j}{2}}(1+z)^{-\frac{(k+j)}{2}} P^\ell_{jk}(z). \tag{$*_{44}$}$$

We get

$$\begin{aligned} D &= 2^j i^{k-j}\sqrt{\frac{(\ell-k)!(\ell+k)!}{(\ell-j)!(\ell+j)!}}\,(1-z)^{\frac{k-j}{2}}(1+z)^{-\frac{(k+j)}{2}} \\ &\quad\times \frac{(-1)^{\ell-k} i^{k-j}}{2^\ell}\sqrt{\frac{(\ell+j)!}{(\ell-k)!(\ell+k)!(\ell-j)!}} \\ &\quad\times (1+z)^{-\frac{(k+j)}{2}}(1-z)^{\frac{k-j}{2}}\,\frac{d^{\ell-j}}{dy^{\ell-j}}[(1-y)^{\ell-k}(1+y)^{\ell+k}]_{y=z} \\ &= \frac{(-1)^{\ell-j}}{2^{\ell-j}(\ell-j)!}\,(1-z)^{k-j}(1+z)^{-(k+j)}\,\frac{d^{\ell-j}}{dz^{\ell-j}}[(1-z)^{\ell-k}(1+z)^{\ell+k}]. \end{aligned}$$

To match D with a Jacobi polynomial we need to find h, λ, μ such that

$$h = \ell - j, \qquad \lambda = -(k-j), \qquad \mu = k+j, \qquad \lambda + h = \ell - k, \qquad \mu + h = \ell + k.$$

We see that h, λ, μ are uniquely determined by

$$h = \ell - j, \quad \lambda = j - k, \quad \mu = k + j$$

and that the last two equations are also satisfied. Observe that λ and μ are integers. Thus we proved the following result.

Proposition 14.29. *The polynomials $P^\ell_{jk}(z)$ and the Jacobi polynomials are related by the equation*

$$P^{j-k,k+j}_{\ell-j}(z) = 2^j i^{k-j} \sqrt{\frac{(\ell-k)!(\ell+k)!}{(\ell-j)!(\ell+j)!}}\,(1-z)^{\frac{k-j}{2}}(1+z)^{-\frac{(k+j)}{2}} P^\ell_{jk}(z). \qquad (*_{45})$$

As we noted earlier, if ℓ is a half integer then j and k cannot be zero. If ℓ is an integer, then $j = 0$ or $k = 0$ is allowed, and so $\lambda = 0$ and $\mu = 0$ are also allowed. In this case the Jacobi polynomial $P^{0,0}_\ell(z)$, simply denoted as $P_\ell(z)$, is given by

$$P_\ell(z) = \frac{(-1)^\ell}{2^\ell \ell!}\frac{d^\ell}{dz^\ell}(1-z^2)^\ell,$$

or equivalently

$$P_\ell(z) = \frac{1}{2^\ell \ell!}\frac{d^\ell}{dz^\ell}(z^2-1)^\ell.$$

This is a *Legendre* polynomial.

Similarly, if ℓ is an integer, then for $k = 0$ the polynomials $P^\ell_{m0}(z)$ are related to polynomials $P^m_\ell(z)$ known as the associated Legendre polynomials.

Definition 14.13. The *Legendre polynomial* $P_\ell(z)$ are defined by

$$P_\ell(z) = \frac{(-1)^\ell}{2^\ell \ell!}\frac{d^\ell}{dz^\ell}(1-z^2)^\ell,$$

and the *associated Legendre polynomials* are defined by

$$P^m_\ell(z) = \frac{(-1)^{m+\ell}}{2^\ell \ell!}(1-z^2)^{\frac{m}{2}}\frac{d^{m+\ell}}{dz^{m+\ell}}(1-z^2)^\ell = (-1)^m(1-z^2)^{\frac{m}{2}}\frac{d^m}{dz^m}P_\ell(z),$$

with $\ell, m \in \mathbb{N}$.

Some authors omit the sign $(-1)^m$ in the definition of the associated Legendre polynomials. We see immediately that

$$P^{\ell}_{00}(z) = P_\ell(z). \tag{$*_{46}$}$$

It is not hard to show that

$$P^k_\ell(z) = i^k \sqrt{\frac{(\ell+k)!}{(\ell-k)!}}\, P^{\ell}_{k0}(z). \tag{$*_{47}$}$$

See Vilenkin [101] (Chapter III, Section 3.9). Since by $(*_{42})$ we have

$$t^{(\ell)}_{k0}(q) = e^{-ik\varphi}\, P^{\ell}_{k0}(\cos\theta),$$

we obtain

$$t^{(\ell)}_{k0}(q) = i^{-k}\sqrt{\frac{(\ell-k)!}{(\ell+k)!}}\, e^{-ik\varphi}\, P^k_\ell(\cos\theta), \quad -\ell \leq k \leq \ell. \tag{$*_{48}$}$$

Recall that ℓ is an integer.

Definition 14.14. The function $t^{(\ell)}_{k0}(q)$ (with $q = r_x(\varphi/2)r_z(\theta/2)$), which does not depend on ψ, can be viewed as a function on the sphere S^2, and is denoted $Y_{\ell k}(\varphi, \theta)$, with $0 \leq \varphi < 2\pi$ and $0 \leq \theta < \pi$. The function $Y_{\ell k}(\varphi, \theta)$ is called a *spherical function*.

Here we are following Vilenkin [101] (Chapter III, Section 3.10). Up to a constant including the term i^{-k}, $Y_{\ell k}(\varphi, \theta)$ is the classical spherical harmonic (unfortunately) denoted $Y^m_\ell(\theta, \varphi)$ and called *Laplace spherical harmonic* by Dieudonné. In special case where $k = 0$ the function $t^{(\ell)}_{00}(q) = P_\ell(\cos\theta)$ depends only on θ and is called a *zonal spherical function*.

More properties of the Legendre and Jacobi polynomials and functional relations and generating functions for the functions $P^{\ell}_{jk}(z)$, can be found in Vilenkin [101], Chapter III, Sections 3-5.

We will now derive an explicit formula for an invariant Haar measure on $\mathbf{SU}(2)$ in terms of the Euler angles φ, θ, ψ. Then since the representations $T_\ell \colon \mathbf{SU}(2) \to \mathbf{GL}(\mathcal{P}^{\mathbb{C}}_\ell)$ form a complete set of irreducible representations of $\mathbf{SU}(2)$, they are equivalent to the representation M_ρ of Peter–Weyl I (with $\rho = \ell$), so we will able to obtain a Hilbert sum decomposition of $\mathrm{L}^2(\mathbf{SU}(2))$ in terms of the functions $t^{(\ell)}_{jk}(q)$, with $q \in \mathbf{SU}(2)$. We will also obtain a Hilbert sum decomposition of $\mathrm{L}^2(S^2)$.

14.12 Integration on $\mathbf{SU}(2)$ and $\mathbf{SO}(3)$

As a first step we will need to derive a formula for an $\mathbf{SU}(2)$-invariant volume form on $\mathbf{SU}(2)$ as a pull-back of the $\mathbf{SO}(4)$-invariant volume form ω_{S^3} on S^3. The reader may want to review volume forms and integration on manifolds before reading this section. These topics are covered in Gallier and Quaintance [39] (Chapter 4 and 6).

Definition 14.15. The bijection $\Sigma\colon \mathbb{H} \to \mathbb{R}^4$ from the space $\mathbb{H}$ of quaternions to $\mathbb{R}^4$ is defined as follows: for every quaternion $A \in \mathbb{H}$, with

$$A = \begin{pmatrix} a+ib & c+id \\ -c+id & a-ib \end{pmatrix}, \quad a,b,c,d \in \mathbb{R},$$

we have

$$\Sigma(A) = (a,b,c,d),$$

where, as usual, we view (a,b,c,d) as a column vector. The bijection Σ restricts to a bijection $\Sigma\colon \mathbf{SU}(2) \to S^3$ from $\mathbf{SU}(2)$ to the sphere S^3 (in $\mathbb{R}^4$).

It is clear that the map $\Sigma\colon \mathbf{SU}(2) \to S^3$ is a homeomorphism. In fact, it is a smooth diffeomorphism. We will compute the tangent map $d\Sigma_A\colon T_A\mathbf{SU}(2) \to T_{\Sigma(A)}S^3$, with $A \in \mathbf{SU}(2)$. Then we will use Σ to define a volume form ω on $\mathbf{SU}(2)$ by pulling back a volume form ω_{S^3} on S^3.

We warn our readers that in this section we do not follow our usual notational convention that a unit quaternion, an element of $\mathbf{SU}(2)$, is denoted by a lower-case letter, typically q. Since we also need to denote points on the sphere S^3, to avoid potential confusion we denote unit quaternions using capital letters, A, A', *etc.*

The volume form ω_{S^3} on S^3 is $\mathbf{SO}(4)$-invariant but, to prove that $\omega = \Sigma^*\omega_{S^3}$ is $\mathbf{SU}(2)$-invariant we need to understand how the left (or right) action of $\mathbf{SU}(2)$ on itself translates into an action on S^3. Here we use the "ancient" fact that left and right translation in $\mathbf{SU}(2)$ translate into a rotation in $\mathbb{R}^4$ restricted to S^3 *via* Σ. This fact can be found as far back as Veblen and Young [100] and also in Gallier [37] (Chapter 9).

Given two matrices $A, A' \in \mathbf{SU}(2)$, if $\Sigma(A) = (a,b,c,d)$ and $\Sigma(A') = (a',b',c,d')$, by multiplying the matrices A and A', we obtain the following identities:

$$\Sigma(AA') = \Sigma(L_A A') = \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix} \begin{pmatrix} a' \\ b' \\ c' \\ d' \end{pmatrix}$$

and

$$\Sigma(AA') = \Sigma(R_{A'} A) = \begin{pmatrix} a' & -b' & -c' & -d' \\ b' & a' & d' & -c' \\ c' & -d' & a' & b' \\ d' & c' & -b' & a' \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}.$$

Definition 14.16. Let $M(L_A)$ and $M(R_{A'})$ be the matrices

$$M(L_A) = \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}, \quad M(R_{A'}) = \begin{pmatrix} a' & -b' & -c' & -d' \\ b' & a' & d' & -c' \\ c' & -d' & a' & b' \\ d' & c' & -b' & a' \end{pmatrix}. \tag{M1}$$

In summary, we proved that

$$\Sigma(AA') = \Sigma(L_A A') = M(L_A)\Sigma(A') = \Sigma(R_{A'}A) = M(R_{A'})\Sigma(A). \tag{M2}$$

Proposition 14.30. *If A and A' are unit quaternions, then $M(L_A)$ and $M(R_{A'})$ belong to $\mathbf{SO}(4)$; that is, they are rotation matrices.*

Proof. Observe that the columns (and the rows) of the matrices $M(L_A)$ and $M(R_{A'})$ are orthogonal. Thus, when A and A' are unit quaternions, both $M(L_A)$ and $M(R_{A'})$ are orthogonal matrices. Furthermore, it is obvious that $M(L_{A^*}) = M(L_A)^\top$, the transpose of $M(L_A)$, and similarly, $M(R_{(A')^*}) = M(R_{A'})^\top$. Since $AA^* = (a^2+b^2+c^2+d^2)I_2 = N(A)I_2$, the matrix $M(L_A)M(L_A)^\top$ is the diagonal matrix $N(A)I$ (where I is the identity 4×4 matrix), and similarly the matrix $M(R_{A'})M(R_{A'})^\top$ is the diagonal matrix $N(A')I$. Since $M(L_A)$ and $M(L_A)^\top$ have the same determinant, we deduce that $\det(M(L_A))^2 = N(A)^4$, and thus $\det(M(L_A)) = \pm N(A)^2$. However, it is obvious that one of the terms in $\det(M(L_A))$ is a^4, and thus

$$\det(M(L_A)) = (a^2+b^2+c^2+d^2)^2.$$

This shows that when A is a unit quaternion, $M(L_A) \in \mathbf{SO}(4)$, that is, $M(L_A)$ a rotation matrix, and similarly when A' is a unit quaternion, $M(R_{A'}) \in \mathbf{SO}(4)$ (see Veblen and Young [100]). ☐

We also need an explicit formula for the derivative $d\Sigma_A\colon T_A\mathbf{SU}(2) \to T_{\Sigma(A)}S^3$.

Proposition 14.31. *For all $A \in \mathbf{SU}(2)$ and all $Y \in T_A\mathbf{SU}(2)$, we have*

$$d\Sigma_A(Y) = \Sigma(Y). \tag{$d\Sigma$}$$

Proof. Since the tangent space $T_A\mathbf{SU}(2)$ is equal to $A\mathfrak{su}(2)$, every $Y \in T_A\mathbf{SU}(2)$ is of the form $Y = A\theta X$ for some $X \in \mathfrak{su}(2)$ given by

$$X = \begin{pmatrix} ix_1 & x_2 + ix_3 \\ -x_2 + ix_3 & -ix_1 \end{pmatrix}$$

with $x_1^2 + x_2^2 + x_3^2 = 1$ and some $\theta \in \mathbb{R}$, and we have a curve

$$c(t) = Ae^{t\theta X}$$

such that $c(0) = A$ and $c'(0) = A\theta X = Y$. But

$$e^{t\theta X} = \cos(t\theta)I_2 + \sin(t\theta)X,$$

so by the chain rule

$$\begin{aligned}
d\Sigma_A(Y) &= d\Sigma_{c(0)}(c'(0)) = (\Sigma(c(t)))'_{t=0} = \\
&= (\Sigma(A(\cos(t\theta)I_2 + \sin(t\theta)X))'_{t=0} \\
&= (M(L_A)(\Sigma(\cos(t\theta)I_2 + \sin(t\theta)X))'_{t=0} \\
&= (M(L_A)(\cos(t\theta),\, \sin(t\theta)x_1,\, \sin(t\theta)x_2,\, \sin(t\theta)x_3))'_{t=0} \\
&= (M(L_A)(-\theta\sin(t\theta),\, \theta\cos(t\theta)x_1,\, \theta\cos(t\theta)x_2,\, \theta\cos(t\theta)x_3))_{t=0} \\
&= M(L_A)(0,\, \theta x_1,\, \theta x_2,\, \theta x_3) \\
&= M(L_A)\Sigma(\theta X) = \Sigma(A\theta X) = \Sigma(AA^{-1}Y) = \Sigma(Y).
\end{aligned}$$

In summary, for all $A \in \mathbf{SU}(2)$ and all $Y \in T_A\mathbf{SU}(2)$, we have

$$d\Sigma_A(Y) = \Sigma(Y),$$

as claimed. $\square$

Since Σ is linear, the pull-back $\Sigma^*\omega_{S^3}$ is given by

$$\Sigma^*(\omega_{S^3})_A(Y) = (\omega_{S^3})_{\Sigma(A)}(d\Sigma_A(Y)) = (\omega_{S^3})_{\Sigma(A)}(\Sigma(Y)).$$

Definition 14.17. The *volume form ω on* $\mathbf{SU}(2)$ is defined as $\omega = \Sigma^*(\omega_{S^3})$; that is, for all $A \in \mathbf{SU}(2)$ and all $Y \in T_A\mathbf{SU}(2)$, we have

$$\omega_A(Y) = (\omega_{S^3})_{\Sigma(A)}(\Sigma(Y)). \tag{ω}$$

Since $\Sigma(A) = \Sigma(AI_2) = M(L_A)\Sigma(I_2) = M(L_A)e_1$, since $Y = A\theta X$ with $X \in \mathfrak{su}(2)$ and $M(L_A)$ is a rotation matrix, $\Sigma(A) = M(L_A)e_1$ and $\Sigma(Y) = M(L_A)\Sigma(\theta X)$ are indeed orthogonal because e_1 and $\Sigma(\theta X)$ are orthogonal since θX has no real part.

Since the volume form ω_{S^3} is invariant under $\mathbf{SO}(4)$, we can prove that the 3-form ω is $\mathbf{SU}(2)$-invariant.

Proposition 14.32. *The volume form ω is invariant under* $\mathbf{SU}(2)$.

Proof. First we verify left-invariance. We need to prove that $L_A^*\omega = \omega$ for all $A \in \mathbf{SU}(2)$. Since L_A given by $L_A(A') = AA'$ is linear, we have $(dL_A)_{A'} = L_A$ for all $A' \in \mathbf{SU}(2)$ and similarly since $L_{M(L_A)}$ is linear, $d(L_{M(L_A)})_Q = L_{M(L_A)}$ for all $Q \in \mathbf{SO}(3)$, and so

$$\begin{aligned}
(L_A^*\omega)_{A'}(Y) &= \omega_{L_A(A')}((dL_A)_{A'}(Y)) = \omega_{L_A(A')}(L_A(Y)) \\
&= (\omega_{S^3})_{\Sigma(AA')}(\Sigma(L_A(Y))) \\
&= (\omega_{S^3})_{M(L_A)\Sigma(A')}(M(L_A)\Sigma(Y)) \\
&= (\omega_{S^3})_{M(L_A)\Sigma(A')}(d(L_{M(L_A)})_{\Sigma(A')}(\Sigma(Y))),
\end{aligned}$$

and since $M(L_A) \in \mathbf{SO}(4)$ and ω_{S^3} is invariant under $\mathbf{SO}(4)$, we conclude that

$$(L_A^*\omega)_{A'}(Y) = (\omega_{S^3})_{M(L_A)\Sigma(A')}(d(L_{M(L_A)})_{\Sigma(A')}(\Sigma(Y))) = (\omega_{S^3})_{\Sigma(A')}(\Sigma(Y)) = \omega_{A'}(Y),$$

as claimed. Next we verify right-invariance, which means that we need to check that $R_A^*\omega = \omega$. Since R_A given by $R_A(A') = A'A$ is linear, we have $(dR_A)_{A'} = R_A$ for all $A' \in \mathbf{SU}(2)$ and similarly since $R_{M(R_A)}$ is linear, $d(R_{M(R_A)})_Q = R_{M(R_A)}$ for all $Q \in \mathbf{SO}(3)$, and so

$$\begin{aligned}(R_A^*\omega)_{A'}(Y) &= \omega_{R_A(A')}((dR_A)_{A'}(Y)) = \omega_{R_A(A')}(R_A(Y)) \\ &= (\omega_{S^3})_{\Sigma(A'A)}(\Sigma(R_A(Y))) \\ &= (\omega_{S^3})_{M(R_A)\Sigma(A')}(M(R_A)\Sigma(Y)) \\ &= (\omega_{S^3})_{M(R_A)\Sigma(A')}(d(R_{M(R_A)})_{\Sigma(A')}(\Sigma(Y))),\end{aligned}$$

and since $M(R_A) \in \mathbf{SO}(4)$ and ω_{S^3} is invariant under $\mathbf{SO}(4)$, we conclude that

$$(R_A^*\omega)_{A'}(Y) = (\omega_{S^3})_{M(R_A)\Sigma(A')}(d(R_{M(R_A)})_{\Sigma(A')}(\Sigma(Y))) = (\omega_{S^3})_{\Sigma(A')}(\Sigma(Y)) = \omega_{A'}(Y),$$

establishing right-invariance. $\square$

It is a standard result of differential geometry that the restriction ω_{S^3} of the differential 3-form $\widetilde{\omega}$ on $\mathbb{R}^4$ to S^3 given by

$$\widetilde{\omega}_p = adx_2 \wedge dx_3 \wedge dx_4 - bdx_1 \wedge dx_3 \wedge dx_4 + cdx_1 \wedge dx_2 \wedge dx_4 - dx_1 \wedge dx_2 \wedge dx_3 \qquad (\widetilde{\omega})$$

is a volume form on S^3, with

$$(\omega_{S^3})_p(u_1, u_2, u_3) = \det(p, u_1, u_2, u_3), \qquad (\widetilde{\omega}_{S^3})$$

where $p = (a, b, c, d) \in S^3$ and $u_1, u_2, u_3 \in T_pS^3$. See Gallier and Quaintance [39] (Chapter 6). Invariance under $\mathbf{SO}(4)$ follows from the fact that the determinant is preserved under $\mathbf{SO}(4)$. Consequently, using the diffeomorphism $\Sigma\colon \mathbf{SU}(2) \to S^3$, we obtain the volume form $\omega_{\mathbf{SU}(2)}$ on $\mathbf{SU}(2)$, for short ω, given by

$$\begin{aligned}(\omega_{\mathbf{SU}(2)})_A = \omega_A = adx_2 \wedge dx_3 \wedge dx_4 - bdx_1 \wedge dx_3 \wedge dx_4 \\ + cdx_1 \wedge dx_2 \wedge dx_4 - dx_1 \wedge dx_2 \wedge dx_3,\end{aligned} \qquad (\omega_{\mathbf{SU}(2)})$$

where

$$A = \begin{pmatrix} a+ib & c+id \\ -c+id & a-ib \end{pmatrix}.$$

In the above formula we abused notation because we identified $T_A\mathbf{SU}(2)$ with $\mathbb{R}^4$ using Σ. As we showed earlier, the 3-form ω is $\mathbf{SU}(2)$-invariant. After all this work, it is nice to see that "things" are basically the same as if we were dealing with S^3, but some justifications are required, in particular invariance under $\mathbf{SU}(2)$. After all, $\mathbf{SU}(2)$ consists of *complex* matrices, but $\mathbf{SO}(4)$ consists of *real* matrices.

Definition 14.18. Let $\Omega \subseteq \mathbb{R}^3$ be the open subset

$$\Omega = (0, 2\pi) \times (0, \pi) \times (-2\pi, 2\pi). \qquad (\Omega)$$

By Proposition 14.4, the map $u\colon \Omega \to \mathbf{SU}(2)$ given by

$$\begin{aligned} u(\varphi,\theta,\psi) &= \begin{pmatrix} e^{\frac{i\varphi}{2}} & 0 \\ 0 & e^{-\frac{i\varphi}{2}} \end{pmatrix} \begin{pmatrix} \cos\frac{\theta}{2} & i\sin\frac{\theta}{2} \\ i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \begin{pmatrix} e^{\frac{i\psi}{2}} & 0 \\ 0 & e^{-\frac{i\psi}{2}} \end{pmatrix} \\ &= \begin{pmatrix} \cos\frac{\theta}{2}\, e^{\frac{i(\varphi+\psi)}{2}} & i\sin\frac{\theta}{2}\, e^{\frac{i(\varphi-\psi)}{2}} \\ i\sin\frac{\theta}{2}\, e^{-\frac{i(\varphi-\psi)}{2}} & \cos\frac{\theta}{2}\, e^{-\frac{i(\varphi+\psi)}{2}} \end{pmatrix} \end{aligned}$$

is a diffeomorphism onto an open subset of $\mathbf{SU}(2)$ that omits a subset of measure zero in its range. We need to find the pull-back $u^*\omega$ of the volume form ω on $\mathbf{SU}(2)$, and since $\omega = \Sigma^*(\omega_{S^3})$, we need to find

$$\omega_\Omega = u^*\omega = u^*(\Sigma^*(\omega_{S^3})) = (\Sigma \circ u)^*(\omega_{S^3}).$$

Definition 14.19. Let $\Phi\colon \Omega \to S^3$ be the composed map $\Phi = \Sigma \circ u$ from Ω onto an open subset of S^3, given by

$$\Phi(\varphi,\theta,\psi) = (\Phi_1(\varphi,\theta,\psi),\ \Phi_2(\varphi,\theta,\psi),\ \Phi_3(\varphi,\theta,\psi),\ \Phi_4(\varphi,\theta,\psi)),$$

with

$$\begin{aligned} \Phi_1(\varphi,\theta,\psi) &= \cos\frac{\theta}{2}\cos\frac{(\varphi+\psi)}{2} & \Phi_2(\varphi,\theta,\psi) &= \cos\frac{\theta}{2}\sin\frac{(\varphi+\psi)}{2} \\ \Phi_3(\varphi,\theta,\psi) &= -\sin\frac{\theta}{2}\sin\frac{(\varphi-\psi)}{2} & \Phi_4(\varphi,\theta,\psi) &= \sin\frac{\theta}{2}\cos\frac{(\varphi-\psi)}{2}. \end{aligned}$$

The map Φ is a diffeomorphism.

Definition 14.20. Let ω_Ω be the pull-back form $\omega_\Omega = \Phi^*\omega_{S^3}$.

The pull-back form $\omega_\Omega = \Phi^*\omega_{S^3}$ is a volume form on Ω, and from the point of integration, since a only subset of measure zero is omitted we can use it to define integration on $\mathbf{SU}(2)$.

Proposition 14.33. *The volume form ω_Ω is given by*

$$\omega_\Omega = \frac{1}{8}\sin\theta\, d\theta \wedge d\varphi \wedge d\psi.$$

Proof. We need to compute

$$\begin{aligned} \Phi^*\omega_{S^3} = {} & \Phi_1(\varphi,\theta,\psi) d\Phi_2 \wedge d\Phi_3 \wedge d\Phi_4 - \Phi_2(\varphi,\theta,\psi) d\Phi_1 \wedge d\Phi_3 \wedge d\Phi_4 \\ & + \Phi_3(\varphi,\theta,\psi) d\Phi_1 \wedge d\Phi_2 \wedge d\Phi_4 - \Phi_4(\varphi,\theta,\psi) d\Phi_1 \wedge d\Phi_2 \wedge d\Phi_3. \end{aligned}$$

It turns out that the computation is simpler if we let $\sigma = \dfrac{\varphi+\psi}{2}$ and $\tau = \dfrac{\varphi-\psi}{2}$. Then we have

$$d\sigma \wedge d\tau = \frac{1}{4}(d\varphi + d\psi) \wedge (d\varphi - d\psi) = -\frac{1}{2} d\varphi \wedge d\psi.$$

Since

$$\begin{aligned} d\Phi_1 &= \frac{\partial \Phi_1}{\partial \sigma} d\sigma + \frac{\partial \Phi_1}{\partial \theta} d\theta + \frac{\partial \Phi_1}{\partial \tau} d\tau & \qquad d\Phi_2 &= \frac{\partial \Phi_2}{\partial \sigma} d\sigma + \frac{\partial \Phi_2}{\partial \theta} d\theta + \frac{\partial \Phi_2}{\partial \tau} d\tau \\ d\Phi_4 &= \frac{\partial \Phi_3}{\partial \sigma} d\sigma + \frac{\partial \Phi_3}{\partial \theta} d\theta + \frac{\partial \Phi_3}{\partial \tau} d\tau & \qquad d\Phi_4 &= \frac{\partial \Phi_4}{\partial \sigma} d\sigma + \frac{\partial \Phi_4}{\partial \theta} d\theta + \frac{\partial \Phi_4}{\partial \tau} d\tau \end{aligned}$$

after a moment of reflexion we see that

$$\Phi^* \omega_{S^3} = \begin{vmatrix} \Phi_1 & \dfrac{\partial \Phi_1}{\partial \sigma} & \dfrac{\partial \Phi_1}{\partial \theta} & \dfrac{\partial \Phi_1}{\partial \tau} \\ \Phi_2 & \dfrac{\partial \Phi_2}{\partial \sigma} & \dfrac{\partial \Phi_2}{\partial \theta} & \dfrac{\partial \Phi_2}{\partial \tau} \\ \Phi_3 & \dfrac{\partial \Phi_3}{\partial \sigma} & \dfrac{\partial \Phi_3}{\partial \theta} & \dfrac{\partial \Phi_3}{\partial \tau} \\ \Phi_4 & \dfrac{\partial \Phi_4}{\partial \sigma} & \dfrac{\partial \Phi_4}{\partial \theta} & \dfrac{\partial \Phi_4}{\partial \tau} \end{vmatrix} d\sigma\, d\theta\, d\tau.$$

We find that

$$\begin{aligned} \frac{\partial \Phi_1}{\partial \theta} &= -\frac{1}{2} \sin \frac{\theta}{2} \cos \sigma & \qquad \frac{\partial \Phi_1}{\partial \sigma} &= -\cos \frac{\theta}{2} \sin \sigma & \qquad \frac{\partial \Phi_1}{\partial \tau} &= 0 \\ \frac{\partial \Phi_2}{\partial \theta} &= -\frac{1}{2} \sin \frac{\theta}{2} \sin \sigma & \qquad \frac{\partial \Phi_2}{\partial \sigma} &= \cos \frac{\theta}{2} \cos \sigma & \qquad \frac{\partial \Phi_2}{\partial \tau} &= 0 \\ \frac{\partial \Phi_3}{\partial \theta} &= -\frac{1}{2} \cos \frac{\theta}{2} \sin \tau & \qquad \frac{\partial \Phi_3}{\partial \sigma} &= 0 & \qquad \frac{\partial \Phi_3}{\partial \tau} &= -\sin \frac{\theta}{2} \cos \tau \\ \frac{\partial \Phi_4}{\partial \theta} &= \frac{1}{2} \cos \frac{\theta}{2} \cos \tau & \qquad \frac{\partial \Phi_4}{\partial \sigma} &= 0 & \qquad \frac{\partial \Phi_4}{\partial \tau} &= -\sin \frac{\theta}{2} \sin \tau. \end{aligned}$$

Since

$$d\sigma \wedge d\theta \wedge d\tau = -d\theta \wedge d\sigma \wedge d\tau = \frac{1}{2} d\theta \wedge d\varphi \wedge d\psi,$$

we obtain

$$\Phi^* \omega_{S^3} = \frac{1}{4} \cos \frac{\theta}{2} \sin \frac{\theta}{2} \begin{vmatrix} \cos \frac{\theta}{2} \cos \sigma & -\sin \frac{\theta}{2} \cos \sigma & -\sin \sigma & 0 \\ \cos \frac{\theta}{2} \sin \sigma & -\sin \frac{\theta}{2} \sin \sigma & \cos \theta & 0 \\ -\sin \frac{\theta}{2} \sin \tau & -\cos \frac{\theta}{2} \sin \tau & 0 & -\cos \tau \\ \sin \frac{\theta}{2} \cos \tau & \cos \frac{\theta}{2} \cos \tau & 0 & -\sin \tau \end{vmatrix} d\theta \wedge d\varphi \wedge d\psi.$$

Observe that the matrix corresponding to the determinant is orthogonal, so

$$\Phi^* \omega_{S^3} = \pm \frac{1}{4} \cos \frac{\theta}{2} \sin \frac{\theta}{2}\, d\theta \wedge d\varphi \wedge d\psi = \pm \frac{1}{8} \sin \theta\, d\theta \wedge d\varphi \wedge d\psi.$$

In fact, it can be verified that the determinant has the value $+1$, so we get

$$\omega_\Omega = \Phi^*\omega_{S^3} = \frac{1}{8}\sin\theta\, d\theta\wedge d\varphi\wedge d\psi,$$

as claimed. $\square$

Finally, given a continuous function $f\colon \Omega\to\mathbb{C}$ the integral $\int_\Omega f\omega_\Omega$ is defined by

$$\int_\Omega f\omega_\Omega = \frac{1}{8}\int_{-2\pi}^{2\pi}\int_0^{2\pi}\int_0^{\pi} f(\varphi,\theta,\psi)\sin\theta\, d\theta\, d\varphi\, d\psi.$$

Since this integral yields the volume $2\pi^2$ (for $f\equiv 1$), we normalize the measure ν associated with ω_Ω so that $\int_\Omega 1\omega_\Omega = 1$. Since $\Sigma\colon \mathbf{SU}(2)\to S^3$ is a diffeomorphism we also have $\omega_{S^3} = (\Sigma^{-1})^*\omega$, and so for any continuous function $f\colon \mathbf{SU}(2)\to\mathbb{C}$,

$$\int_{\mathbf{SU}(2)} f\omega = \int_{S^3}(f\circ\Sigma^{-1})\omega_{S^3},$$

and since for any continuous function $f\colon S^3\to\mathbb{C}$ we have

$$\int_{S^3} f\omega_{S^3} = \int_\Omega (f\circ\Phi)\omega_\Omega = \int_\Omega (f\circ\Sigma\circ u)\omega_\Omega,$$

we obtain

$$\int_{\mathbf{SU}(2)} f\omega = \int_{S^3}(f\circ\Sigma^{-1})\omega_{S^3} = \int_\Omega (f\circ\Sigma^{-1}\circ\Sigma\circ u)\omega_\Omega = \int_\Omega (f\circ u)\omega_\Omega.$$

In summary we obtained the following result.

Proposition 14.34. *The pull-back volume form $\omega_\Omega = \Phi^*\omega_{S^3}$ on Ω is given by*

$$\omega_\Omega = \frac{1}{8}\sin\theta\, d\theta\wedge d\varphi\wedge d\psi. \tag{ω_Ω}$$

For a continuous function $f\colon\Omega\to\mathbb{C}$, the integral $\int_\Omega f\omega_\Omega$ is given by

$$\int_\Omega f\omega_\Omega = \frac{1}{16\pi^2}\int_{-2\pi}^{2\pi}\int_0^{2\pi}\int_0^{\pi} f(\varphi,\theta,\psi)\sin\theta\, d\theta\, d\varphi\, d\psi. \tag{INT-Ω}$$

For any continuous function $f\colon \mathbf{SU}(2)\to\mathbb{C}$, the integral $\int_{\mathbf{SU}(2)} f\omega = \int_\Omega (f\circ u)\omega_\Omega$ is given by

$$\int_{\mathbf{SU}(2)} f\omega = \frac{1}{16\pi^2}\int_{-2\pi}^{2\pi}\int_0^{2\pi}\int_0^{\pi} f(u(\varphi,\theta,\psi))\sin\theta\, d\theta d\varphi d\psi. \tag{INT-$\mathbf{SU}(2)$}$$

For any continuous function $f\colon S^3\to\mathbb{C}$, the integral $\int_{S^3} f\omega = \int_\Omega (f\circ\Phi)\omega_\Omega$ is given by

$$\int_{S^3} f\omega_{S^3} = \frac{1}{16\pi^2}\int_{-2\pi}^{2\pi}\int_0^{2\pi}\int_0^{\pi} f(\Phi(\varphi,\theta,\psi))\sin\theta\, d\theta\, d\varphi\, d\psi. \tag{INT-S^3}$$

We also write $\int_{\mathbf{SU}(2)} f\, d\nu$ for $\int_{\mathbf{SU}(2)} f\omega$.

Remark: Formula (INT$-\Omega$) is stated in Vilenkin [101] (Chapter III, Section 6) and in Kosmann-Schwarzbach [59], see Exercise 5.6.

It is remarkable that we can also obtain the normalized Haar measure on $\mathbf{SO}(3)$ in terms of the Euler angles from the normalized Haar measure on $\mathbf{SU}(2)$ *without any additional computation*. Recall that as a corollary of Proposition 14.4, the map $\rho \circ u_0$ from $[0, 2\pi) \times [0, \pi] \times [0, 2\pi)$ to $\mathbf{SO}(3)$ is surjective, where $\rho\colon \mathbf{SU}(2) \to \mathbf{SO}(3)$ is a surjective homomorphism whose kernel is $\{-I, I\}$ and u_0 is the restriction of $u\colon [0, 2\pi) \times [0, \pi] \times [-2\pi, 2\pi) \to \mathbf{SU}(2)$ to $[0, 2\pi) \times [0, \pi] \times [0, 2\pi)$.

Definition 14.21. Let $\Omega_0 \subseteq \mathbb{R}^3$ be the open subset

$$\Omega_0 = (0, 2\pi) \times (0, \pi) \times (0, 2\pi), \tag{Ω_0}$$

a proper open subset of $\Omega = (0, 2\pi) \times (0, \pi) \times (-2\pi, 2\pi)$, and let $u_0\colon \Omega_0 \to \mathbf{SU}(2)$ be the restriction of $u\colon \Omega \to \mathbf{SU}(2)$ to Ω_0.

Definition 14.22. The map $R_0\colon \Omega_0 \to \mathbf{SO}(3)$ is given by

$$R_0(\varphi, \theta, \psi) = R_x(\varphi) R_z(\theta) R_x(\psi),$$

or more explicitly,

$$R_0(\varphi, \theta, \psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{pmatrix}.$$

Proposition 14.4 implies that u is injective on $\Omega = (0, 2\pi) \times (0, \pi) \times (-2\pi, 2\pi)$, and thus u_0 is injective on Ω_0.

Definition 14.23. Let $U_0 = u_0(\Omega_0)$ be the image of Ω_0 by u_0, an open subset of $\mathbf{SU}(2)$.

Proposition 14.35. *The restriction ρ_0 of ρ to U_0 is injective. As a consequence, the map $R_0\colon \Omega_0 \to \mathbf{SO}(3)$ is also injective.*

Proof. This is because if $\rho(q_1) = \rho(q_2)$ for some $q_1, q_2 \in \Omega_0$ such that $q_1 \neq q_2$, then $q_2 = -q_1$, but if $q_1 = r_x(\varphi/2) r_z(\theta/2) r_x(\psi/2)$, then

$$\begin{aligned} r_x(\varphi/2) r_z(\theta/2) r_x((\psi - 2\pi)/2) &= r_x(\varphi/2) r_z(\theta/2) r_x(\psi/2 - \pi) \\ &= -r_x(\varphi/2) r_z(\theta/2) r_x(\psi/2) = -q_1, \end{aligned}$$

so $q_2 = r_x(\varphi/2) r_z(\theta/2) r_x((\psi - 2\pi)/2)$ with $-2\pi < \psi < 0$ contradicting the fact that $q_2 \in \Omega_0$. $\square$

As consequence if we let $V_0 = \rho_0(U_0) \subseteq \mathbf{SO}(3)$, the bijective map $\rho_0 \colon U_0 \to V_0$ has an inverse $s_0 \colon V_0 \to U_0$.

Definition 14.24. Let $V_0 = \rho_0(U_0) \subseteq \mathbf{SO}(3)$ and let $s_0 \colon V_0 \to U_0$ be the inverse of $\rho_0 \colon U_0 \to V_0$. Also let Σ_0 be the composition $\Sigma_0 = \Sigma \circ s_0$.

Consider the commutative diagram

$$\begin{tikzcd}
\Omega_0 \arrow[r, "u_0"] \arrow[dr, "R_0"'] & U_0 \subseteq \mathbf{SU}(2) \arrow[r, "\Sigma"] \arrow[d, shift right, "\rho_0"'] & S^3 \\
& V_0 \subseteq \mathbf{SO}(3) \arrow[u, shift right, "s_0"'] \arrow[ur, "\Sigma_0"'] &
\end{tikzcd}$$

Since s_0 is a diffeomorphism, we can pull back the volume form ω on $\mathbf{SU}(2)$ (really on U_0) to $\mathbf{SO}(3)$ (really V_0).

Proposition 14.36. *The 3-form $s_0^*\omega$ is $\mathbf{SO}(3)$-invariant.*

Proof. We check left-invariance, leaving right-invariance as an exercise. For this, for any $Q, R \in V_0 \subseteq \mathbf{SO}(3)$ and any $Y \in T_R\mathbf{SO}(3)$, we compute

$$\begin{aligned}
(L_Q^*(s_0^*\omega))_R(Y) &= ((s_0 \circ L_Q)^*\omega)_R(Y) = \omega_{(s_0 \circ L_Q)(R)}(d(s_0 \circ L_Q)_R(Y)) \\
&= \omega_{s_0(QR)}(d(s_0 \circ L_Q)_R(Y)) \\
&= \omega_{s_0(Q)s_0(R)}(d(s_0 \circ L_Q)_R(Y)).
\end{aligned}$$

But since $(s_0 \circ L_Q)(R) = s_0(QR) = s_0(Q)s_0(R) = L_{s_0(Q)}(s_0(R))$, we see that $d(s_0 \circ L_Q)_R = d(L_{s_0(Q)})_{s_0(R)} \circ (ds_0)_R$, so we have

$$(L_Q^*(s_0^*\omega))_R(Y) = \omega_{s_0(Q)s_0(R)}(d(s_0 \circ L_Q)_R(Y)) = \omega_{L_{s_0(Q)}(s_0(R))}(d(L_{s_0(Q)})_{s_0(R)}((ds_0)_R(Y))).$$

Since ω is left-invariant, we obtain

$$\begin{aligned}
(L_Q^*(s_0^*\omega))_R(Y) &= \omega_{L_{s_0(Q)}(s_0(R))}(d(L_{s_0(Q)})_{s_0(R)}((ds_0)_R(Y))) \\
&= \omega_{(s_0(R)}((ds_0)_R(Y)) = (s_0^*\omega)_R(Y),
\end{aligned}$$

establishing left-invariance. ∎

It follows that $s_0^*\omega$ is an invariant volume form on V_0, and thus a volume form $\omega_{\mathbf{SO}(3)}$ on $\mathbf{SO}(3)$ up to a set of measure zero.

Remark: It can be shown that there is an invariant volume form $\omega_{\mathbf{SO}(3)}$ on $\mathbf{SO}(3)$ such that $\omega_{\mathbf{SO}(3)}$ and the volume form $\omega_{\mathbf{SU}(2)}$ on $\mathbf{SU}(2)$ are related by $\omega_{\mathbf{SU}(2)} = \rho^*\omega_{\mathbf{SO}(3)}$, where ρ is the covering homomorphism $\rho \colon \mathbf{SU}(2) \to \mathbf{SO}(3)$. This is because $\mathbf{SO}(3)$ is orientable. The

proof is essentially the same as the proof that $\mathbb{RP}^n$ is orientable iff n is odd; see Madsen and Tornehave [67], Example 9.19. Thus the volume form $s_0^*\omega$ on $V_0 \subseteq \mathbf{SO}(3)$ is a piece of the volume form $\omega_{\mathbf{SO}(3)}$ corresponding to a section s_0 of ρ. The complement of the domain of s_0 is a subset of measure zero in $\mathbf{SO}(3)$. The volume form $\omega_{\mathbf{SO}(3)}$ defines the Haar measure on $\mathbf{SO}(3)$, and up to a subset of measure zero, so does $s_0^*\omega$.

The pull-back $\omega_{\Omega_0} = R_0^*(s_0^*\omega) = (s_0 \circ R_0)^*\omega$ of the volume form $s_0^*\omega$ by R_0 is the volume form on Ω_0 that we seek. However, the commutativity of the above diagram and the fact that by definition $\omega = \Sigma^*\omega_{S^3}$ show that

$$\omega_{\Omega_0} = (s_0 \circ R_0)^*\omega = u_0^*\omega = u_0^*(\Sigma^*\omega_{S^3}) = (\Sigma \circ u_0)^*\omega_{S^3}.$$

Definition 14.25. Let $\Phi_0 \colon \Omega_0 \to S^3$ be the composed map $\Phi_0 = \Sigma \circ u_0$ and let ω_{Ω_0} be the volume form on Ω_0 defined as the pull-back $\Phi_0^*\omega_{S^3}$ of ω_{S^3}.

Since u_0 is the restriction of u to Ω_0, we conclude that $\omega_{\Omega_0} = \Phi_0^*\omega_{S^3}$ is given as in the $\mathbf{SU}(2)$ case, by

$$\omega_{\Omega_0} = \frac{1}{8}\sin\theta\, d\theta \wedge d\varphi \wedge d\psi.$$

Proposition 14.37. *The pull-back volume form $\omega_{\Omega_0} = \Phi_0^*\omega_{S^3}$ on Ω_0 is given by*

$$\omega_{\Omega_0} = \frac{1}{8}\sin\theta\, d\theta \wedge d\varphi \wedge d\psi. \tag{ω_{Ω_0}}$$

For a continuous function $f \colon \Omega_0 \to \mathbb{C}$, the integral $\int_{\Omega_0} f\omega_{\Omega_0}$ is given by

$$\int_{\Omega_0} f\omega_{\Omega_0} = \frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi} f(\varphi,\theta,\psi)\sin\theta\, d\theta d\varphi d\psi. \tag{INT-Ω_0}$$

For a continuous function $f \colon \mathbf{SO}(3) \to \mathbb{C}$, the integral $\int_{\mathbf{SO}(3)} f\omega_{\mathbf{SO}(3)}$ is given by

$$\int_{\mathbf{SO}(3)} f\omega_{\mathbf{SO}(3)} = \frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi} f(R_0(\varphi,\theta,\psi))\sin\theta\, d\theta\, d\varphi\, d\psi. \tag{INT-$\mathbf{SO}(3)$}$$

We also write $\int_{\mathbf{SO}(3)} f\, d\nu_0$ for $\int_{\mathbf{SO}(3)} f\omega_{\mathbf{SO}(3)}$. Observe that the measure ν_0 associated with ω_{Ω_0} is already normalized.

Remark: Formula (INT$-\Omega_0$) is stated in Kosmann-Schwarzbach [59], see Exercise 5.5.

14.13 Series Expansion of Functions in $\mathrm{L}^2(\mathbf{SU}(2))$ Using the t^ℓ_{jk}

In the preceding sections we computed explicitly several matrix representations $t^{(\ell)}(q)$ ($q \in \mathbf{SU}(2)$) for the irreducible representations $T_\ell \colon \mathbf{SU}(2) \to \mathbf{U}(\mathcal{P}_\ell^{\mathbb{C}})$ with respect to an invariant

hermitian inner product on $\mathcal{P}_\ell^{\mathbb{C}}$. In terms of the general results presented in Sections 13.1–13.3, $\rho = \ell$, $n_\rho = 2\ell + 1$, $M_\ell(q) = t^{(\ell)}(q)$, and since

$$M_\ell(q) = \left(\frac{1}{n_\ell} m_{ij}^{(\ell)}(q)\right),$$

the functions $m_{ij}^{(\ell)}(q)$ are given by $m_{ij}^{(\ell)}(q) = (2\ell+1)t_{ij}^{(\ell)}(q)$, where ℓ ranges through the set $R = \{0, 1/2, 1, 3/2, 2, 5/2, 3, \ldots\}$ of all nonnegative integer and half-integer values. By Peter–Weyl I (Theorem 13.2), the $n_\ell^2 = (2\ell+1)^2$ functions $\frac{1}{\sqrt{n_\ell}} m_{ij}^{(\ell)} = \sqrt{2\ell+1}\, t_{ij}^{(\ell)}$ in the matrix $\sqrt{2\ell+1}\, t^{(\ell)}$ form an orthonormal basis of the minimal two-sided ideal $\mathfrak{a}_\ell$ arising in the Hilbert sum

$$\mathrm{L}^2(\mathbf{SU}(2)) = \bigoplus_\ell \mathfrak{a}_\ell,$$

and thus the family of function

$$\left(\sqrt{2\ell+1}\, t_{ij}^{(\ell)}\right)_{-\ell \le i,j \le \ell,\, \ell \in R}$$

with $R = \{0, 1/2, 1, 3/2, 2, \ldots\}$, is a Hilbert basis of $\mathrm{L}^2(\mathbf{SU}(2))$. It follows that every function $f \in \mathrm{L}^2(\mathbf{SU}(2))$ can be expressed as the Fourier series

$$f(q) = \sum_{\ell \in R} \sum_{j=-\ell}^{\ell} \sum_{k=-\ell}^{\ell} \alpha_{jk}^\ell\, t_{jk}^{(\ell)}(q), \quad q \in \mathbf{SU}(2), \tag{FS1}$$

where the Fourier coefficients are given by

$$\alpha_{jk}^\ell = (2\ell+1) \int_{\mathbf{SU}(2)} f(q) \overline{t_{jk}^{(\ell)}(q)}\, d\nu, \tag{FC1}$$

where ν is the normalized Haar measure on $\mathbf{SU}(2)$. Using the Euler angles, Proposition 14.27 (in particular, $(*_{42})$), Proposition 14.34, and the fact that $\overline{P_{jk}^\ell(\cos\theta)} = (-1)^{j-k} P_{jk}^\ell(\cos\theta)$ (left as an exercise), we obtain the following series expansion for the functions in $\mathrm{L}^2(\mathbf{SU}(2))$.

Proposition 14.38. *Every function $f \in \mathrm{L}^2(\mathbf{SU}(2))$ expressed in terms of the Euler angles $(0 \le \varphi < 2\pi, 0 \le \theta < \pi, -2\pi \le \psi < 2\pi)$ can be written as the Fourier series*

$$f(u(\varphi, \theta, \psi)) = \sum_{\ell \in R} \sum_{j=-\ell}^{\ell} \sum_{k=-\ell}^{\ell} \alpha_{jk}^\ell\, e^{-i(j\varphi + k\psi)} P_{jk}^{(\ell)}(\cos\theta), \tag{FS2}$$

where the Fourier coefficients are given by

$$\alpha_{jk}^\ell = \frac{(-1)^{j-k}(2\ell+1)}{16\pi^2} \int_{-2\pi}^{2\pi} \int_0^{2\pi} \int_0^{\pi} f(u(\varphi, \theta, \psi)) e^{i(j\varphi + k\psi)} P_{jk}^\ell(\cos\theta) \sin\theta\, d\theta\, d\varphi\, d\psi. \tag{FC2}$$

Remark: The Parseval identity is given by

$$\sum_\ell \frac{1}{2\ell+1}\sum_{j=-\ell}^{\ell}\sum_{k=-\ell}^{\ell}|\alpha^\ell_{jk}|^2 = \frac{1}{16\pi^2}\int_{-2\pi}^{2\pi}\int_0^{2\pi}\int_0^{\pi}|f(u(\varphi,\theta,\psi))|^2\sin\theta\,d\theta\,d\varphi\,d\psi. \tag{PS1}$$

Vilenkin investigates the expansion in Fourier series for two subspaces $\mathfrak{L}^2_k$ and ${}_j\mathfrak{L}^2$ of $\mathrm{L}^2(\mathbf{SU}(2))$.

Definition 14.26. The subspace $\mathfrak{L}^2_k$ consists of the functions $f\in \mathrm{L}^2(\mathbf{SU}(2))$ such that

$$f(qH(\psi)) = e^{-ik\psi}f(q),\quad q\in\mathbf{SU}(2),$$

for all $H(\psi)$ in the subgroup Ω_x of $\mathbf{SU}(2)$ given by

$$\Omega_x = \left\{ H(t) = r_x(t/2) = \begin{pmatrix} e^{\frac{it}{2}} & 0\\ 0 & e^{-\frac{it}{2}}\end{pmatrix} \;\middle|\; 0\le t\le 2\pi\right\}. \tag{Ω_x}$$

It is easy to see that $\mathfrak{L}^2_k$ consists of the functions $f\in \mathrm{L}^2(\mathbf{SU}(2))$ such that

$$f(u(\varphi,\theta,\psi)) = e^{-ik\psi}f(u(\varphi,\theta,0))$$

for all φ,θ,ψ.

Definition 14.27. The subspace ${}_j\mathfrak{L}^2$ consists of the functions $f\in \mathrm{L}^2(\mathbf{SU}(2))$ such that

$$f(H(\varphi)q) = e^{-ij\varphi}f(q),\quad q\in\mathbf{SU}(2),$$

for all $H(\varphi)$ in the subgroup Ω_x of $\mathbf{SU}(2)$.

It is easy to see that ${}_j\mathfrak{L}^2$ consists of the functions $f\in \mathrm{L}^2(\mathbf{SU}(2))$ such that

$$f(u(\varphi,\theta,\psi)) = e^{-ij\varphi}f(u(0,\theta,\psi))$$

for all φ,θ,ψ.

Observe that the functions $t^{(\ell)}_{jk}(q)$ (with q expressed in terms of the Euler angles) for k fixed,

$$t^{(\ell)}_{jk}(q) = e^{-ik\psi}e^{-ij\varphi}P^\ell_{jk}(\cos\theta)$$

for $\ell = |k|, |k|+1,\ldots,|k|+m,\ldots,$ $-\ell\le j\le \ell$, belong to $\mathfrak{L}^2_k$. In fact, it is shown in Vilenkin [101] (Chapter III, Section 6.4) that these functions form an orthogonal basis of $\mathfrak{L}^2_k$, more precisely, every function $f\in\mathfrak{L}^2_k$ can written as the Fourier series

$$f(u(\varphi,\theta,\psi)) = e^{-ik\psi}\sum_{\ell=|k|}^{\infty}\sum_{j=-\ell}^{\ell}\alpha^\ell_j\, e^{-ij\varphi}P^{(\ell)}_{jk}(\cos\theta), \tag{FS3}$$

where the Fourier coefficients are given by

$$\alpha_j^\ell = \frac{(-1)^{j-k}(2\ell+1)}{4\pi}\int_0^{2\pi}\int_0^{\pi} f(u(\varphi,\theta,0))e^{ij\varphi}P_{jk}^\ell(\cos\theta)\sin\theta\, d\theta\, d\varphi. \tag{FC3}$$

In particular, for functions $f \in \mathrm{L}^2(\mathbf{SU}(2))$ such that

$$f(qH) = f(q), \quad \text{for all } q \in \mathbf{SU}(2) \text{ and all } H \in \Omega_x,$$

namely, functions which do not depend on the Euler angle ψ, we have

$$f(u(\varphi,\theta,0)) = \sum_{\ell=0}^{\infty}\sum_{j=-\ell}^{\ell} \alpha_j^\ell\, e^{-ij\varphi}P_{j0}^{(\ell)}(\cos\theta), \tag{FS4}$$

with

$$\alpha_j^\ell = \frac{(-1)^{j}(2\ell+1)}{4\pi}\int_0^{2\pi}\int_0^{\pi} f(u(\varphi,\theta,0))e^{ij\varphi}P_{j0}^\ell(\cos\theta)\sin\theta\, d\theta\, d\varphi. \tag{FC4}$$

It is shown in Vilenkin [101] (Chapter III, Section 3.9) that $(*_{47})$ implies that we have

$$P_\ell^{-j}(z) = (-1)^j\frac{(\ell-j)!}{(\ell+j)!}P_\ell^j(z),$$

so we obtain an expansion in terms of the associated Legendre functions $P_\ell^j(\cos\theta)$,

$$f(u(\varphi,\theta,0)) = \sum_{\ell=0}^{\infty}\sum_{j=-\ell}^{\ell} \beta_\ell^j\, e^{-ij\varphi}P_\ell^j(\cos\theta), \tag{FS5}$$

with

$$\beta_\ell^j = \frac{(2\ell+1)}{4\pi}\frac{(\ell-j)!}{(\ell+j)!}\int_0^{2\pi}\int_0^{\pi} f(u(\varphi,\theta,0))e^{ij\varphi}P_\ell^j(\cos\theta)\sin\theta\, d\theta\, d\varphi. \tag{FC5}$$

Similary, it can be shown (see Vilenkin [101], Chapter III, Section 6.4) that every function $f \in {}_j\mathfrak{L}^2$ can written as the Fourier series

$$f(u(\varphi,\theta,\psi)) = e^{-ij\varphi}\sum_{\ell=|j|}^{\infty}\sum_{k=-\ell}^{\ell} \alpha_k^\ell\, e^{-ik\psi}P_{jk}^{(\ell)}(\cos\theta), \tag{FS6}$$

where the Fourier coefficients are given by

$$\alpha_k^\ell = \frac{(-1)^{j-k}(2\ell+1)}{8\pi}\int_{-2\pi}^{2\pi}\int_0^{\pi} f(u(0,\theta,\psi))e^{ik\psi}P_{jk}^\ell(\cos\theta)\sin\theta\, d\theta\, d\psi. \tag{FC6}$$

In particular, for functions $f \in \mathrm{L}^2(\mathbf{SU}(2))$ such that

$$f(Hq) = f(q), \quad \text{for all } q \in \mathbf{SU}(2) \text{ and all } H \in \Omega_x,$$

namely, functions which do not depend on the Euler angle φ, we have

$$f(u(0,\theta,\psi)) = \sum_{\ell=0}^{\infty}\sum_{k=-\ell}^{\ell} \beta_\ell^k\, e^{-ik\psi} P_\ell^k(\cos\theta), \tag{FS7}$$

with

$$\beta_\ell^k = \frac{(2\ell+1)}{8\pi}\frac{(\ell-k)!}{(\ell+k)!}\int_{-2\pi}^{2\pi}\int_0^{\pi} f(u(0,\theta,\psi))e^{ik\psi}P_\ell^k(\cos\theta)\sin\theta\, d\theta\, d\psi. \tag{FC7}$$

Fourier expansion formulae for the functions in ${}_j\mathfrak{L}^2 \cap \mathfrak{L}^2_k$ can also be obtained, as well as formulae for functions $f \in \mathrm{L}^2(\mathbf{SU}(2))$ such that

$$f(H_1 q H_2) = f(q), \quad \text{for all } q \in \mathbf{SU}(2) \text{ and all } H_1, H_2 \in \Omega_x,$$

namely functions that do not depend on the Euler angles φ and ψ, but we leave these as exercises (see Vilenkin [101], Chapter III, Section 6.4).

The expansion in Fourier series of the function in $\mathrm{L}^2(\mathbf{SU}(2))$ that are independent of ψ yield a Fourier series expansion of functions in $\mathrm{L}^2(S^2)$. This is because $\mathbf{SU}(2)/\Omega_x$ is a homogeneous space homeomorphic to S^2 and the functions $f \in \mathrm{L}^2(\mathbf{SU}(2))$ such that $f(qH) = f(q)$ for all $q \in \mathbf{SU}(2)$ and all $H \in \Omega_x$ correspond bijectively to the functions in $\mathrm{L}^2(S^2)$. The group $\mathbf{SU}(2)$ acts on the sphere S^2 by rotations, which means that for any skew-hermitian matrix

$$X = \begin{pmatrix} ix & y+iz \\ -y+iz & -ix \end{pmatrix}, \quad (x,y,z) \in S^2$$

and any $q \in \mathbf{SU}(2)$, we have the action

$$q \cdot X = qXq^*.$$

Since this action is a rotation of S^2, it is transitive. The stabilizer of $e_1 = (1,0,0)$ is the subgroup consisting of all unit quaternions

$$q = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}$$

such that

$$\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}\begin{pmatrix} \overline{\alpha} & -\beta \\ \overline{\beta} & \alpha \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

or equivalently

$$\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix},$$

which means that

$$\begin{pmatrix} i\alpha & -i\beta \\ -i\overline{\beta} & -i\overline{\alpha} \end{pmatrix} = \begin{pmatrix} i\alpha & i\beta \\ i\overline{\beta} & -i\overline{\alpha} \end{pmatrix},$$

and so $\beta = 0$. Therefore the stabilizer of $e_1 = (1, 0, 0)$ is indeed the subgroup Ω_x. From Section 14.3, since every unit quaternion q can be factored as

$$q = r_x(\varphi/2) r_z(\theta/2) r_x(\psi/2),$$

with $r_x(\varphi/2), r_x(\psi/2) \in \Omega_x$, we see that a representative of the left coset $q\Omega_x$ is given by

$$r_x(\varphi/2) r_z(\theta/2).$$

Therefore the points of S^2 are the orbit of $e_1 = (1, 0, 0)$ under all rotations $r_x(\varphi/2) r_z(\theta/2)$, and from Section 14.3, since the corresponding rotation matrices are

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

by reading of the first column of the matrix Q, we see that the corresponding orbit points on the sphere S^2 have coordinates

$$(\cos\theta,\ \sin\theta\cos\varphi,\ \sin\theta\sin\varphi).$$

According to the physical convention, the spherical coordinates of a point p with respect to the (azimuthal) angle φ measured from the x-axis in the xy-plane and (polar) angle θ measured from the z-axis in the plane containing the z-axis an passing through the point p are given by

$$(\sin\theta\cos\varphi,\ \sin\theta\sin\varphi,\ \cos\theta).$$

Thus we see that the coordinates

$$(\cos\theta,\ \sin\theta\cos\varphi,\ \sin\theta\sin\varphi)$$

are "funny" spherical coordinates for which the x-axis and the z-axis are swapped and φ is changed to $\pi/2 - \varphi$. We can use (FS5) and (FC5) to obtain the following Fourier series expansion for every function $f \in \mathrm{L}^2(S^2)$ in terms of the associated Legendre functions,

$$f(\varphi, \theta) = \sum_{\ell=0}^{\infty} \sum_{j=-\ell}^{\ell} \beta_\ell^j \, e^{-ij\varphi} P_\ell^j(\cos\theta), \tag{FS8}$$

with

$$\beta_\ell^j = \frac{(2\ell+1)}{4\pi} \frac{(\ell-j)!}{(\ell+j)!} \int_0^{2\pi} \int_0^{\pi} f(\varphi, \theta) e^{ij\varphi} P_\ell^j(\cos\theta) \sin\theta \, d\theta \, d\varphi. \tag{FC8}$$

We also have the Parseval identity

$$(2\ell+1) \int_0^{2\pi} \int_0^{\pi} |f(\varphi, \theta)|^2 d\nu = \sum_{\ell=0}^{\infty} \frac{1}{2\ell+1} \sum_{j=-\ell}^{\ell} \frac{(\ell+j)!}{(\ell-j)!} |\beta_\ell^j|^2, \tag{PS2}$$

where $d\nu = (1/4)\sin\theta\, d\theta\, d\varphi$ is the normalized measure on S^2 in spherical coordinates; among other sources, see Gallier and Quaintance [39] (Section 6.4).

Recall from Definition 14.14 and $(*_{48})$ that

$$Y_{\ell k}(\varphi,\theta) = t^{(\ell)}_{k0}(q) = i^{-k}\sqrt{\frac{(\ell-k)!}{(\ell+k)!}}\, e^{-ik\varphi}\, P^k_\ell(\cos\theta), \quad -\ell \leq k \leq \ell,$$

with $\ell \in \mathbb{N}$, so we have

$$\sqrt{\frac{(2\ell+1)(\ell-k)!}{(\ell+k)!}} e^{-ik\varphi} P^k_\ell(\cos\theta) = i^k\sqrt{2\ell+1}\, Y_{\ell k}(\varphi,\theta),$$

for $\ell \in \mathbb{N}$ and $-\ell \leq k \leq \ell$, and in view of (FS8) and (PS2), the above functions form a Hilbert basis for the functions in $\mathrm{L}^2(S^2)$. The functions $i^k\sqrt{2\ell+1}\, Y_{\ell k}(\varphi,\theta)$ are (a version of) the Laplace spherical harmonics $Y^m_\ell(\theta,\varphi)$, namely

$$Y^m_\ell(\theta,\varphi) = \sqrt{\frac{(2\ell+1)(\ell-k)!}{(\ell+k)!}} e^{-ik\varphi} P^k_\ell(\cos\theta).$$

Remark: Some authors include $1/\sqrt{4\pi}$ in the leading constant.

The associated Legendre functions can be computed starting with the Legendre polynomials using some recurrence equations; see Gallier and Quaintance [39] (Section 7.3).

14.14 Decomposition of Fields on the Sphere S^2

In various applications it is necessary to decompose not only scalar-valued but also vector-valued functions on the sphere S^2 into Fourier series that behave nicely under rotations of the sphere. Vilenkin suggests a way to do this that we now discuss (see Vilenkin [101] (Chapter III, Section 6.6).

Let $T_\ell \colon \mathbf{SU}(2) \to \mathcal{P}_\ell$ be the irreducible representation of $\mathbf{SU}(2)$ associated with $\ell \in R = \{0, 1/2, 1, 3/2, 2, 5/2, 3, \ldots\}$.

Definition 14.28. Let $\mathfrak{F}^S_\ell$ be the Hilbert space of functions $f \colon S^2 \to \mathcal{P}_\ell$ defined by the isomorphism

$$\mathfrak{F}^S_\ell \simeq \bigoplus_{j=-\ell}^{\ell} \mathrm{L}^2(S^2)\psi_j,$$

where the ψ_j constitute an orthonormal basis of $\mathcal{P}_\ell$ for an $\mathbf{SU}(2)$-invariant hermitian inner product defined in Section 14.7 ($\mathcal{P}_\ell$ is a complex vector space of dimension $2\ell+1$). More precisely, the inner product of two functions $f, g \in \mathfrak{F}^S_\ell$ is given by

$$\langle f, g\rangle = \int_{S^2} \langle f(\xi), g(\xi)\rangle\, d\sigma(\xi),$$

where $\langle -, - \rangle$ is the $\mathbf{SU}(2)$-invariant hermitian inner product on $\mathcal{P}_\ell$ defined earlier and σ is the normalized $\mathbf{SO}(3)$-invariant measure on S^2.

As in the previous section, $\mathbf{SU}(2)$ acts on S^2 by rotations, but for simplicity of notation, we write qX instead of $q \cdot X = qXq^*$, where $q \in \mathbf{SU}(2)$, X is the skew-hermitian matrix corresponding to the point (x, y, z) on the sphere S^2, and we write $X \in S^2$. We define the following representation.

Definition 14.29. For every $\ell \in R$, let $V_\ell \colon \mathbf{SU}(2) \to \mathbf{GL}(\mathfrak{F}_\ell^S)$ be the representation given by

$$[V_\ell(q)(f)](X) = [T_\ell(q)](f(q^{-1}X)), \quad q \in \mathbf{SU}(2),\ (f \colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}_\ell^S,\ X \in S^2.$$

Vilenkin calls the functions in $\mathfrak{F}_\ell^S$ *fields of quantities on the sphere transforming according to the irreducible representation* T_ℓ. For example, for $\ell = 1$, since $2\ell + 1 = 3$, we get a vector field on the sphere.

Observe that for any two functions $(f \colon S^2 \to \mathcal{P}_\ell), (g \colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}_\ell^S$, since T_ℓ is unitary and σ is rotation invariant, we have

$$\begin{aligned}
\langle V_\ell(q)(f), V_\ell(q)(g) \rangle &= \int_{S^2} \langle [V_\ell(q)(f)](X), [V_\ell(q)(g)](X) \rangle \, d\sigma(X) \\
&= \int_{S^2} \langle [T_\ell(q)](f(q^{-1}X), [T_\ell(q)](g(q^{-1}X) \rangle \, d\sigma(X) \\
&= \int_{S^2} \langle f(q^{-1}X), g(q^{-1}X) \rangle \, d\sigma(X) \\
&= \int_{S^2} \langle f(X), g(X) \rangle \, d\sigma(X) = \langle f, g \rangle.
\end{aligned}$$

Therefore the representation V_ℓ is unitary, that is, we have $V_\ell \colon \mathbf{SU}(2) \to \mathbf{U}(\mathfrak{F}_\ell^S)$.

For technical reasons, we need to convert the functions in $\mathfrak{F}_\ell^S$, which are functions on the sphere, to functions on $\mathbf{SU}(2)$. Let X_0 be the skew-hermitian matrix

$$X_0 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

corresponding to $e_1 = (1, 0, 0) \in S^2$.

Definition 14.30. For every function $(f \colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}_\ell^S$, let $\widehat{f} \colon \mathbf{SU}(2) \to \mathcal{P}_\ell$ be the function defined by

$$\widehat{f}(q) = [V_\ell(q^{-1})(f)](X_0) = [T_\ell(q^{-1})](f(qX_0)), \quad q \in \mathbf{SU}(2).$$

The functions $\widehat{f} \colon \mathbf{SU}(2) \to \mathcal{P}_\ell$ belong to the Hilbert space defined below.

Definition 14.31. Let $\mathfrak{F}_\ell^{\mathbf{SU}}$ be the Hilbert space of functions $f\colon \mathbf{SU}(2) \to \mathcal{P}_\ell$ defined by the isomorphism

$$\mathfrak{F}_\ell^{\mathbf{SU}} \simeq \bigoplus_{j=-\ell}^{\ell} \mathrm{L}^2(\mathbf{SU}(2))\psi_j.$$

More precisely, the inner product of two functions $f, g \in \mathfrak{F}_\ell^{\mathbf{SU}}$ is given by

$$\langle f, g\rangle = \int_{\mathbf{SU}(2)} \langle f(q), g(q)\rangle \, d\nu(q),$$

where $\langle -, -\rangle$ is the $\mathbf{SU}(2)$-invariant hermitian inner product on $\mathcal{P}_\ell$ defined earlier and ν is the normalized Haar measure on $\mathbf{SU}(2)$.

Proposition 14.39. *The map $f \mapsto \widehat{f}$ is an injection from $\mathfrak{F}_\ell^S$ to $\mathfrak{F}_\ell^{\mathbf{SU}}$.*

Proof. Indeed, if $\widehat{f}(q) = \widehat{g}(q)$ for all $q \in \mathbf{SU}(2)$, then $[T_\ell(q^{-1})](f(qX_0)) = [T_\ell(q^{-1})](g(qX_0))$ for all $q \in \mathbf{SU}(2)$, and since $T_\ell(q^{-1})$ is a bijection, $f(qX_0) = g(qX_0)$ for all $q \in \mathbf{SU}(2)$, and since the action of $\mathbf{SU}(2)$ on S^2 is transitive, we must have $f = g$. $\square$

Definition 14.32. The image of $\mathfrak{F}_\ell^S$ in $\mathfrak{F}_\ell^{\mathbf{SU}}$ by the map $\widehat{\ }$ is denoted by $\widehat{\mathfrak{F}_\ell^S}$.

Observe that f can be recovered from $\widehat{f}$ as follows:

$$f(qX_0) = [T_\ell(q)](\widehat{f}(q)), \quad q \in \mathbf{SU}(2).$$

Since Ω_x is the stabilizer of X_0, for every $h \in \Omega_x$ we have $hX_0 = X_0$, and so

$$\begin{aligned}
\widehat{f}(qh) &= [T_\ell((qh)^{-1})](f(qhX_0)) \\
&= T_\ell(h^{-1})([T_\ell(q^{-1})](f(qX_0)) \\
&= [T_\ell(h)^{-1}](\widehat{f}(q)),
\end{aligned}$$

which we record as the equation

$$\widehat{f}(qh) = [T_\ell(h)^{-1}](\widehat{f}(q)), \quad (f\colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}_\ell^S,\ h \in \Omega_x. \tag{fhat}$$

Let us figure out what is the function in $\widehat{\mathfrak{F}_\ell^S} \subseteq \mathfrak{F}_\ell^{\mathbf{SU}}$ corresponding to the function $[V_\ell(q_0)](f) \in \mathfrak{F}_\ell^S$, with $(f\colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}_\ell^S$. We have

$$\begin{aligned}
([V_\ell(q_0)](f))\widehat{\ }(q) &= [V_\ell(q^{-1})(V_\ell(q_0)(f))](X_0) \\
&= [V_\ell(q^{-1}q_0)(f)](X_0) = \widehat{f}(q_0^{-1}q),
\end{aligned}$$

for all $q_0, q \in \mathbf{SU}(2)$. This suggests the following definition.

Definition 14.33. The representation $\widehat{V}_\ell\colon \mathbf{SU}(2) \to \mathbf{U}(\widehat{\mathfrak{F}^S_\ell})$ is given by

$$[\widehat{V}_\ell(q_0)(\widehat{f}\,)](q) = \widehat{f}(q_0^{-1}q) = [T_\ell(q^{-1}q_0)](f(q_0^{-1}qX_0)), \quad (f\colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}^S_\ell,\ q_0, q \in \mathbf{SU}(2).$$

The definition of $\widehat{V}_\ell$ implies that

$$\widehat{V}_\ell(q_0)(\widehat{f}\,) = ([V_\ell(q_0)](f))\widehat{\ }$$

so that the following diagram commutes,

$$\begin{tikzcd} \mathfrak{F}^S_\ell \arrow[r, "V_\ell(q_0)"] \arrow[d, "\widehat{\ }"'] & \mathfrak{F}^S_\ell \arrow[d, "\widehat{\ }"] \\ \widehat{\mathfrak{F}^S_\ell} \arrow[r, "\widehat{V}_\ell(q_0)"'] & \widehat{\mathfrak{F}^S_\ell}, \end{tikzcd}$$

and since $\widehat{\ }$ is an isomorphism between $\mathfrak{F}^S_\ell$ and $\widehat{\mathfrak{F}^S_\ell}$, the representations $V_\ell\colon \mathbf{SU}(2) \to \mathbf{U}(\mathfrak{F}^S_\ell)$ and $\widehat{V}_\ell\colon \mathbf{SU}(2) \to \mathbf{U}(\widehat{\mathfrak{F}^S_\ell})$ are equivalent.

The trick is now to decompose the space $\widehat{\mathfrak{F}^S_\ell}$ into a direct sum of $2\ell + 1$ subspaces $(\widehat{\mathfrak{F}^S_\ell})_k$ which are related to the spaces $\mathfrak{L}^2_{-k}$ introduced in Section 14.13.

Definition 14.34. For k with $-\ell \leq k \leq \ell$, for every function $(f\colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}^S_\ell$, define the function $\widehat{f}_k \in \widehat{\mathfrak{F}^S_\ell}$ as

$$\widehat{f}_k(q) = \frac{1}{2\pi}\int_0^{2\pi} \widehat{f}(qh(t))e^{-ikt}\, dt, \quad q \in \mathbf{SU}(2),$$

with

$$h(t) = \begin{pmatrix} e^{\frac{it}{2}} & 0 \\ 0 & e^{-\frac{it}{2}} \end{pmatrix} \in \Omega_x.$$

Proposition 14.40. *For k with $-\ell \leq k \leq \ell$, for every function $(f\colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}^S_\ell$, the following properties hold:*

(1)
$$\widehat{f}_k(qh(s)) = e^{iks}\widehat{f}_k(q), \quad q \in \mathbf{SU}(2),\ h(s) \in \Omega_x.$$

Consequently, $\widehat{f}_k \in \bigoplus_{j=-\ell}^{\ell} \mathfrak{L}^2_{-k}\psi_j$.

(2)
$$\widehat{f}(q) = \sum_{k=-\ell}^{\ell} \widehat{f}_k(q), \quad q \in \mathbf{SU}(2).$$

(3)

$$[\widehat{V}_\ell(q_0)(\widehat{f}\,)]_k(q) = \widehat{f}_k(q_0^{-1}q), \quad q_0, q \in \mathbf{SU}(2).$$

Proof. (1) We have

$$\begin{aligned}
\widehat{f}_k(qh(s)) &= \frac{1}{2\pi}\int_0^{2\pi} \widehat{f}(qh(s)h(t))e^{-ikt}\,dt \\
&= e^{iks}\frac{1}{2\pi}\int_0^{2\pi} \widehat{f}(qh(s+t))e^{-ik(s+t)}\,dt \\
&= e^{iks}\frac{1}{2\pi}\int_0^{2\pi} \widehat{f}(qh(t_1))e^{-ikt_1}\,dt_1 \\
&= e^{iks}\widehat{f}_k(q),
\end{aligned}$$

as claimed.

(2) Using (fhat), we have

$$\begin{aligned}
\sum_{k=-\ell}^{\ell} \widehat{f}_k(q) &= \frac{1}{2\pi}\sum_{k=-\ell}^{\ell}\int_0^{2\pi} \widehat{f}(qh(t))e^{-ikt}\,dt \\
&= \frac{1}{2\pi}\int_0^{2\pi}\left(\sum_{k=-\ell}^{\ell} e^{-ikt}T_\ell(h(t)^{-1})\right)(\widehat{f}(q))\,dt.
\end{aligned}$$

Here we need to recall from Proposition 14.15 that in the basis $(\psi_m)_{-\ell\leq m\leq \ell}$, since $h(t) = r_x(t/2)$, the matrix of $T_\ell(h(t)^{-1})$ is the diagonal matrix

$$\begin{pmatrix}
e^{-i\ell t} & & & & \\
& e^{-i(\ell-1)t} & & & \\
& & \ddots & & \\
& & & e^{i(\ell-1)t} & \\
& & & & e^{i\ell t}
\end{pmatrix}.$$

Consequently the entries of $e^{-ikt}T_\ell(h(t)^{-1})$ are of the form $e^{-i(\ell-j+k)t}$ with $j = 0, 1, \ldots, 2\ell$ and $\ell + k - j$ an integer, but

$$\frac{1}{2\pi}\int_0^{2\pi} e^{-i(\ell+k-j)t}\,dt = \delta_{\ell+k,j}$$

so the only entry that survives corresponds to $j = \ell + k$ and its contribution is 1, so in fact

$$\sum_{k=-\ell}^{\ell} e^{-ikt}T_\ell(h(t)^{-1}) = I_{2\ell+1},$$

and thus

$$\frac{1}{2\pi}\int_0^{2\pi}\left(\sum_{k=-\ell}^{\ell} e^{-ikt}T_\ell(h(t)^{-1})\right)(\widehat{f}(q))\,dt = \widehat{f}(q),$$

as claimed.

(3) We have

$$\begin{aligned}
[\widehat{V}(q_0)(\widehat{f}\,)]_k(q) &= \frac{1}{2\pi}\int_0^{2\pi}[(\widehat{V}(q_0)(\widehat{f}\,)](qh(t))e^{-ikt}\,dt\\
&= \frac{1}{2\pi}\int_0^{2\pi}\widehat{f}(q_0^{-1}qh(t))e^{-ikt}\,dt\\
&= \widehat{f}_k(q_0^{-1}q),
\end{aligned}$$

which concludes the proof of the proposition. $\square$

As a corollary we have the following result.

Proposition 14.41. *Denote by $(\widehat{\mathfrak{F}^S_\ell})_k$ the image of $\widehat{\mathfrak{F}^S_\ell}$ by the linear map $\widehat{f} \mapsto \widehat{f}_k$, with $(f\colon S^2 \to \mathcal{P}_\ell) \in \mathfrak{F}^S_\ell$.*

(1) We have a direct sum

$$\widehat{\mathfrak{F}^S_\ell} = \bigoplus_{k=-\ell}^{\ell}(\widehat{\mathfrak{F}^S_\ell})_k,$$

where every function $\widehat{f}_k \in (\widehat{\mathfrak{F}^S_\ell})_k$ satisfies the equation

$$\widehat{f}_k(qh(s)) = e^{iks}\widehat{f}_k(q), \quad q \in \mathbf{SU}(2),\ h(s) \in \Omega_x.$$

(2) The map $(\widehat{V}_\ell)_k$ defined by

$$[\widehat{V}_\ell(q_0)(\widehat{f}\,)]_k(q) = \widehat{f}_k(q_0^{-1}q), \quad q_0, q \in \mathbf{SU}(2)$$

is a representation $(\widehat{V}_\ell)_k\colon \mathbf{SU}(2) \to \mathbf{U}((\widehat{\mathfrak{F}^S_\ell})_k)$, and we have

$$\widehat{V}_\ell(q_0) = \bigoplus_{k=-\ell}^{\ell}(\widehat{V}_\ell)_k(q_0).$$

Proof. Since by (2), $\widehat{\mathfrak{F}^S_\ell}$ is the sum of the subspaces $(\widehat{\mathfrak{F}^S_\ell})_k$, and by (1), if there was a nonzero function such that $\widehat{f}_{k_1} = \widehat{f}_{k_2}$ for some $k_1 \neq k_2$, then we would have

$$\widehat{f}_{k_1}(qh(s)) = e^{ik_1s}\widehat{f}_{k_1}(q) = \widehat{f}_{k_2}(qh(s)) = e^{ik_2s}\widehat{f}_{k_2}(q),$$

and so we would have $e^{ik_1 s} = e^{ik_2 s}$ for all s, which implies $k_1 = k_2$, a contradiction.

The fact that

$$\widehat{V}_\ell(q_0) = \bigoplus_{k=-\ell}^{\ell} (\widehat{V}_\ell)_k(q_0),$$

follows from Parts (2) and (3). $\square$

For every function $\sum_{j=-\ell}^{\ell} \widehat{(f_j)}\psi_j$ in $(\widehat{\mathfrak{F}_\ell^S})_k$, we have $\widehat{(f_j)} \in \mathfrak{L}^2_{-k}$ (with $f_j \in \mathrm{L}^2(S^2)$), so as shown in Section 14.13, the function $\widehat{(f_j)}$ can be expanded in Fourier series according to Formulae (FS3) and (FC3) (with k changed to $-k$).

We leave it as an exercise to the reader to use the isomorphism $\widehat{\ }\colon \mathfrak{F}_\ell^S \to \widehat{\mathfrak{F}_\ell^S}$ to define a direct sum decomposition of $\mathfrak{F}_\ell^S$ of the form

$$\mathfrak{F}_\ell^S = \bigoplus_{k=-\ell}^{\ell} (\mathfrak{F}_\ell^S)_k$$

and to translate the results obtained for functions in $\widehat{\mathfrak{F}_\ell^S}$ and the representations $\widehat{V}_\ell$ to the functions in $\mathfrak{F}_\ell^S$ and to the representations V_ℓ.

14.15 The Clebsch–Gordan Coefficients

The Clebsch–Gordan coefficients have to do with tensor products of the irreducible representations T_ℓ of $\mathbf{SU}(2)$ (see Definition 13.11 for the definition of the tensor product of representations). In general, the tensor product $T_{\ell_1} \otimes T_{\ell_2}$ of two irreducible representations T_{ℓ_1} and T_{ℓ_2} of $\mathbf{SU}(2)$ is not irreducible, so according to the Peter–Weyl theorem (Theorem 13.16) it splits as a direct sum of irreducible representations. Since the character associated with the representation $T_{\ell_1} \otimes T_{\ell_2}$ is equal to the product $\chi_{T_{\ell_1}}\chi_{T_{\ell_2}}$ of the characters $\chi_{T_{\ell_1}}$ and $\chi_{T_{\ell_2}}$ associated with T_{ℓ_1} and T_{ℓ_2}, by Proposition 13.18, this splitting as a direct sum decomposition translates into a decomposition

$$\chi_{T_{\ell_1}}\chi_{T_{\ell_2}} = \sum_{\ell} c_{\ell_1,\ell_2}^{\ell} \chi_{T_\ell},$$

where c_{ℓ_1,ℓ_2}^{ℓ} is the number of times that the irreducible representation T_ℓ occurs in the representation $T_{\ell_1} \otimes T_{\ell_2}$ (see Section 13.3).

The natural numbers c_{ℓ_1,ℓ_2}^{ℓ} can be determined from the expression of the characters that was obtained in Section 14.1. However this expression was obtained for the representations U_m in the space $\mathcal{P}_m^{\mathbb{C}}(2)$ of homogeneous polynomials of degree m in two variables, so we work out the expression of the characters in terms of the representations T_ℓ in the spaces $\mathcal{P}_\ell^{\mathbb{C}}$ of polynomials of degree 2ℓ in one variable (in particular, ℓ is now an integer or a half integer). To simplify notation, we will write $\mathcal{P}_\ell$ instead of $\mathcal{P}_\ell^{\mathbb{C}}$.

Recall that we showed in the proof of Proposition 14.1 that every unitary matrix $q \in \mathbf{SU}(2)$ is diagonalizable as

$$q = Rr_x(t/2)R^*$$

for some unitary matrix $R \in \mathbf{SU}(2)$, where

$$r_x(t/2) = \begin{pmatrix} e^{\frac{it}{2}} & 0 \\ 0 & e^{-\frac{it}{2}} \end{pmatrix}$$

is uniquely determined if $0 \leq t \leq 2\pi$. If the matrix q is given by

$$q = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}, \quad |\alpha|^2 + |\beta|^2 = 1,$$

then its eigenvalues are the zeros of the equation

$$\begin{vmatrix} \lambda - \alpha & -\beta \\ \overline{\beta} & \lambda - \overline{\alpha} \end{vmatrix} = 0,$$

that is,

$$\lambda^2 - 2\Re(\alpha)\lambda + 1 = 0$$

(since $\alpha + \overline{\alpha} = 2\Re(\alpha)$), whose zeros are

$$\lambda = \Re(\alpha) \pm i\sqrt{1 - (\Re(\alpha))^2}.$$

Since we assumed that the eigenvalues of q are $e^{\pm\frac{it}{2}}$, we have

$$\Re(\alpha) = \cos\frac{t}{2}.$$

If q is expressed in terms of the Euler angles as $q = u(\varphi, \theta, \psi)$, then from the fomulae just before Proposition 14.4 we have

$$\alpha = \cos\frac{\theta}{2}\, e^{\frac{i(\varphi+\psi)}{2}},$$

and so

$$\Re(\alpha) = \cos\frac{t}{2} = \cos\frac{\theta}{2}\, \cos\frac{(\varphi+\psi)}{2}.$$

Since the characters are central functions, that is, constant on conjugacy classes, we have

$$\chi_{T_\ell}(q) = \chi_{T_\ell}(r_x(t/2)) = \operatorname{tr}(T_\ell(r_x(t/2))).$$

Since we showed in Proposition 14.15 that in the basis $(z^{\ell-k})_{-\ell\leq k\leq \ell}$, the matrix of $T_\ell(r_x(t/2))$ is the diagonal matrix

$$\begin{pmatrix} e^{i\ell t} & & & & \\ & e^{i(\ell-1)t} & & & \\ & & \ddots & & \\ & & & e^{-i(\ell-1)t} & \\ & & & & e^{-i\ell t} \end{pmatrix},$$

we obtain

$$\chi_{T_\ell}(q) = \chi_{T_\ell}(r_x(t/2)) = \sum_{k=-\ell}^{\ell} e^{-ikt} = \frac{\epsilon^{\ell+1} - \epsilon^{-\ell}}{\epsilon - 1},$$

with $\epsilon = e^{-it}$, and we showed in the proof of Proposition 14.1 that we obtain the following expression (in that formula we make $m = 2\ell$ and $\varphi = t/2$):

$$\chi_{T_\ell}(q) = \frac{\epsilon^{\ell+1} - \epsilon^{-\ell}}{\epsilon - 1} = \frac{\sin\left(\ell + \frac{1}{2}\right) t}{\sin \frac{t}{2}},$$

with $\epsilon = e^{-it}$. Compare Vilenkin [101], Chapter III, Section 7.1. Using the above formula we obtain the following result.

Proposition 14.42. *For any two irreducible representations T_{ℓ_1} and T_{ℓ_2} of $\mathbf{SU}(2)$, we have*

$$\chi_{T_{\ell_1}}(q)\chi_{T_{\ell_2}}(q) = \sum_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \chi_{T_\ell}(q), \quad q \in \mathbf{SU}(2). \tag{CG1}$$

Proof. We follow Vilenkin [101], Chapter III, Section 8.1. First assume that $\ell_1 \geq \ell_2$. With $\epsilon = e^{-it}$ as above, we have

$$\begin{aligned}
\chi_{T_{\ell_1}}(q)\chi_{T_{\ell_2}}(q) &= \sum_{k=-\ell_2}^{\ell_2} \epsilon^k \frac{(\epsilon^{\ell_1+1} - \epsilon^{-\ell_1})}{\epsilon - 1} \\
&= \sum_{k=-\ell_2}^{\ell_2} \frac{\epsilon^{\ell_1+k+1} - \epsilon^{k-\ell_1}}{\epsilon - 1} \\
&= \frac{1}{\epsilon - 1}(\epsilon^{\ell_1+\ell_2+1} + \cdots + \epsilon^{\ell_1-\ell_2+1} - \epsilon^{\ell_2-\ell_1} - \cdots - \epsilon^{-\ell_1-\ell_2}) \\
&= \frac{1}{\epsilon - 1}(\epsilon^{\ell_1+\ell_2+1} - \epsilon^{\ell_2-\ell_1} + \cdots + \epsilon^{\ell_1-\ell_2+1} - \epsilon^{\ell_2-\ell_1}),
\end{aligned}$$

where the last line is obtained by combining pairwise positive and negative terms, the sum of whose indices is equal to 1, we obtain

$$\chi_{T_{\ell_1}}(q)\chi_{T_{\ell_2}}(q) = \sum_{\ell=\ell_1-\ell_2}^{\ell_1+\ell_2} \frac{\epsilon^{\ell+1} - \epsilon^{-\ell}}{\epsilon - 1} = \sum_{\ell=\ell_1-\ell_2}^{\ell_1+\ell_2} \chi_{T_\ell}(q).$$

If $\ell_2 \geq \ell_1$, the proof is similar but the sum starts with $\ell_2 - \ell_1$. ∎

The above proposition shows the somewhat unexpected fact that in the decomposition of the tensor product representation $T_{\ell_1} \otimes T_{\ell_2}$, those representations T_ℓ that occur correspond

to values of ℓ such that $|\ell_1 - \ell_2| \leq \ell \leq \ell_1 + \ell_2$ where ℓ is an integer or a half integer as $\ell_1 + \ell_2$ is, and each such representation occurs exactly once. Thus

$$T_{\ell_1} \otimes T_{\ell_2} = \bigoplus_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} T_\ell. \tag{CG2}$$

We also have an isomorphism

$$\mathcal{P}_{\ell_1} \otimes \mathcal{P}_{\ell_2} \simeq \bigoplus_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \mathcal{P}_\ell. \tag{CG3}$$

The space $\mathcal{P}_{\ell_1} \otimes \mathcal{P}_{\ell_2}$ has dimension $(2\ell_1+1)(2\ell_2+1)$ and each summand $\mathcal{P}_\ell$ has dimension $2\ell+1$. The reader should check that

$$(2\ell_1+1)(2\ell_2+1) = \sum_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} (2\ell+1).$$

Recall from Proposition 14.16 that each vector space $\mathcal{P}_\ell$ has an orthonormal basis (ψ_k) $(-\ell \leq k \leq \ell)$ invariant under the action of $\mathbf{SU}(2)$. Following Vilenkin [101] (Chapter III, Section 8.2), we denote the basis of $\mathcal{P}_{\ell_1}$ as $(\mathbf{f}_j)$ $(-\ell_1 \leq j \leq \ell_1)$ and the basis of $\mathcal{P}_{\ell_2}$ as $(\mathbf{h}_k)$ $(-\ell_2 \leq k \leq \ell_2)$. Then the family of tensor products

$$\mathbf{f}_j \otimes \mathbf{h}_k, \quad -\ell_1 \leq j \leq \ell_1,\ -\ell_2 \leq k \leq \ell_2$$

is a basis of $\mathcal{P}_{\ell_1} \otimes \mathcal{P}_{\ell_2}$. If we give $\mathcal{P}_{\ell_1} \otimes \mathcal{P}_{\ell_2}$ the inner product defined in Definition 13.10 induced by the inner products associated with the bases $(\mathbf{f}_j)$ and $(\mathbf{h}_k)$, then the vectors $(\mathbf{f}_j \otimes \mathbf{h}_k)$ form an orthonormal basis of $\mathcal{P}_{\ell_1} \otimes \mathcal{P}_{\ell_2}$.

Since we have the direct sum

$$\mathcal{P}_{\ell_1} \otimes \mathcal{P}_{\ell_2} \simeq \bigoplus_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \mathcal{P}_\ell,$$

we also have a basis of $\mathcal{P}_{\ell_1} \otimes \mathcal{P}_{\ell_2}$ consisting of the union of the bases associated with each of the summand $\mathcal{P}_l$, which Vilenkin denotes by

$$\mathbf{a}^\ell_m, \quad |\ell_1 - \ell_2| \leq \ell \leq \ell_1 + \ell_2,\ -\ell \leq m \leq \ell,$$

where for ℓ fixed, $(\mathbf{a}^\ell_m)$ $(-\ell \leq m \leq \ell)$ is the basis of $\mathcal{P}_\ell$. Since both bases are orthonormal bases of $\mathcal{P}_{\ell_1} \otimes \mathcal{P}_{\ell_2}$ there is a unitary matrix C expressing the basis $(\mathbf{f}_j \otimes \mathbf{h}_k)$ in terms of the basis $(\mathbf{a}^\ell_m)$, and the entries of the matrix C are called the *Clebsch–Gordan coefficients*.

Amazingly, these coefficients can be computed explicitly, but the formulae are very complicated and the technical details of the computations are quite involved. Complete details can be found in Vilenkin [101] (Chapter III, Section 8). We will content ourselves by providing an outline of these computations.

The first observation is that the matrix of $(T_{\ell_1} \otimes T_{\ell_2})(q) = T_{\ell_1}(q) \otimes T_{\ell_2}(q)$ with respect to the basis $(\mathbf{f}_j \otimes \mathbf{h}_k)$ is the Kronecker product of the matrices $t^{(\ell_1)}(q)$ and $t^{(\ell_2)}(q)$. Following Vilenkin we denote this matrix as $\alpha(q) = (\alpha_{(jk),(j'k')}(q))$, and we have

$$\alpha_{(jk),(j'k')}(q) = t^{(\ell_1)}_{jj'}(q)t^{(\ell_2)}_{kk'}(q), \tag{CG4}$$

with $-\ell_1 \leq j, j' \leq \ell_1$, $-\ell_2 \leq k, k' \leq \ell_2$.

On the other hand, in the basis $(\mathbf{a}^{\ell}_m)$, the matrix representing $T_{\ell_1}(q) \otimes T_{\ell_2}(q)$ is a block-diagonal matrix whose blocks are the matrices $t^{(\ell)}(q)$. Again, following Vilenkin, we denote this matrix as $\beta(q) = (\beta_{(\ell m),(\ell' m')}(q))$, with $|\ell_1 - \ell_2| \leq \ell, \ell' \leq \ell_1 + \ell_2$, $-\ell \leq m \leq \ell$ and $-\ell' \leq m' \leq \ell'$. Since this matrix is block-diagonal we must have

$$\beta_{(\ell m),(\ell' m')}(q) = 0 \quad \text{if } \ell \neq \ell',$$

and if $\ell = \ell'$, then $\beta_{(\ell m),(\ell m')}(q) = t^{(\ell)}_{mm'}(q)$, so we have

$$\beta_{(\ell m),(\ell' m')}(q) = \delta_{\ell\ell'} t^{(\ell)}_{mm'}(q), \tag{CG5}$$

with $-\ell \leq m \leq \ell$, $-\ell' \leq m' \leq \ell'$.

The change of basis matrix $C = (C_{(\ell m),(jk)})$ is the unitary matrix defined such that the (jk)th column of C consists of the coefficients of $\mathbf{f}_j \otimes \mathbf{h}_k$ over the basis $(\mathbf{a}^{\ell}_m)$, namely

$$\mathbf{f}_j \otimes \mathbf{h}_k = \sum_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \sum_{m=-\ell}^{\ell} C_{(\ell m),(jk)} \mathbf{a}^{\ell}_m, \tag{CG6}$$

with $-\ell_1 \leq j \leq \ell_1$, $-\ell_2 \leq k \leq \ell_2$. Since $\beta(q)$ is the matrix of $T_{\ell_1}(q) \otimes T_{\ell_2}(q)$ in the "old" basis $(\mathbf{a}^{\ell}_m)$ and $\alpha(q)$ is the the matrix of $T_{\ell_1}(q) \otimes T_{\ell_2}(q)$ in the "new" basis $(\mathbf{f}_j \otimes \mathbf{h}_k)$, we have

$$\alpha(q) = C^* \beta(q) C. \tag{CG7}$$

It turns out that it is often desirable to indicate explicitly the dependence of C on the indices ℓ_1 and ℓ_2, so we also write $C(\ell_1, \ell_2, \ell; j, k, m)$ instead of $C_{(\ell m),(jk)}$. To be more concise we introduce the following notation.

Definition 14.35. The coefficients $C(\ell_1, \ell_2, \ell; j, k, m)$ are also written as $C(\mathbf{l}, \mathbf{j})$, with $\mathbf{l} = (\ell_1, \ell_2, \ell)$ and $\mathbf{j} = (j, k, m)$ and are called the *Clebsch–Gordan coefficients*.

In terms of matrix elements, (CG7) yields

$$\alpha_{(jk),(j'k')}(q) = \sum_{\ell,\ell'=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell'}^{\ell'} \overline{C_{(\ell m),(jk)}} \beta_{(\ell m),(\ell' m')} C_{(\ell' m'),(j'k')}. \tag{CG8}$$

Using (CG4) and (CG5), we obtain

$$t^{(\ell_1)}_{jj'}(q)t^{(\ell_2)}_{kk'}(q) = \sum_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} \sum_{m,m'=-\ell}^{\ell} C(\mathbf{l},\mathbf{j}')\overline{C(\mathbf{l},\mathbf{j})}t^{(\ell)}_{mm'}(q), \tag{CG9}$$

with $\mathbf{l} = (\ell_1, \ell_2, \ell)$, $\mathbf{j} = (j, k, m)$, $\mathbf{j}' = (j', k', m')$.

Equation (CG9) is the key to the computation of the coefficients $C(\mathbf{l},\mathbf{j})$. By a clever use of the fact that the functions $\sqrt{2\ell+1}\, t^{(\ell)}_{mm'}(q)$ $(-\ell \leq m, m' \leq \ell)$ form a Hilbert basis of $\mathrm{L}^2(\mathbf{SU}(2))$ (see the beginning of Section 14.13) and the expression of $t^{(\ell)}_{mm'}(q)$ in terms of the Euler angles given by Proposition 14.27 as

$$t^{(\ell)}_{mm'}(q) = e^{-i(m\varphi+m'\psi)}\, P^{\ell}_{mm'}(\cos\theta) \tag{$*$}$$

(with $q = u(\varphi,\theta,\psi)$), it is possible to find (more or less explicit) formulae for $C(\mathbf{l},\mathbf{j})$.

The first step is to multiply both sides of (CG9) by $\overline{t^{(\ell)}_{mm'}(q)}$ and integrate over $\mathbf{SU}(2)$. Since the $\sqrt{2\ell+1}\, t^{(\ell)}_{mm'}(q)$ $(-\ell \leq m, m' \leq \ell)$ form a Hilbert basis we obtain

$$C(\mathbf{l},\mathbf{j}')\overline{C(\mathbf{l},\mathbf{j})} = (2\ell+1)\int_{\mathbf{SU}(2)} t^{(\ell_1)}_{jj'}(q)t^{(\ell_2)}_{kk'}(q)\overline{t^{(\ell)}_{mm'}(q)}\, d\mu(q). \tag{CG10}$$

Using $(*)$ and the volume form

$$\frac{1}{16\pi^2}\, \sin\theta\, d\theta\, d\varphi\, d\psi$$

(see Proposition 14.34), we find that *the integral in (CG10) is nonzero if and only if* $j+k = m$ *and* $j' + k' = m'$. Therefore we only need to compute the Clebsch–Gordan coefficients $C(\ell_1, \ell_2, \ell; j, k, j{+}k)$. Among those, it turns out that the coefficients $C(\ell_1, \ell_2, \ell; \ell_1, -\ell_2, \ell_1{-}\ell_2)$ play a special role. They can be computed before the arbitrary coefficients $C(\ell_1, \ell_2, \ell; j, k, j+k)$ and it can be arranged that $C(\ell_1, \ell_2, \ell; \ell_1, -\ell_2, \ell_1 - \ell_2) \geq 0$, which implies that all $C(\ell_1, \ell_2, \ell; j, k, j+k)$ are real, even though a priori they are complex numbers. The points is that for each ℓ, we can multiply all the basis vectors $\mathbf{a}^{\ell}_m$ by a complex number of unit length and still obtain a Hilbert basis, and ensure that $C(\ell_1, \ell_2, \ell; \ell_1, -\ell_2, \ell_1 - \ell_2) \geq 0$. From now on, we assume that this normalization has been made.

We now go back to (CG10) in which we set $m = j + k$ and $m' = j' + k'$ and use $(*)$ to integrate (making the substitution $x = \cos\theta$) to obtain

$$C(\mathbf{l},\mathbf{j}')\overline{C(\mathbf{l},\mathbf{j})} = \frac{(2\ell+1)}{2}\int_{-1}^{1} P^{\ell_1}_{jj'}(x)P^{\ell_2}_{kk'}(x)\overline{P^{\ell}_{j+k\, j'+k'}(x)}\, dx, \tag{CG11}$$

with $\mathbf{l} = (\ell_1, \ell_2, \ell)$, $\mathbf{j} = (j, k, j + k)$, $\mathbf{j}' = (j', k', j' + k')$.

In order to compute $C(\ell_1, \ell_2, \ell; \ell_1, -\ell_2, \ell_1 - \ell_2)$ we set $j' = \ell_1$ and $k' = -\ell_2$ in (CG11). Then we can use the special case of $(*_{40})$ in which $j = \ell$, the symmetry equations $P^{\ell}_{mn}(z) =$

$P^{\ell}_{-m-n}(z)$ and $P^{\ell}_{mn}(z) = P^{\ell}_{nm}(z)$ (just after Proposition 14.28), and ($*_{43}$), and after a bit of work on (CG11) (see Vilenkin [101], Chapter III, Section 8.3, Page 179), we obtain the following formidable equation:

$$\begin{aligned}
&C(\ell_1,\ell_2,\ell;\ell_1,-\ell_2,\ell_1-\ell_2)\overline{C(\ell_1,\ell_2,\ell;j,k,j+k)}\\
&\qquad = \frac{(-1)^{-\ell+\ell_1+k}(2\ell+1)}{2^{\ell+\ell_1+\ell_2+1}}\sqrt{\frac{(2\ell_1)!(2\ell_2)!(\ell+j+k)!}{(\ell_1-j)!(\ell_1+j)!(\ell_2-k)!(\ell_2+k)!}}\\
&\qquad\qquad \times\sqrt{\frac{1}{(\ell-j-k)!(\ell+\ell_1-\ell_2)!(\ell-\ell_1+\ell_2)!}}\\
&\qquad \times\int_{-1}^{1}(1-x)^{\ell_1-j}(1+x)^{\ell_2-k}\frac{d^{\ell-j-k}}{dx^{\ell-j-k}}[(1-x)^{\ell-\ell_1+\ell_2}(1+x)^{\ell+\ell_1-\ell_2}]\,dx. \qquad \text{(CG12)}
\end{aligned}$$

To find $C(\ell_1,\ell_2,\ell;\ell_1,-\ell_2,\ell_1-\ell_2)$ we set $j=\ell_1$ and $k=-\ell_2$. Sparing the reader some details found in Vilenkin and using integration by parts $\ell-\ell_1+\ell_2$ times, we find that

$$|C(\ell_1,\ell_2,\ell;\ell_1,-\ell_2,\ell_1-\ell_2)|^2 = \frac{(2\ell+1)(2\ell_1)!(2\ell_2)!}{(\ell_1+\ell_2-\ell)!(\ell_1+\ell_2+\ell+1)!}.$$

Since we normalized our bases so that $C(\ell_1,\ell_2,\ell;\ell_1,-\ell_2,\ell_1-\ell_2)\geq 0$, we obtain

$$C(\ell_1,\ell_2,\ell;\ell_1,-\ell_2,\ell_1-\ell_2) = \sqrt{\frac{(2\ell+1)(2\ell_1)!(2\ell_2)!}{(\ell_1+\ell_2-\ell)!(\ell_1+\ell_2+\ell+1)!}}. \qquad \text{(CG13)}$$

Observe that (CG13) implies that $C(\ell_1,\ell_2,\ell;j,k,j+k)$ is real, so plugging (CG13) into (CG12), we finally obtain the "master equation"

$$\begin{aligned}
&C(\ell_1,\ell_2,\ell;j,k,j+k) = \frac{(-1)^{-\ell+\ell_1+k}}{2^{\ell+\ell_1+\ell_2+1}}\\
&\quad\times\sqrt{\frac{(2\ell+1)(\ell+j+k)!(\ell_1+\ell_2-\ell)!(\ell_1+\ell_2+\ell+1)!}{(\ell_1-j)!(\ell_1+j)!(\ell_2-k)!(\ell_2+k)!(\ell-j-k)!(\ell+\ell_1-\ell_2)!(\ell-\ell_1+\ell_2)!}}\\
&\qquad \times\int_{-1}^{1}(1-x)^{\ell_1-j}(1+x)^{\ell_2-k}\frac{d^{\ell-j-k}}{dx^{\ell-j-k}}[(1-x)^{\ell-\ell_1+\ell_2}(1+x)^{\ell+\ell_1-\ell_2}]\,dx. \qquad \text{(CG14)}
\end{aligned}$$

Remark: Another master equation is obtained from (CG11) as follows. When we set $j'=\ell_1$ and $k'=-\ell_2$, the polynomial $P^{\ell}_{j+k,\ell_1-\ell_2}(x)$ appears, but $P^{\ell}_{j+k,\ell_1-\ell_2}(x) = P^{\ell}_{\ell_1-\ell_2,j+k}(x)$, so we can use $P^{\ell}_{\ell_1-\ell_2,j+k}(x)$ and obtain another version of (CG1); see Vilenkin [101], Chapter III, Section 8.3, Page 181.

Equation (CG14) still does not give an explicit formula but such formulae can be obtained. By using the product rule in (CG14) and integrating term by term, it is shown in Vilenkin

[101] (also Page 181) that we have

$$
\begin{aligned}
C(\ell_1, \ell_2, \ell; j, k, j+k) &= (-1)^{-\ell+\ell_1+k} \\
&\times \sqrt{\frac{(2\ell+1)(\ell+j+k)!(\ell_1+\ell_2-\ell)!(\ell-j-k)!(\ell+\ell_1-\ell_2)!(\ell-\ell_1+\ell_2)!}{(\ell_1+\ell_2+\ell+1)!(\ell_1-j)!(\ell_1+j)!(\ell_2-k)!(\ell_2+k)!}} \\
&\times \sum_{s=M}^{N} \frac{(-1)^s(\ell+\ell_2-j-s)!(\ell_1+j+s)!}{s!(\ell-j-k-s)!(\ell-\ell_1+\ell_2-s)!(\ell_1-\ell_2+j+k+s)!} \qquad \text{(CG15)}
\end{aligned}
$$

with $M = \max(0, \ell_2 - \ell_1 - j - k)$, $N = \min(\ell - j - k, \ell - \ell_1 + \ell_2)$.

Two more explicit formulae for $C(\ell_1, \ell_2, \ell; j, k, j+k)$ in terms of sums are given in Vilenkin [101], Chapter III, Section 8.3, Pages 181-182.

The Clebsch–Gordan coefficients enjoy several symmetry relations. These relations are discussed in Vilenkin [101], Chapter III, Section 8.4. For example, it can be shown that

$$
\begin{aligned}
C(\ell_1, \ell_2, \ell; j, k, j+k) &= (-1)^{\ell-\ell_1-\ell_2} C(\ell_1, \ell_2, \ell; -j, -k, -j-k) \\
C(\ell_1, \ell_2, \ell; j, k, j+k) &= (-1)^{\ell-\ell_1-\ell_2} C(\ell_1, \ell_2, \ell; k, j, j+k).
\end{aligned}
$$

Wigner came up with an ingenious device to formulate these symmetry relations. The *Wigner symbol* (also known as *3j-symbol*)

$$
\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix},
$$

which is zero unless $m_3 = -m_1 - m_2$, is defined in terms of the Clebsch–Gordan coefficients by the equation

$$
C(\ell_1, \ell_2, \ell; j, k, j+k) = (-1)^{\ell_1-\ell_2+j+k}\sqrt{2\ell+1} \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ j & k & -j-k \end{pmatrix}. \qquad \text{(CG16)}
$$

The Wigner symbol enjoys a total of 72 symmetries that can be formulated as follows. If we associate to the Wigner symbol the 3×3 matrix show below,

$$
\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \mapsto \begin{pmatrix} -j_1+j_2+j_3 & j_1-j_2+j_3 & j_1+j_2-j_3 \\ j_1-m_1 & j_2-m_2 & j_3-m_3 \\ j_1+m-1 & j_2+m_2 & j_3+m_3 \end{pmatrix},
$$

then for an even permutation of the rows or columns of the 3×3 matrix or under transposition, the Wigner symbol is unchanged, and for an odd permutation of the rows or columns it is multiplied by $(-1)^{j_1+j_2+j_3}$; see Vilenkin [101], Chapter III, Section 8.4.

More properties of the Clebsch–Gordan coefficients, including special values, expansions of products of the functions $P^{\ell}_{mn}(z)$, connections with Jacobi polynomials, recurrence formulae, and generating functions, can be found in Vilenkin [101], Chapter III, Sections 8.5-8.9.

Chapter 15

Induced Representations

If G is a locally compact group and if H is a closed subgroup of G, under certain conditions, it is possible to construct a Hilbert space $\mathcal{H}$ and a unitary representation $\Pi\colon G \to \mathbf{U}(\mathcal{H})$ of G in $\mathcal{H}$ from a unitary representation $U\colon H \to \mathbf{U}(E)$ of H in a (separable) Hilbert space E. The representation Π is called an *induced representation*. In particular, this construction can be used to define unitary representations of the group $\mathbf{SL}(2, \mathbb{R})$ which would be hard to find if we did not have this method.

There are two approaches for the construction of the Hilbert space $\mathcal{H}$:

1. The Hilbert space $\mathcal{H}$ is a set of functions from $X = G/H$ to E.

2. The Hilbert space $\mathcal{H}$ is a set of functions from G to E.

In the first approach, we will construct unitary representations of G in $\mathcal{H}$ using certain functions $\alpha\colon G \times (G/H) \to \mathbf{GL}(E)$ called *cocycles*. In the second approach, the construction of the Hilbert space $\mathcal{H}$ is more complicated, but the definition of the operator Π_s is simpler.

The general construction (in the first approach) consists of seven steps, where the first four are purely algebraic and do not deal with continuous unitary representations, but instead linear representations (group homomorphisms $U\colon G \to \mathbf{GL}(E)$, where G is a group not equipped with any topology and E is just a vector space with no additional structure):

(1) Let G be a group acting (on the left) on a set X and let E be a vector space. In Section 15.1 we define the notion of *equilinear action* of G on $X \times E$, which is an action of the form

$$s \cdot (x, z) = (s \cdot x, \alpha(s, x)(z)), \qquad s \in G, x \in X, z \in E,$$

where $\alpha(s, x)$ is a linear automorphism of E satisfying the conditions

(a) For all $x \in X$

$$\alpha(e, x) = \mathrm{id}_E.$$

(b) For all $x \in X$ and all $s, t \in G$,

$$\alpha(st, x) = \alpha(s, t \cdot x) \circ \alpha(t, x).$$

A map $\alpha\colon G \times X \to \mathbf{GL}(E)$ satisfying Conditions (a) and (b) is called a *cocycle of* G *with values in* $\mathbf{GL}(E)$. Conversely, an action of G on X and a cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$ determines an equilinear action of G on $X \times E$. Then we show that an equilinear action of G on $X \times E$ induces a homomorphism $\Pi\colon G \to \mathbf{GL}(E^X)$, where E^X is the vector space of all functions from X to E. More precisely, for every function $f\colon X \to E$, for every $s \in G$, $\Pi_s(f)\colon X \to E$ is function given by

$$(\Pi_s(f))(x) = \alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x)), \qquad \text{for every } x \in X.$$

(2) In Section 15.2 we specialize the construction to the homogeneous space $X = G/H$ of left cosets. Then G acts on G/H on the left by

$$s \cdot (gH) = sgH.$$

By choosing a set of representatives in the cosets of G/H, a cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$ determines a homomorphism $\sigma\colon H \to \mathbf{GL}(E)$ and a map $\beta\colon X \to \mathbf{GL}(E)$. Conversely, a homomorphism $\sigma\colon H \to \mathbf{GL}(E)$ and a map $\beta\colon X \to \mathbf{GL}(E)$ determine a cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$. This step is the most important application of Step 1, and E is an arbitrary vector space.

(3) For a given homomorphism $\sigma\colon H \to \mathbf{GL}(E)$, the homomorphisms $\Pi\colon G \to \mathbf{GL}(E^X)$ corresponding to cocycles associated with different maps β are equivalent.

(4) In Section 15.3 we show that a cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$ determines a bijection between E^X and a subspace L^α of the set E^G of maps from G to E. As a consequence, the homomorphism $\Pi\colon G \to \mathbf{GL}(E^X)$ corresponding to a cocycle α determines a homomorphism $\Pi_{L^\alpha}\colon G \to \mathbf{GL}(L^\alpha)$. This completes the purely algebraic construction. The next steps use topology and analysis to construct *unitary* representations.

(5) In Section 15.4 we assume that G is a locally compact group and H is a closed subgroup of G, in which case G/H is also locally compact. Let μ be a positive measure on $X = G/H$, and assume that E is a separable Hilbert space. Then we define a Hilbert space $\mathcal{L}^2_\mu(X; E)$ consisting of measurable functions from X to E.

(6) In Section 15.5, given a unitary representation U of H in E, we assume that the measure μ on $X = G/H$ is G-invariant and that the cocycle α satisfies the conditions:

(i) The linear automorphisms $\alpha(s, x)$ of E are unitary operators of E for all $s \in G$ and all $x \in G/H$, and $\alpha(h, x_0) = U(h)$ for all $h \in H$ (where x_0 denotes the coset H).

(ii) For every $s \in G$, for every $f \in \mathcal{L}^2_\mu(X; E)$, the map $x \mapsto \alpha(s, x)(f(x))$ from X to E is μ-measurable.

(iii) For every $f \in \mathcal{L}^2_\mu(X; E)$, the map $s \mapsto [\Pi_s(f)]$ from G to $\mathrm{L}^2_\mu(X; E)$ is continuous.

Then the homomorphism $s \mapsto \Pi_s([f]) = [\Pi_s(f)]$ is a unitary representation of G in $\mathrm{L}^2_\mu(X; E) = \mathcal{H}$.

(7) In Sections 15.6 and 15.7 we generalize the previous construction to certain measure called *quasi-invariant*. If the measure μ on G/H is quasi-invariant and another technical condition is satisfied, then the homomorphism $s \mapsto \Pi_s([f]) = [\Pi_s(f)]$ is a unitary representation of G in $\mathrm{L}^2_\mu(X; E)$. Quasi-invariant measures on G/H always exist and can be constructed using rho-functions. We conclude by showing how to construct unitary representations of $\mathbf{SL}(2, \mathbb{R})$ using induced representations. One example involves the action of $\mathbf{SL}(2, \mathbb{R})$ on the projective line $\mathbb{RP}^1$, and the other example involves the action of $\mathbf{SL}(2, \mathbb{R})$ on the upper half plane.

In Section 15.10 we consider a compact (metrizable) group G and a closed subgroup H of G, and our goal is to determine the canonical (unitary) representation of G in $\mathrm{L}^2_\mu(G/H; \mathbb{C})$ induced by the trivial representation of H in $E = \mathbb{C}$ (see Definition 15.13), where μ is the G-invariant measure on G/H induced by a Haar measure λ on G. For simplicity of notation we write $\mathrm{L}^2_\mu(G/H)$ instead of $\mathrm{L}^2_\mu(G/H; \mathbb{C})$. To do this it is necessary to understand what is the restriction of the representation $M_\rho \colon G \to \mathbf{U}(\mathbb{C}^{n_\rho})$ to H, with $\rho \in R(G)$.

In Proposition 15.18, we show that the space $\mathrm{L}^2_\mu(G/H)$ is the Hilbert sum of subspaces $L_\rho \subseteq \mathfrak{a}_\rho$. If the trivial representation σ_0 of H is contained $d = (\rho : \sigma_0) \geq 1$ times in the restriction of M_ρ to H, then L_ρ is the direct sum of the first d columns of M_ρ,

$$L_\rho = \bigoplus_{j=1}^{d} \mathfrak{l}^{(\rho)}_j \quad \text{and} \quad \mathfrak{l}^{(\rho)}_j = \bigoplus_{k=1}^{n_\rho} \mathbb{C} m^{(\rho)}_{kj}.$$

If $d = 0$, then $L_\rho = (0)$.

Then we consider the space $H\backslash G$ of right cosets Hs of G ($s \in G$). If $\pi \colon G \to H\backslash G$ is the quotient map $\pi(s) = Hs$, the fact that the Haar measure λ on a compact group is left and right invartiant implies immediately that there is a G-invariant measure μ' on $H\backslash G$. We show in Proposition 15.19 that the space $\mathrm{L}^2_{\mu'}(H\backslash G)$ is the Hilbert sum of subspaces $\check{\overline{L}}_\rho \subseteq \mathfrak{a}_\rho$. If the trivial representation σ_0 of H is contained $d = (\rho : \sigma_0) \geq 1$ times in the restriction of M_ρ to H, then $\check{\overline{L}}_\rho$ is the direct sum of the first d rows of M_ρ; that is,

$$\check{\overline{L}}_\rho = \bigoplus_{i=1}^{d} \bigoplus_{j=1}^{n_\rho} \mathbb{C} m^{(\rho)}_{ij}.$$

In preparation for Chapter 17 we consider the intersection $\mathrm{L}^2_\mu(G/H) \cap \mathrm{L}^2_{\mu'}(H\backslash G)$. This is a closed involutive subalgebra of $\mathrm{L}^2(G)$, thus a complete Hilbert algebra. We can view a function $g \in \mathrm{L}^2_\mu(G/H) \cap \mathrm{L}^2_{\mu'}(H\backslash G)$ as a function $g \in \mathrm{L}^2(G)$ such that

$$g(tst') = g(s) \qquad \text{for all } t, t' \in H \text{ and all } s \in G. \tag{$*_{H\backslash G/H}$}$$

We can also think of the functions $g \in \mathrm{L}^2_\mu(G/H) \cap \mathrm{L}^2_{\mu'}(H\backslash G)$ as functions defined on the *double classes (or double cosets) HsH* of G with respect to H.

We denote the algebra of functions in $\mathrm{L}^2(G)$ satisfying $(*_{H\backslash G/H})$ as $\mathrm{L}^2(H\backslash G/H)$. Then we show in Proposition 15.20 that the algebra $\mathrm{L}^2(H\backslash G/H)$ is the Hilbert sum of the minimal two-sided ideals

$$\mathfrak{a}_{\rho,\sigma_0} = L_\rho \cap \check{\overline{L}}_\rho = \bigoplus_{i=1}^{d} \bigoplus_{j=1}^{d} \mathbb{C} m_{ij}^{(\rho)}.$$

Each $\mathfrak{a}_{\rho,\sigma_0}$ is a matrix algebra of dimension d^2 having the family $(m_{ij}^{(\rho)})_{1\le i,j\le d}$ as a basis.

Again, in preparation for Chapter 17 on Gelfand pairs, we show in Proposition 15.21 that the algebra $\mathrm{L}^2(H\backslash G/H)$ is commutative if and only if $(\rho : \sigma_0) \le 1$ for all $\rho \in R(G)$. If so, then for every $\rho \in R(G)$ such that $(\rho : \sigma_0) = 1$, the ideal $\mathfrak{a}_{\rho,\sigma_0}$ is one-dimensional and is spanned by the function

$$\omega_\rho(s) = \frac{1}{n_\rho} m_{11}^{(\rho)}(s),$$

which is continuous and of positive type. Thus

$$\mathrm{L}^2(H\backslash G/H) = \bigoplus_{\rho \mid (\rho:\sigma_0)=1} \mathbb{C}\omega_\rho.$$

The function ω_ρ also satisfies the following equations:

$$\begin{aligned} \omega_\rho(tst') &= \omega_\rho(s), \qquad \text{for all } s \in G \text{ and all } t, t' \in H \\ \omega_\rho(e) &= 1. \end{aligned}$$

The function ω_ρ is called a *(zonal) spherical function*. Such functions are crucial in generalizing the notion of Fourier transform to a homogeneous space G/H. In Chapter 17 we will see how to achieve this when G is not compact (but H is compact). The key point is to consider pairs (G, H) for which $\mathrm{L}^2(H\backslash G/H)$ is commutative. Actually, we can't quite work with $\mathrm{L}^2(H\backslash G/H)$ because this space is not closed under convolution, but we will be able to work with another commutative algebra $\mathrm{L}^1(H\backslash G/H)$.

15.1 Cocycles and Induced Representations

As a warm up and as an example of the second approach, we consider the case where G is compact, H is a closed subgroup of G, and U is a linear representation of G in a *finite-dimensional* vector space E. This means that U is a homomorphism $U\colon G \to \mathbf{GL}(E)$ and that Condition (C) of Definition 12.1 is dropped.

Consider the Hilbert space $\mathrm{L}^2(G; E)$ consisting of all functions $f\colon G \to E$ such that for any orthonormal basis $(e_1, \ldots, e_n)$ of E, $f = f_1 e_1 + \cdots + f_n e_n$, where the f_i are functions in $\mathrm{L}^2(G)$; equivalently, $\mathrm{L}^2(G; E)$ is the finite Hilbert sum $\mathrm{L}^2(G; E) = \bigoplus_{i=1}^n \mathrm{L}^2(G) e_i$. The inner product of two functions $f = \sum_{i=1} f_i e_i$ and $g = \sum_{i=1} g_i e_i$ is

$$\langle f, g \rangle = \sum_{i=1}^n \int_G f_i(s) \overline{g_i(s)}\, d\lambda(s),$$

where λ is a Haar measure on G. This construction will be generalized in Section 15.4 to an infinite-dimensional Hilbert space. Consider the subspace $\mathcal{H}$ of $\mathrm{L}^2(G; E)$ consisting of all functions f such that

$$f(sh) = U(h^{-1})(f(s)), \qquad \text{for all } s \in G \text{ and all } h \in H. \tag{$*$}$$

It is easy to check that $\mathcal{H}$ is closed in $\mathrm{L}^2(G; E)$, so it is a Hilbert space. For any $f \in \mathcal{H}$, as before, let $\lambda_s f$ be the function given by

$$(\lambda_s f)(t) = f(s^{-1} t), \qquad s, t \in G.$$

For $s \in G$ fixed, the map $f \mapsto \lambda_s f$ is obviously linear. Observe that by $(*)$, for all $s, t \in G$, all $h \in H$, and all $f \in \mathcal{H}$, we have

$$(\lambda_t f)(sh) = f(t^{-1} sh) = U(h^{-1})(f(t^{-1} s)) = U(h^{-1})((\lambda_t f)(s)),$$

so $\lambda_t f \in \mathcal{H}$. For all $s, t, t' \in G$, we also have

$$(\lambda_{tt'} f)(s) = f((tt')^{-1} s) = f(t'^{-1} t^{-1} s) = (\lambda_{t'} f)(t^{-1} s) = \lambda_t((\lambda_{t'} f))(s).$$

If we define the map $\Pi\colon G \to \mathbf{GL}(\mathcal{H})$ by

$$\Pi_s(f) = \lambda_s f, \qquad s \in G,\ f \in \mathcal{H},$$

equivalently

$$(\Pi_s(f))(t) = f(s^{-1} t), \qquad s, t \in G,\ f \in \mathcal{H},$$

then we see that Π is a linear representation of G in $\mathcal{H}$ (Condition (C) of Definition 12.1 may fail, but here we are not considering continuous representations). Since the Haar measure is left and right invariant, the maps $\lambda_t f$ are unitary ($f \in \mathcal{H}$), so Π is a unitary representation of G in $\mathcal{H}$, called the representation *induced* by U.

It is easy to see that if we replace U by an equivalent representation $h \mapsto PU(h)P^{-1}$, where P is a unitary transformation $P\colon E \to E'$, then the corresponding induced representation is $s \mapsto f_P \Pi_s f_P^{-1}$, a unitary representation equivalent to Π, where f_P is the linear map from $\mathcal{H}$ to $\mathcal{H}'$ given by $f_P(f) = P \circ f$. Therefore, the above construction defines a class of unitary representations of G induced by a class of linear representations of H.

Let us now consider a more general situation. Our first construction is purely algebraic and does not assume that the group G or the vector space E have any topology. As a consequence, until Section 15.4 we consider linear representations of G in E; these are simply homomorphisms $U\colon G \to \mathbf{GL}(E)$, with no continuity requirement.

Definition 15.1. If we have a left group action $\cdot\colon G \times X \to X$ of a group G on a set X, for any vector space E, a left action $\cdot\colon G \times (X \times E) \to X \times E$ is *equilinear* if there is some function $\alpha\colon G \times X \to \mathbf{GL}(E)$ such that

$$s \cdot (x, z) = (s \cdot x, \alpha(s, x)(z)), \qquad \text{for all } s \in G, \text{ all } x \in X, \text{ and all } z \in E.$$

The crucial property of an equilinear action is that the second component $pr_2(s \cdot (x, z))$ of the action of $s \in G$ on $(x, z) \in X \times E$ given by

$$s \cdot (x, z) = (s \cdot x, pr_2(s \cdot (x, z)))$$

is *linear* in z. This is the reason for introducing the linear isomorphism $\alpha(s, x)$ given by $\alpha(s, x)(z) = pr_2(s \cdot (x, z))$.

If we have an equilinear action $\cdot\colon G \times (X \times E) \to X \times E$, then the conditions for being a left action are

$$\begin{aligned} e \cdot (x, z) &= (x, z) \\ (st) \cdot (x, z) &= s \cdot (t \cdot (x, z)), \end{aligned}$$

which translate to

$$\begin{aligned} (e \cdot x, \alpha(e, x)(z)) &= (x, z) \\ ((st) \cdot x, \alpha(st, x)(z)) &= s \cdot (t \cdot x, \alpha(t, x)(z)) \\ &= (s \cdot (t \cdot x), \alpha(s, t \cdot x)(\alpha(t, x)(z))), \end{aligned}$$

so we must have

$$\begin{aligned} \alpha(e, x)(z) &= z \\ \alpha(st, x)(z) &= (\alpha(s, t \cdot x) \circ \alpha(t, x))(z), \end{aligned}$$

for all $s, t \in G$, all $x \in X$, and all $z \in E$. By reversing the above computations, we see that if a function $\alpha\colon G \times X \to \mathbf{GL}(E)$ satisfies the above two conditions, then the map given by

$$s \cdot (x, z) = (s \cdot x, \alpha(s, x)(z)), \qquad \text{for all } s \in G, \text{ all } x \in X, \text{ and all } z \in E$$

is an equilinear action. In summary, we proved the following proposition.

Proposition 15.1. *Given a left group action $\cdot\colon G \times X \to X$ and a vector space E, for any function $\alpha\colon G \times X \to \mathbf{GL}(E)$, the map $\cdot\colon G \times (X \times E) \to X \times E$ given by*

$$s \cdot (x, z) = (s \cdot x, \alpha(s, x)(z)), \qquad \textit{for all } s \in G, \textit{ all } x \in X, \textit{ and all } z \in E$$

is an equilinear action if and only if the following two conditions hold:

(a) For all $x \in X$

$$\alpha(e, x) = \mathrm{id}_E.$$

(b) *For all $x \in X$ and all $s, t \in G$,*

$$\alpha(st, x) = \alpha(s, t \cdot x) \circ \alpha(t, x).$$

In view or Proposition 15.1, we make the following definition.

Definition 15.2. Let G be a left action of a group G on a set X, and let E be a vector space. Let $\alpha \colon G \times X \to \mathbf{GL}(E)$ be a function and assume that the following conditions hold:

(a) For all $x \in X$

$$\alpha(e, x) = \mathrm{id}_E.$$

(b) For all $x \in X$ and all $s, t \in G$,

$$\alpha(st, x) = \alpha(s, t \cdot x) \circ \alpha(t, x).$$

A map $\alpha \colon G \times X \to \mathbf{GL}(E)$ satisfying Conditions (a) and (b) is called a *cocycle of G with values in* $\mathbf{GL}(E)$.

The point of equilinear actions is that they yield homomorphisms $\Pi \colon G \to \mathbf{GL}(E^X)$, that is, linear representations of G in the vector space $[X \to E] = E^X$. We just explained before Definition 15.2 how a cocycle defines an equilinear action. The reader may wonder where cocycles come from. The answer will be given in the next section; they are induced by linear representations of subgroups of G.

Given an equilinear action $\cdot \colon G \times (X \times E) \to X \times E$, we obtain an action Π of G on E^X as follows: for every $s \in G$, for every $f \in E^X$, the function $\Pi_s(f) \in E^X$ is given by the equation

$$s \cdot (x, f(x)) = (s \cdot x, (\Pi_s(f))(s \cdot x)) \quad \text{for all } x \in X.$$

Using Definition 15.1, the above equation is equivalent to

$$(\Pi_s(f))(s \cdot x) = \alpha(s, x)(f(x)), \quad \text{for all } x \in X,$$

which is equivalent to

$$(\Pi_s(f))(x) = \alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x)), \quad \text{for all } x \in X.$$

We are led to the following definition.

Definition 15.3. Let G be a left action of a group G on a set X, and let E be a vector space. For every equilinear action $\cdot \colon G \times (X \times E) \to X \times E$ defined by a cocycle $\alpha \colon G \times X \to \mathbf{GL}(E)$, for every function $f \colon X \to E$, for every $s \in G$, let $\Pi^\alpha_s(f) \colon X \to E$ be the function given by

$$(\Pi^\alpha_s(f))(x) = \alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x)), \qquad \text{for every } x \in X. \tag{Π^α_s}$$

The above equation defines a map $\Pi^\alpha_s \colon E^X \to E^X$. The map $\Pi^\alpha \colon G \to \mathbf{GL}(E^X)$ given by $s \mapsto \Pi^\alpha_s$ is the *(linear) representation of G in E^X induced by the cocycle α*. For simplicity of notation, we write Π instead of Π^α.

The following proposition confirms that the map Π is a linear representation of G in the vector space E^X.

Proposition 15.2. *Let G be a left action of a group G on a set X, and let E be a vector space. For every equilinear action $\cdot\colon G \times (X \times E) \to X \times E$ defined by a cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$, for every $s \in G$, the map $\Pi_s\colon E^X \to E^X$ is a linear isomorphism, and the map $\Pi\colon G \to \mathbf{GL}(E^X)$ given by $s \mapsto \Pi_s$ is a homomorphism, that is, a linear representation of G in the vector space E^X.*

Proof. Since $\alpha(s, s^{-1} \cdot x)$ is a linear automorphism of E, we have

$$\begin{aligned}
(\Pi_s(f_1 + f_2))(x) &= \alpha(s, s^{-1} \cdot x)((f_1 + f_2)(s^{-1} \cdot x)) \\
&= \alpha(s, s^{-1} \cdot x)(f_1(s^{-1} \cdot x) + f_2(s^{-1} \cdot x)) \\
&= \alpha(s, s^{-1} \cdot x)(f_1(s^{-1} \cdot x)) + \alpha(s, s^{-1} \cdot x)(f_2(s^{-1} \cdot x)) \\
&= (\Pi_s(f_1))(x) + (\Pi_s(f_2))(x)),
\end{aligned}$$

and for every $\lambda \in \mathbb{C}$,

$$\begin{aligned}
(\Pi_s(\lambda f))(x) &= \alpha(s, s^{-1} \cdot x)((\lambda f)(s^{-1} \cdot x)) \\
&= \alpha(s, s^{-1} \cdot x)(\lambda f(s^{-1} \cdot x)) \\
&= \lambda \alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x)) \\
&= \lambda(\Pi_s(f))(x),
\end{aligned}$$

so the map $f \mapsto \Pi_s(f)$ from E^X to itself is linear. Given any fixed $s \in G$, for every function $g\colon X \to E$, we have $\Pi_s(f) = g$ iff $(\Pi_s(f))(x) = g(x)$ for all $x \in X$ iff

$$\alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x)) = g(x) \quad \text{for all } x \in X,$$

and since $\alpha(s, s^{-1} \cdot x)$ is an invertible linear map, we must have

$$f(s^{-1} \cdot x) = (\alpha(s, s^{-1} \cdot x))^{-1}(g(x)) \quad \text{for all } x \in X,$$

so if we write $y = s^{-1} \cdot x$, then $x = s \cdot y$ and since the map $y \mapsto s \cdot y$ is a bijection (because $\cdot$ is a group action of G on X), we have

$$f(y) = (\alpha(s, y))^{-1}(g(s \cdot y)) \quad \text{for all } y \in X,$$

which shows that f is uniquely determined and thus that Π_s is a bijection.

For all $s, t \in G$, we have

$$g(y) = (\Pi_t(f))(y) = \alpha(t, t^{-1} \cdot y)(f(t^{-1} \cdot y)),$$

so

$$\begin{aligned}
\big(\Pi_s(\Pi_t(f))\big)(x) &= (\Pi_s(g))(x) \\
&= \alpha(s, s^{-1} \cdot x)(g(s^{-1} \cdot x)) \\
&= \big(\alpha(s, s^{-1} \cdot x) \circ \alpha(t, t^{-1} \cdot (s^{-1} \cdot x))\big)(f(t^{-1} \cdot (s^{-1} \cdot x))),
\end{aligned}$$

and we also have

$$\begin{aligned}
(\Pi_{st}(f))(x) &= \alpha(st, (st)^{-1}\cdot x)(f((st)^{-1}\cdot x)) \\
&= \big(\alpha(s, t\cdot((t^{-1}s^{-1})\cdot x)) \circ \alpha(t, (t^{-1}s^{-1})\cdot x)\big)(f(t^{-1}s^{-1}\cdot x)) \\
&= \big(\alpha(s, s^{-1}\cdot x) \circ \alpha(t, t^{-1}\cdot(s^{-1}\cdot x))\big)(f(t^{-1}\cdot(s^{-1}\cdot x))) \\
&= \big(\Pi_s(\Pi_t(f))\big)(x),
\end{aligned}$$

which proves that $\Pi_{st}(f) = (\Pi_s \circ \Pi_t)(f)$, that is, Π is a homomorphism. $\square$

If we let $t = s^{-1}$ in (b) of Definition 15.2, we obtain

$$\alpha(s^{-1}, x) = (\alpha(s, s^{-1}\cdot x))^{-1},$$

so $\Pi_s(f)$ can also be written as

$$(\Pi_s(f))(x) = (\alpha(s^{-1}, x))^{-1}(f(s^{-1}\cdot x)). \tag{Π_s}$$

15.2 Cocycles on a Homogeneous Space $X = G/H$

We now consider the special case where $X = G/H$ is the homogeneous space of left cosets for some subgroup H of G, and the left action of G acts on G/H given by

$$s\cdot(gH) = sgH.$$

Definition 15.4. Given a group G and a subgroup H of G, a *set of representatives* $(r_x)_{x\in G/H}$ for the cosets of G/H is the choice for every coset $x \in G/H$ of some element $r_x \in G$ so that $x = r_xH$. Then every element g of $x = r_xH$ is written uniquely as $g = r_xh$, with $h \in H$. We denote the coset H by x_0 and pick $r_{x_0} = e$. For any $s \in G$, the representative of $s\cdot x = s\cdot r_xH = sr_xH$ is denoted by $r_{s\cdot x}$.

If we denote the quotient map by $\pi\colon G \to G/H$, then picking a set of representatives $(r_x)_{x\in G/H}$ in the cosets of G/H is equivalent to picking a *section of* π, that is, a map $r\colon G/H \to G$ such that $\pi\circ r = \mathrm{id}_{G/H}$.

Since for every coset $x \in H$ we have $x = r_x\cdot x_0$ (the class r_xH, which is x), Condition (b) of Definition 15.2 yields

$$\alpha(sr_x, x_0) = \alpha(s, r_x\cdot x_0)\circ\alpha(r_x, x_0) = \alpha(s, x)\circ\alpha(r_x, x_0),$$

and so

$$\alpha(s, x) = \alpha(sr_x, x_0)\circ(\alpha(r_x, x_0)))^{-1}. \tag{$*_1$}$$

Equation $(*_1)$ shows that the automorphisms $\alpha(s, x_0)$ of E determine the $\alpha(s, x)$ for all $x \in X$.

Denote $\alpha(s, x_0)$ by $\alpha_0(s)$. Conditions (a) and (b) of Definition 15.2 imply that

$$\begin{aligned}\alpha_0(e) &= \mathrm{id}_E \\ \alpha(sh, x_0) &= \alpha(s, h \cdot x_0) \circ \alpha(h, x_0),\end{aligned}$$

for all $s \in G$ and all $h \in H$, and since $h \cdot x_0 = x_0$, we get

$$\alpha_0(sh) = \alpha_0(s) \circ \alpha_0(h) \qquad \text{for all } s \in G \text{ and all } h \in H. \tag{$*_2$}$$

Now sr_x belongs to the coset $sr_xH = s \cdot r_xH = s \cdot x = r_{s\cdot x}H$, so there is a unique element of H, denoted $u(s, x)$, such that

$$sr_x = r_{s\cdot x}u(s, x), \tag{$\dagger$}$$

and by $(*_2)$,

$$\alpha_0(sr_x) = \alpha_0(r_{s\cdot x}u(s, x)) = \alpha_0(r_{s\cdot x}) \circ \alpha_0(u(s, x)),$$

so $(*_1)$ can be written as

$$\alpha(s, x) = \alpha_0(r_{s\cdot x}) \circ \alpha_0(u(s, x)) \circ (\alpha_0(r_x))^{-1}. \tag{$*_3$}$$

Definition 15.5. Given $\alpha\colon G \times X \to \mathbf{GL}(E)$ as in Definition 15.2, for all $s \in G$, all $h \in H$, and all $x \in X$, define $\alpha_0(s), \sigma(h), \beta(x)$ and $u(s, x)$ by

$$\begin{aligned}\alpha_0(s) &= \alpha(s, x_0) \\ \sigma(h) &= \alpha(h, x_0) = \alpha_0(h) \\ \beta(x) &= \alpha(r_x, x_0) = \alpha_0(r_x) \\ u(s, x) &= r_{s\cdot x}^{-1} s r_x \in H.\end{aligned} \tag{u}$$

Then $(*_3)$ becomes

$$\alpha(s, x) = \beta(s \cdot x) \circ \sigma(u(s, x)) \circ (\beta(x))^{-1}, \tag{$*_4$}$$

and $(*_2)$ implies that

$$\sigma(h_1h_2) = \sigma(h_1) \circ \sigma(h_2) \qquad \text{for all } h_1, h_2 \in H, \tag{$*_5$}$$

which shows that $\sigma\colon H \to \mathbf{GL}(E)$ is a homomorphism. Thus the restriction of the cocycle α to $H \times \{x_0\}$ is a representation of H in E.

Conversely, let $\sigma\colon H \to \mathbf{GL}(E)$ be any homomorphism, and let $\beta\colon X \to \mathbf{GL}(E)$ be any function. Then we define the function $u\colon G \times G/H \to H$ using the equation

$$u(s, x) = r_{s\cdot x}^{-1} s r_x$$

of Definition 15.5, and the function $\alpha\colon G \times X \to \mathbf{GL}(E)$ given by $(*_4)$, namely

$$\alpha(s, x) = \beta(s \cdot x) \circ \sigma(u(s, x)) \circ (\beta(x))^{-1}.$$

Observe that for all $s, t \in G$ and all $x \in X$, we have

$$u(st, x) = u(s, t \cdot x)u(t, x), \tag{$*_h$}$$

because by (†)

$$\begin{aligned} str_x &= r_{(st)\cdot x} u(st, x) \\ sr_{t\cdot x} &= r_{s\cdot(t\cdot x)} u(s, t\cdot x) \\ &= r_{(st)\cdot x} u(s, t\cdot x) \\ tr_x &= r_{t\cdot x} u(t, x), \end{aligned}$$

so we have

$$r_{(st)\cdot x} u(st, x) = str_x = sr_{t\cdot x} u(t, x) = r_{(st)\cdot x} u(s, t\cdot x) u(t, x),$$

and since $r_{(st)\cdot x} \in G$, it is invertible, which proves $(*_h)$. The verification that $\alpha(e, x) = \mathrm{id}_E$ is immediate, since $e \cdot x = x$, so $u(e, x) = e$, $\beta(e \cdot x) = \beta(x)$, and $\sigma(e) = \mathrm{id}_E$. Using $(*_h)$ and the fact that σ is a homomorphism, we have

$$\begin{aligned} \alpha(st, x) &= \beta(st\cdot x) \circ \sigma(u(st, x)) \circ (\beta(x))^{-1} \\ &= \beta(st\cdot x) \circ \sigma(u(s, t\cdot x)u(t, x)) \circ (\beta(x))^{-1} \\ &= \beta(s\cdot(t\cdot x)) \circ \sigma(u(s, t\cdot x)) \circ \sigma(u(t, x)) \circ (\beta(x))^{-1} \\ &= \beta(s\cdot(t\cdot x)) \circ \sigma(u(s, t\cdot x)) \circ \beta(t\cdot x)^{-1} \circ \beta(t\cdot x) \circ \sigma(u(t, x)) \circ (\beta(x))^{-1} \\ &= \alpha(s, t\cdot x) \circ \alpha(t, x), \end{aligned}$$

which shows that α is a cocycle. In summary, we obtained the following result.

Proposition 15.3. *Let G be a group, H be a subgroup of G, and E be a vector space. Choose a set $(r_x)_{x\in G/H}$ of representatives for the cosets of $X = G/H$ as explained above, with $x_0 = H$ and $r_{x_0} = e$. Every cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$ determines a homomorphism $\sigma\colon H \to \mathbf{GL}(E)$ with $\sigma(h) = \alpha(h, x_0)$ for all $h \in H$, a map $\beta\colon X \to \mathbf{GL}(E)$ given by $\beta(x) = \alpha(r_x, x_0)$ for all $x \in X$, and a map $u\colon G \times G/H \to H$ given by $u(s, x) = r_{s\cdot x}^{-1} sr_x \in H$, such that*

$$\alpha(s, x) = \beta(s\cdot x) \circ \sigma(u(s, x)) \circ (\beta(x))^{-1}.$$

Conversely, given a homomorphism $\sigma\colon H \to \mathbf{GL}(E)$ and a map $\beta\colon X \to \mathbf{GL}(E)$, if we set

$$u(s, x) = r_{s\cdot x}^{-1} sr_x \tag{u}$$

and

$$\alpha(s, x) = \beta(s\cdot x) \circ \sigma(u(s, x)) \circ (\beta(x))^{-1}, \tag{α}$$

then $\alpha\colon G \times X \to \mathbf{GL}(E)$ is a cocycle.

Remark: Kirillov [54] (Appendix V, Section 2.1) calls (u) the *Master equation*. See also Proposition 5, Lemma 2, and Lemma 3. This material is also discussed in Kririllov [53] (Sections 13.1 and 13.2).

In view of Proposition 15.3 we make the following definition.

Definition 15.6. Given a homomorphism $\sigma\colon H \to \mathbf{GL}(E)$ and a map $\beta\colon X \to \mathbf{GL}(E)$, if α is the cocycle associated with σ and β, we say that the representation Π^α of G in E^X defined by α is the *representation induced by* σ *and* β.

Remarkably, for a given homomorphism $\sigma\colon H \to \mathbf{GL}(E)$, the representations $\Pi_1\colon G \to \mathbf{GL}(E^X)$ and $\Pi_2\colon G \to \mathbf{GL}(E^X)$ corresponding to the cocycles α_1 and α_2 associated with two maps β_1 and β_2 are equivalent, in the sense that there is an automorphism γ of E^X such that

$$\Pi_2 = \gamma \circ \Pi_1 \circ \gamma^{-1}.$$

This is proven as follows.

Proposition 15.4. *Let G be a group, H be a subgroup of G, and E be a vector space. Choose a set $(r_x)_{x\in G/H}$ of representatives for the cosets of $X = G/H$ as explained above, with $x_0 = H$ and $r_{x_0} = e$. Let $\sigma\colon H \to \mathbf{GL}(E)$ be a homomorphism, let $\beta\colon X \to \mathbf{GL}(E)$ be a map, and let α be the cocycle determined by σ and β as in Proposition 15.3, and let $\Pi\colon G \to \mathbf{GL}(E^X)$ be the corresponding representation. If $c(x) = \beta(x)^{-1}$ for all $x \in X$, then define the automorphism γ of E^X by*

$$(\gamma(f))(x) = c(x)(f(x)), \qquad f \in E^X,\ x \in X.$$

Then the representation

$$\Pi' = \gamma \circ \Pi \circ \gamma^{-1}$$

is associated with the cocycle α' given by

$$\alpha'(s, x) = \sigma(u(s, x)), \tag{α'}$$

with

$$u(s, x) = r_{s\cdot x}^{-1} s r_x. \tag{u}$$

Thus, the representation Π induced by σ and β is equivalent to the representation induced by σ and β', with $\beta'(x) = \mathrm{id}_E$ for all $x \in X$. The induced representation Π' associated with α' is given by

$$(\Pi'_s(f))(x) = \sigma(u(s^{-1}, x)^{-1})(f(s^{-1}\cdot x)) = \sigma(u(s, s^{-1}\cdot x))(f(s^{-1}\cdot x)),\ \ f \in E^X,\ x \in X. \tag{Π'}$$

Proof. Consider any map $c\colon X \to \mathbf{GL}(E)$ such that $c(x_0) = \mathrm{id}_E$. Define the automorphism γ of E^X by

$$(\gamma(f))(x) = c(x)(f(x)), \qquad f \in E^X,\ x \in X,$$

and let

$$\Pi' = \gamma \circ \Pi \circ \gamma^{-1}.$$

Since γ is an automorphism of E^X, the map Π' is a linear representation of G in E^X. Clearly, the inverse of γ is given by

$$(\gamma^{-1}(f))(x) = c(x)^{-1}(f(x)), \qquad f \in E^X,\ x \in X,$$

and since for any $g \in E^X$, we have

$$(\Pi_s(g))(x) = \alpha(s, s^{-1} \cdot x)(g(s^{-1} \cdot x)),$$

with $g = \gamma^{-1}(f)$, we obtain

$$\begin{aligned}(\Pi_s(\gamma^{-1})(f))(x) &= \alpha(s, s^{-1} \cdot x)(\gamma^{-1}(f)(s^{-1} \cdot x)) \\ &= \alpha(s, s^{-1} \cdot x)\big(c(s^{-1} \cdot x)^{-1}(f(s^{-1} \cdot x))\big),\end{aligned}$$

and so

$$c(x)\big((\Pi_s(\gamma^{-1})(f))(x)\big) = c(x)\big(\alpha(s, s^{-1} \cdot x)\big(c(s^{-1} \cdot x)^{-1}(f(s^{-1} \cdot x))\big)\big);$$

that is,

$$(\Pi'_s(f))(x) = \big(c(x) \circ \alpha(s, s^{-1} \cdot x) \circ c(s^{-1} \cdot x)^{-1}\big)(f(s^{-1} \cdot x)), \tag{$*_6$}$$

which shows that Π' is obtained from α' as Π is obtained from α, with

$$\alpha'(s, x) = c(s \cdot x) \circ \alpha(s, x) \circ c(x)^{-1}. \tag{$*_7$}$$

If we write $\alpha'_0(s) = \alpha'(s, x_0)$ and $\beta'(x) = \alpha'_0(r_x)$ as before, then the hypothesis $c(x_0) = \mathrm{id}_E$ implies that

$$\alpha'_0(h) = \alpha'(h, x_0) = c(h \cdot x_0) \circ \alpha(h, x_0) \circ c(x_0)^{-1} = c(x_0) \circ \sigma(h) \circ c(x_0)^{-1} = \sigma(h)$$

for all $h \in H$, and

$$\begin{aligned}\beta'(x) &= \alpha'(r_x, x_0) \\ &= c(r_x \cdot x_0) \circ \alpha(r_x, x_0) \circ c(x_0)^{-1} \\ &= c(x) \circ \alpha(r_x, x_0) \\ &= c(x) \circ \beta(x);\end{aligned}$$

that is,

$$\beta'(x) = c(x) \circ \beta(x). \tag{$*_8$}$$

Since $\beta(x_0) = \mathrm{id}_E$, we can pick $c(x) = \beta(x)^{-1}$, and then

$$\beta'(x) = \mathrm{id}_E(x) \qquad \text{for all } x \in X, \tag{$*_9$}$$

and since

$$\alpha(s, x) = \beta(s \cdot x) \circ \sigma(u(s, x)) \circ (\beta(x))^{-1},$$

from $(*_7)$ we obtain

$$\alpha'(s, x) = \sigma(u(s, x)). \tag{$*_{10}$}$$

Therefore, Π is equivalent to Π' with $\beta'(x) = \mathrm{id}_E$ for all $x \in X$. $\square$

It is also easy to check that if σ is replaced by an equivalent representation σ' of H in E^X, then the corresponding representations Π and Π' of G in E^X are equivalent.

Therefore, the process for making a representation Π of G in E^X from a representation σ of H in E and a function $\beta\colon X \to \mathbf{GL}(E)$ defines a class of representations of G in E^X. Furthermore, there is a special representation associated with σ and the constant function β given by $\beta(x) = \mathrm{id}_E$, for all $x \in X$.

In summary, the method is find a set $(r_x)_{x\in G/H}$ of representatives for the cosets of G/H, then to construct u given by $u(s,x) = r_{s\cdot x}^{-1} s r_x$ as in Equation (u), and then to define α by $\alpha(s,x) = \sigma(u(s,x))$. The induced representation is given by

$$(\Pi_s(f))(x) = \sigma(u(s, s^{-1}\cdot x))(f(s^{-1}\cdot x)), \;\; f \in E^X,\; x \in X \tag{$*$}$$

Vilenkin [101] (Chapter 1, Section 7) calls such a representation a *representation with operator factor*.

From a theoretical point of view, a cocycle α is equivalent to a pair (σ, β) as in Proposition 15.3, but from a practical point of view, it may be very hard (if not impossible) to find constructively a set $(r_x)_{x\in G/H}$ of representatives for the cosets of G/H. Thus we use cocycles α that agree with a given representation $\sigma\colon H \to \mathbf{GL}(E)$, in the sense that $\alpha(h, x_0) = \sigma(h)$ for all $h \in H$.

A case of practical case interest in equivariant machine learning is the case where $G = \mathbf{SE}(3)$ and $H = \mathbf{SO}(3)$.

Example 15.1. Let $G = \mathbf{SE}(3)$ and $H = \mathbf{SO}(3)$. The group $\mathbf{SE}(3)$ is the group of affine rigid motions of $\mathbb{R}^3$ consisting of rotations and translations. Here we view $\mathbf{SE}(3)$ as the group of matrices

$$s = \begin{pmatrix} Q & a \\ 0 & 1 \end{pmatrix}, \quad Q \in \mathbf{SO}(3),\; a \in \mathbb{R}^3$$

under multiplication. For short we denote the above matrix by (a, Q). The group $\mathbf{SE}(3)$ acts on $\mathbb{R}^3$ by

$$(a, Q)\cdot x = Qx + a, \quad x \in \mathbb{R}^3.$$

Multiplication in $\mathbf{SE}(n)$ is given by

$$(a, Q)(b, R) = (a + Qb, QR),$$

and the inverse of (a, Q) is

$$(a, Q)^{-1} = (-Q^\top a, Q^\top).$$

For details on $\mathbf{SE}(3)$ and the fact that it is a semi-direct product of $\mathbb{R}^3$ and $\mathbf{SO}(3)$, see Example 16.1. It is easy to see that the homogeneous space $\mathbf{SE}(3)/\mathbf{SO}(3)$ is $\mathbb{R}^3$. Indeed $\mathbf{SE}(3)$ acts on $\mathbb{R}^3$, and the stabilizer of the origin 0_3 is $\mathbf{SO}(3)$ viewed as the set of matrices

$$\begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}, \quad Q \in \mathbf{SO}(3).$$

We now use the method based on Proposition 15.3 and Proposition 15.4 to construct an induced representation of $\mathbf{SE}(3)$ from a representation $\sigma\colon \mathbf{SO}(3) \to \mathbf{GL}(E)$ of $\mathbf{SO}(3)$. For this we need to find a set of representative for the cosets of $\mathbb{R}^3 = \mathbf{SE}(3)/\mathbf{SO}(3)$ in order to define u, and then $\alpha(s, x)$ is given by $\alpha(s, x) = \sigma(u(s, x))$ and the induced representation Π is given by $(*)$. This is a case where it is easy to pick a set of coset representatives, namely for each $x \in \mathbb{R}^3$, $r_x \in \mathbf{SE}(3)$ is the matrix

$$\begin{pmatrix} I_3 & x \\ 0 & 1 \end{pmatrix},$$

the translation by x. The coset $x\mathbf{SO}(3)$ consists of the matrices

$$\begin{pmatrix} Q & x \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} I_3 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}$$

with x fixed. Let us compute $u(s, x) = r_{s\cdot x}^{-1} s r_x$. First $s \cdot x = (a, Q) \cdot x = Qx + a$, so

$$r_{s\cdot x} = \begin{pmatrix} I_3 & Qx + a, \\ 0 & 1 \end{pmatrix}, \quad r_{s\cdot x}^{-1} = \begin{pmatrix} I_3 & -Qx - a, \\ 0 & 1 \end{pmatrix},$$

and finally

$$\begin{aligned} u(s, x) = r_{s\cdot x}^{-1} s r_x &= \begin{pmatrix} I_3 & -Qx - a, \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_3 & x \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} Q & -Qx \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_3 & x \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Consequently, if $\sigma\colon \mathbf{SO}(3) \to \mathbf{GL}(E)$ is any representation of $\mathbf{SO}(3)$ is a finite-dimensional (nontrivial) vector space E, the above shows that $u(s, x)$ is independent of x and given by

$$u(s, x) = u((a, Q), x) = Q$$

and so $\alpha((a, Q), x)$ is given by

$$\alpha((a, Q), x) = \sigma(u((a, Q), x)) = \sigma(Q).$$

Then by $(*)$ we obtain the representation $\Pi\colon \mathbf{SE}(3) \to \mathbf{GL}(E^{\mathbb{R}^3})$ of $\mathbf{SE}(3)$ in $E^{\mathbb{R}^3}$ given by

$$\begin{aligned} (\Pi_{(a,Q)}(f))(x) &= \sigma(u(s, s^{-1} \cdot x))(f(s^{-1} \cdot x)) \\ &= \sigma(Q) f((a, Q)^{-1} \cdot x) = \sigma(Q) f(Q^\top (x - a)), \end{aligned}$$

that is,

$$(\Pi_{(a,Q)}(f))(x) = \sigma(Q) f(Q^\top (x - a)), \quad f \in E^{\mathbb{R}^3},\ x \in \mathbb{R}^3.$$

Since the vector space $E^{\mathbb{R}^3}$ is infinite-dimensional, even if σ is irreducible, this representation is reducible because its restriction to $\mathbf{SO}(3)$ is reducible (since the irreducible representations of $\mathbf{SO}(3)$ are finite-dimensional).

15.3 Converting Induced Representations of G From E^X to E^G

We can also show that a cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$ defines an isomorphism τ between the space E^X and a subspace L^α of the space E^G.

Definition 15.7. Let G be a group, H be a subgroup of G, E be a vector space, and write $X = G/H$. Given any cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$, for any function $f\colon X \to E$, the function $f^\alpha\colon G \to E$ is given by

$$f^\alpha(s) = \alpha(s^{-1}, s \cdot x_0)(f(s \cdot x_0)) = (\alpha(s, x_0))^{-1}(f(s \cdot x_0)), \qquad \text{for all } s \in G, \qquad (*_{\alpha_1})$$

with $x_0 = H$.

Recall from Definition 15.5 that $\sigma(h) = \alpha(h, x_0)$ for all $h \in H$.

Proposition 15.5. *With the hypotheses of Definition 15.7, the function f^α satisfies the equation*

$$f^\alpha(sh) = \sigma(h^{-1})(f^\alpha(s)), \qquad \text{for all } h \in H \text{ and all } s \in G. \qquad (*_{\alpha_2})$$

Proof. By (b) of Definition 15.2 and since $h \cdot x_0 = x_0$, we have

$$\begin{aligned}
f^\alpha(sh) &= \alpha((sh)^{-1}, (sh) \cdot x_0)(f((sh) \cdot x_0)) \\
&= \alpha(h^{-1}s^{-1}, (sh) \cdot x_0)(f((sh) \cdot x_0)) \\
&= \big(\alpha(h^{-1}, s^{-1} \cdot (s \cdot (h \cdot x_0))) \circ \alpha(s^{-1}, s \cdot (h \cdot x_0))\big)(f(s \cdot (h \cdot x_0))) \\
&= \big(\alpha(h^{-1}, x_0) \circ \alpha(s^{-1}, s \cdot x_0)\big)(f(s \cdot x_0)) \\
&= \sigma(h^{-1})(f^\alpha(s)),
\end{aligned}$$

establishing the proposition. ∎

Definition 15.8. Let G be a group, H be a subgroup of G, E be a vector space, and write $X = G/H$. Given any cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$, let L^α be the subspace of E^G consisting of all functions $g\colon G \to E$ such that

$$g(sh) = \sigma(h^{-1})(g(s)), \qquad \text{for all } s \in G \text{ and all } h \in H, \qquad (*_{\alpha_3})$$

where $\sigma(h) = \alpha(h, x_0)$, for all $h \in H$ (with $x_0 = H$).

Proposition 15.6. *With the hypotheses of Definition 15.7, for every $g \in L^\alpha$, there is a unique function $f\colon E \to X$ such that $g = f^\alpha$. Therefore, the map $\tau\colon E^X \to L^\alpha$ given by $\tau(f) = f^\alpha$ is an isomorphism.*

Proof. Note that the function $s \mapsto \alpha(s, x_0)(g(s))$ has the same value if s is replaced by sh for every $h \in H$, since by (b) of Definition 15.2, $(*_{\alpha 3})$, and the facts that $\sigma(h) = \alpha(h, x_0)$ and $h \cdot x_0 = x_0$ for $h \in H$,

$$\begin{aligned} \alpha(sh, x_0)(g(sh)) &= \big(\alpha(s, h \cdot x_0) \circ \alpha(h, x_0)\big)(g(sh)) \\ &= \big(\alpha(s, x_0) \circ \alpha(h, x_0)\big)\big(\sigma(h^{-1})(g(s))\big) \\ &= \big(\alpha(s, x_0) \circ \sigma(h) \circ \sigma(h^{-1})\big)(g(s)) \\ &= \alpha(s, x_0)(g(s)). \end{aligned}$$

Therefore, we have a well-defined function $f\colon X \to E$ given by

$$f(x) = f(s \cdot x_0) = \alpha(s, x_0)(g(s)), \tag{$*_f$}$$

and by definition of f^α, we have

$$\begin{aligned} f^\alpha(s) &= (\alpha(s, x_0))^{-1}(f(s \cdot x_0)) \\ &= (\alpha(s, x_0))^{-1}\big(\alpha(s, x_0)(g(s))\big) \\ &= g(s), \end{aligned}$$

that is, $f^\alpha = g$, which shows that τ is surjective.

Since $\alpha(s, x)$ is an automorphism and since the map $s \mapsto s \cdot x_0$ from G to G/H is surjective, for any two functions $f_1, f_2 \in E^X$, if $f_1^\alpha = f_2^\alpha$, then

$$\alpha(s^{-1}, s \cdot x_0)(f_1(s \cdot x_0)) = \alpha(s^{-1}, s \cdot x_0)(f_2(s \cdot x_0))$$

for all $s \in G$, so $f_1 = f_2$, which shows that τ is injective. $\square$

Observe that in the proof of Proposition 15.6, Equation $(*_f)$ and the fact that $\tau(f) = f^\alpha = g$ show that if $g \in L^\alpha$, then

$$(\tau^{-1}(g))(s \cdot x_0) = \alpha(s, x_0)(g(s)). \tag{$*_{\tau^{-1}(g)}$}$$

For any cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$, we can use the isomorphism $\tau\colon E^X \to L^\alpha$ to convert the representation $\Pi\colon G \to \mathbf{GL}(E^X)$ defined by α into the equivalent representation Π_{L^α} given by $\Pi_{L^\alpha}(s) = \tau \circ \Pi(s) \circ \tau^{-1}$.

Proposition 15.7. *For every cocycle $\alpha\colon G \times X \to \mathbf{GL}(E)$,if $\Pi\colon G \to \mathbf{GL}(E^X)$ is the representation defined by α, then the equivalent representation $\Pi_{L^\alpha}\colon G \to \mathbf{GL}(L^\alpha)$ defined by $\Pi_{L^\alpha}(s) = \tau \circ \Pi(s) \circ \tau^{-1}$ is given by*

$$((\Pi_{L^\alpha})_s(g))(t) = g(s^{-1}t) \qquad \text{for all } g \in L^\alpha \text{ and all } s, t \in G. \tag{Π_{L^α}}$$

Proof. For any $g \in L^\alpha$, since by $(*_{\tau^{-1}(g)})$
$$(\tau^{-1}(g))(u \cdot x_0) = \alpha(u, x_0)(g(u)),$$
and
$$(\Pi_s(f))(x) = \alpha(s, s^{-1} \cdot x_0)(f(s^{-1} \cdot x)),$$
with $x = t \cdot x_0$, we have
$$(\Pi_s(f))(t \cdot x_0) = \alpha(s, s^{-1} \cdot x_0)(f(s^{-1} \cdot (t \cdot x_0))),$$
and by setting $f = \tau^{-1}(g)$, we get
$$\begin{aligned}\big(\Pi_s(\tau^{-1}(g))\big)(t \cdot x_0) &= \alpha(s, s^{-1} \cdot (t \cdot x_0))\big((\tau^{-1}(g))(s^{-1} \cdot (t \cdot x_0))\big)\\ &= \alpha(s, (s^{-1}t) \cdot x_0)\big((\tau^{-1}(g))((s^{-1}t) \cdot x_0\big)\\ &= \alpha(s, (s^{-1}t) \cdot x_0)\big(\alpha(s^{-1}t, x_0)(g(s^{-1}t))\big)\\ &= \alpha(ss^{-1}t, x_0)(g(s^{-1}t))\\ &= \alpha(t, x_0)(g(s^{-1}t)).\end{aligned}$$
Since
$$(\tau(h))(t) = (\alpha(t, x_0))^{-1}(h(t \cdot x_0)),$$
for any $h \in E^X$, with $h = \Pi_s(\tau^{-1}(g))$, we obtain
$$\tau\big(\Pi_s(\tau^{-1}(g))\big)(t) = (\alpha(t, x_0))^{-1}\big(\alpha(t, x_0)(g(s^{-1}t))\big) = g(s^{-1}t),$$
as claimed. $\square$

Remark: Observe that L^α only depends on σ, so we may write L^σ instead of L^α, and Π_{L^α} depends only on σ, so we may also write Π_{L^σ} instead of Π_{L^α}.

We have concluded our discussion of algebraic methods for constructing representations of G from representations of a subgroup H of G.

15.4 Construction of the Hilbert Space $\mathrm{L}^2_\mu(X; E)$

We now assume that G is a locally compact group and that H is a closed subgroup of G. By Proposition 8.6(1), the space $X = G/H$ is also locally compact. If G is separable, then so is G/H, and if G is metrizable, then so G/H; see Dieudonné [24] (Chapter XII, Sections 10 and 11).

Given a unitary representation $U \colon H \to \mathbf{U}(E)$ of H we would like to construct a unitary representation $\Pi \colon G \to \mathbf{U}(\mathcal{H})$ of G. This is possible under certain conditions on H and G and on measures on $X = G/H$. Note that unlike in the previous sections we are now considering *continuous* unitary representations.

The first step is to construct a Hilbert space $\mathcal{H}$ that will be the representation space of a unitary representation of G. There are two approaches:

1. The Hilbert space $\mathcal{H}$ is a set of functions from $X = G/H$ to E.

2. The Hilbert space $\mathcal{H}$ is a set of functions from G to E, analogous to the space L^α of Section 15.3.

The second step is to define the operators Π_s (for $s \in G$) so that they are unitary operators of $\mathcal{H}$. This involves defining an inner product in $\mathcal{H}$ that makes the operators Π_s unitary. In the first approach that makes use of cocycles, the definition of the inner product on $\mathcal{H}$ is straightforward. To ensure that the operators Π_s are unitary, a Borel measure μ on $X = G/H$ is needed, and the cocycles must satisfy some additional conditions with respect to the measure μ. The case where the measure μ is G-invariant is simpler than the case where μ is only quasi-invariant.

In the second approach, the definition of the Hilbert space $\mathcal{H}$ is more complicated and requires a completion. We will sketch two variants of this method at the end of Section 15.7.

A good candidate for the first approach is a subspace $\mathrm{L}^2_\mu(X;E)$ of the vector space E^X, where μ is positive Borel measure on G/H. In the special case where H is compact, given a cocycle α on $G \times X$ satisfying some suitable conditions, the space L^α will be a subspace of $\mathrm{L}^2_\lambda(G;E) \subseteq E^G$, where λ is a left-invariant Haar measure on G.

Whether μ is G-invariant is an issue that will come up later, but for the time being we can ignore it.

Let E be a separable Hilbert space, and let (a_n) be a Hilbert basis of E. Every function $f\colon X \to E$ can be written uniquely as $f = \sum_n f_n a_n$, where $f_n\colon X \to \mathbb{C}$, and such that the series $\sum_n |f_n(x)|^2$ converges for all $x \in X$. By definition, we let

$$\|f(x)\|_E^2 = \sum_n |f_n(x)|^2.$$

We claim that a function $f\colon X \to E$ is μ-measurable iff all the f_n are μ-measurable.

If f is μ-measurable, since $f_n(x) = \langle f(x), a_n\rangle$, the f_n are μ-measurable. Conversely, if the f_n are μ-measurable, then Egoroff's theorem implies that f is μ-measurable; see Dieudonné [24] (Chapter XIII, Theorem 13.9.10).

Definition 15.9. Let G be a locally compact group, let H be a closed subgroup of G, let μ be a positive Borel measure on $X = G/H$, and let E be a separable Hilbert space. For any Hilbert basis (a_n) of E, let $\mathcal{L}^2_\mu(X;E)$ be the space of all μ-measurable functions $f\colon X \to E$ with $f = \sum_n f_n a_n$, such that the function $x \mapsto \sum_n |f_n(x)|^2 = \|f(x)\|_E^2$ is μ-integrable.

It is easy to see that if $f = \sum_n f_n a_n$, then $f_n \in \mathcal{L}^2_\mu(X;\mathbb{C})$, and

$$\int_{G/H} \|f\|_E^2 \, d\mu = \sum_n \int_{G/H} |f_n|^2 \, d\mu = \sum_n \|f_n\|_2^2;$$

see Dieudonné [24] (Chapter XIII, Sections 8 and 9). As a consequence, given two functions $f = \sum_n f_n a_n$ and $g = \sum_n g_n a_n$ in $\mathcal{L}^2_\mu(X; E)$, by Proposition 5.41, the function $x \mapsto \langle f(x), g(x)\rangle$ is integrable and

$$\int_{G/H} \langle f(x), g(x)\rangle \, d\mu(x) = \sum_n \int_{G/H} f_n(x)\overline{g_n(x)}\, d\mu(x).$$

Definition 15.10. We say that a function $f \in \mathcal{L}^2_\mu(X; E)$ is *negligeable* if the function $x \mapsto \|f(x)\|_E^2$ is zero almost everywhere.

The quotient of the space $\mathcal{L}^2_\mu(X; E)$ by the subspace of negligeable functions is denoted by $\mathrm{L}^2_\mu(X; E)$. It is a hermitian space under the inner product

$$\langle [f], [g]\rangle = \int_{G/H} \langle f(x), g(x)\rangle \, d\mu(x),$$

and we have the norm N_2[1] given by

$$N_2([f]) = \sqrt{\langle [f], [f]\rangle}.$$

If $[f]$ is represented by $f = \sum_n f_n a_n$, then

$$N_2([f])^2 = \int_{G/H} \|f\|_E^2 \; d\mu = \sum_n \|f_n\|_2^2 \, .$$

Actually, it turns out that the hermitian space $\mathrm{L}^2_\mu(X; E)$ is complete, that is, it is a Hilbert space. In fact, it is a separable Hilbert space.

Proposition 15.8. *Let G be a locally compact group, let H be a closed subgroup of G, let μ be a positive Borel measure on $X = G/H$, and let E be a separable Hilbert space. The space $\mathrm{L}^2_\mu(X; E)$ is a separable Hilbert space.*

Proof. Let $(f^{(m)})$ be a Cauchy sequence in $\mathrm{L}^2_\mu(X; E)$, with $f^{(m)} = \sum_n f_n^{(m)} a_n$. For every $\epsilon > 0$, there is some m_0 such that for all $p, q \geq m_0$, we have

$$N_2(f^{(p)} - f^{(q)})^2 = \sum_n \int_{G/H} |f_n^{(p)} - f_n^{(q)}|^2 \, d\mu \leq \epsilon, \tag{$*_1$}$$

and this implies that for every n, the sequence $(f_n^{(m)})_{m\geq 1}$ is a Cauchy sequence in $\mathrm{L}^2_\mu(X; \mathbb{C})$. Therefore, each sequence $(f_n^{(m)})_{m\geq 1}$ has a limit $g_n \in \mathrm{L}^2_\mu(X; \mathbb{C})$, since $\mathrm{L}^2_\mu(X; \mathbb{C})$ is complete by Fischer–Riesz. For every integer $N > 0$, if we let q tend to $+\infty$ in $(*_1)$, we see that

$$\sum_{n=1}^{N} \left\| g_n - f_n^{(p)} \right\|_2^2 \leq \epsilon, \tag{$*_2$}$$

[1] We are using the notation N_2 for the norm on $\mathrm{L}^2_\mu(X; E)$ to avoid a confusion with the norm $\| \, \|_2$ on $\mathrm{L}^2_\mu(X; \mathbb{C})$.

so

$$\sum_{n=1}^{N} \|g_n\|_2^2 \leq \sum_{n=1}^{N} \left\|g_n - f_n^{(p)}\right\|_2^2 + \sum_{n=1}^{N} \left\|f_n^{(p)}\right\|_2^2 \leq \epsilon + \left\|f^{(p)}\right\|_2^2,$$

which proves that the series $\sum_{n=1}^{\infty} \|g_n\|_2^2$ converges. Since (by definition)

$$\sum_{n=1}^{\infty} \|g_n\|_2^2 = N_2(g)^2,$$

it follows that $g = \sum_n g_n a_n \in \mathrm{L}^2_\mu(X; E)$, and by $(*_2)$

$$N_2(g - f^{(p)})^2 = \sum_n \left\|g_n - f_n^{(p)}\right\|_2^2 \leq \epsilon$$

for all $p \geq m_0$, and so g is the limit of the sequence $(f^{(m)})$ in $\mathrm{L}^2_\mu(X; E)$.

If D is a countable dense subset of $\mathrm{L}^2_\mu(X; \mathbb{C})$, then we can check that the set of functions $f = \sum f_n a_n$ such that $f_n \in D$ for all n and and $f_n = 0$ but all for finitely many values of n is dense in $\mathrm{L}^2_\mu(X; E)$. □

15.5 Induced Representations, I; G/H has a G-Invariant Measure

In the rest of this chapter, by unitary representation, we mean *continuous* unitary representation.

We will now assume that the positive Borel measure μ on $X = G/H$ is G-invariant. Recall from Section 8.10 (Definition 8.18) that

$$(\lambda_s(\mu))(A) = \mu(s^{-1} \cdot A),$$

for every Borel subset A of X, so μ is G-invariant if for every Borel subset A of X,

$$\mu(s^{-1} \cdot A) = \mu(A) \qquad \text{for all } s \in G.$$

In this case,

$$\int_{G/H} f(s \cdot x)\, d\mu(x) = \int_{G/H} f(x)\, d\mu(x), \qquad \text{for all } s \in G.$$

Let E be a separable Hilbert space, and let $U\colon H \to \mathbf{U}(E)$ be a unitary representation of H.

Theorem 15.9. *Let G be a locally compact group, H be a closed subgroup of G, E be a separable Hilbert space, and $U\colon H \to \mathbf{U}(E)$ be a unitary representation of H. If $X = G/H$ admits a G-invariant σ-Radon measure μ, and for any cocycle $\alpha\colon G \times X \to \mathbf{U}(E)$, if the following conditions hold*

(1) We have $\alpha(h, x_0) = U(h)$ for all $h \in H$;

(2) For every $s \in G$, for every $f \in \mathrm{L}^2_\mu(X; E)$, the map $x \mapsto \alpha(s, x)(f(x))$ from X to E is μ-measurable;

(3) For every $f \in \mathrm{L}^2_\mu(X; E)$, the map $s \mapsto \Pi_s(f)$ is a continuous map from G to $\mathrm{L}^2_\mu(X; E)$, where Π is the homomorphism $\Pi \colon G \to \mathbf{GL}(E^X)$ induced by the cocycle α;

then the homomorphism $\Pi \colon G \to \mathbf{U}(\mathrm{L}^2_\mu(X; E))$ induced by the cocycle α given by

$$(\Pi_s(f))(x) = (\alpha(s^{-1}, x))^{-1}(f(s^{-1} \cdot x)), \qquad f \in \mathrm{L}^2_\mu(X; E),\, x \in X,$$

(see Definition 15.3) is a unitary representation of G.

Proof. We simply have to prove that

$$N_2(\Pi_s(f)) = N_2(f), \qquad \text{for all } f \in \mathcal{L}^2_\mu(X; E) \text{ and all } s \in G,$$

which implies that $\Pi_s(f) \in \mathcal{L}^2_\mu(X; E)$, and the other conditions imply that the homomorphism $\Pi \colon G \to \mathbf{GL}(\mathrm{L}^2_\mu(X; E))$ induced by α is a unitary representation of G. Since by hypothesis $\alpha(s, s^{-1} \cdot x)$ is a unitary operator, we have

$$\|(\Pi_s(f))(x)\|_E = \left\|\alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x))\right\|_E = \left\|f(s^{-1} \cdot x)\right\|_E,$$

and since μ is G-invariant,

$$\int_{G/H} \left\|f(s^{-1} \cdot x)\right\|^2_E d\mu(x) = \int_{G/H} \|f(x)\|^2_E \, d\mu(x),$$

and so $N_2(\Pi_s(f)) = N_2(f)$. ☐

Definition 15.11. The unitary representation $\Pi \colon G \to \mathbf{U}(\mathrm{L}^2_\mu(X; E))$ induced by the cocycle α (and the unitary representation $U \colon H \to \mathbf{U}(E)$) is denoted $\mathrm{Ind}_H^G \alpha$, or by abuse of notation even $\mathrm{Ind}_H^G U$.

Remark: To be very precise, the representing space $\mathrm{L}^2_\mu(X; E)$ of this representation should be specified, for example as in $\mathrm{Ind}^G_{H, \mathrm{L}^2_\mu(X;E)} \alpha$, because there are variants of this construction that use a different representation space.

If U is the trivial representation of H in E, and if we choose $\alpha(s, x) = \mathrm{id}_E$ for all $(s, x) \in G \times (G/H)$, then it can be verified that the hypotheses of Theorem 15.9 are satisfied. To verify Condition (3), we use the fact that the family of maps $f \mapsto \Pi_s(f)$ ($s \in G$) is equicontinuous; see Proposition 2.13. Then we use Proposition 2.12; for details, see Dieudonné [22], (Chapter XXII, Section 3). In this case, the subspace L^α corresponding to $\mathcal{L}^2_\mu(X; E)$ consists of all functions of the form $f \circ \pi$ with $f \in \mathcal{L}^2_\mu(X; E)$, where $\pi \colon G \to G/H$ is the projection map.

If H is a (closed) compact subgroup of G, then by Proposition 8.43, the space G/H has G-invariant measures (unique up to a scalar). This is a special case of particular interest. A good illustration of this situation is provided by Example 15.1 that we now revisit.

Example 15.2. As in Example 15.1 consider the groups $G = \mathbf{SE}(3)$ and $H \approx \mathbf{SO}(3)$, where G is locally compact and H is compact and closed in G. Consequently $X = G/H \approx \mathbb{R}^3$ has an $\mathbf{SE}(3)$-invariant Radon measure μ. Consider any unitary representation $\sigma\colon \mathbf{SO}(3) \to \mathbf{U}(E)$ of $\mathbf{SO}(3)$ in a separable Hilbert space E. We showed in Example 15.1 that we have a cocycle $\alpha\colon \mathbf{SE}(3) \times \mathbb{R}^3 \to \mathbf{U}(E)$ given by

$$\alpha((a, Q), x) = \sigma(Q), \quad a, x \in \mathbb{R}^3,\ Q \in \mathbf{SO}(3),$$

and the homomorphism $\Pi\colon \mathbf{SE}(3) \to \mathbf{GL}(E^{\mathbb{R}^3})$ induced by α is given by

$$(\Pi_{(a,Q)}(f))(x) = \sigma(Q)f(Q^\top(x - a)), \quad f \in E^{\mathbb{R}^3},\ x \in \mathbb{R}^3.$$

We leave it as an exercise to check that Conditions (1)-(3) of Theorem 15.9 are satisfied, and so Π is a unitary representation $\Pi\colon \mathbf{SE}(3) \to \mathbf{U}(\mathbf{L}^2_\mu(\mathbb{R}^3; E))$ of $\mathbf{SE}(3)$ in the Hilbert space $\mathbf{L}^2_\mu(\mathbb{R}^3; E)$. If E is finite-dimensional, say of dimension $n \geq 1$, then the Hilbert space $\mathbf{L}^2_\mu(\mathbb{R}^3; E)$ is isomorphic to the direct sum of n copies of $\mathbf{L}^2_\mu(\mathbb{R}^3; \mathbb{C})$. Then every function $f \in \mathbf{L}^2_\mu(\mathbb{R}^3; E)$ is identified with the n-tuple $f = (f_1, \ldots, f_n)$ where $f_i \in \mathbf{L}^2_\mu(\mathbb{R}^3; \mathbb{C})$, with the inner product of $f = (f_1, \ldots, f_n)$ and $g = (g_1, \ldots, g_n)$ given by

$$\langle f, g\rangle = \sum_{i=1}^n \int_{\mathbb{R}^3} f_n(x)\overline{g(x)}\, d\mu(x).$$

Another example of induced representations of $G = \mathbf{SE}(n)$ arises from the normal abelian subgroup $H = \mathbb{R}^n$.

Example 15.3. Consider the groups $G = \mathbf{SE}(n)$ and $H \approx \mathbb{R}^n$, where G is locally compact and H is a closed normal abelian group in G. Here $G = \mathbf{SE}(n)$ consists of all matrices

$$s = \begin{pmatrix} Q & a \\ 0 & 1 \end{pmatrix} \quad \text{with } Q \in \mathbf{SO}(n) \text{ and } a \in \mathbb{R}^n,$$

$H \approx \mathbb{R}^n$ is the normal subgroup of $\mathbf{SE}(n)$ consisting of all matrices

$$h = \begin{pmatrix} I_n & b \\ 0 & 1 \end{pmatrix} \quad \text{with } b \in \mathbb{R}^n,$$

and $X = G/H \approx \mathbf{SO}(n)$ is the compact subgroup of $\mathbf{SE}(n)$ consisting of all matrices

$$\begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix} \quad \text{with } R \in \mathbf{SO}(n).$$

Recall that since $\mathbb{R}^n$ is abelian, its irreducible representations are one-dimensional. Therefore the irreducible representations of $\mathbb{R}^n$ are determined by the characters of $\mathbb{R}^n$, which by Corollary 10.11 are of the form $\chi_y\colon \mathbb{R}^n \to \mathbb{T}$ for any $y \in \mathbb{R}^n$, with

$$\chi_y(x) = e^{iy\cdot x}, \quad x \in \mathbb{R}^n.$$

Consequently the irreducible representations $\rho\colon \mathbb{R}^n \to \mathbf{U}(1)$ of $\mathbb{R}^n$ are of the form

$$(\rho(x))(z) = \chi_y(x)z, \quad x \in \mathbb{R}^n,\ z \in \mathbb{C}$$

for any fixed $y \in \mathbb{R}^n$, namely, multiplication by $\chi_y(x)$. Since for

$$s = (a, Q) = \begin{pmatrix} Q & a \\ 0 & 1 \end{pmatrix} \in \mathbf{SE}(n) \quad \text{and} \quad h = (b, I) = \begin{pmatrix} I & b \\ 0 & 1 \end{pmatrix} \in H \approx \mathbb{R}^n$$

we have

$$sH = (a, Q)H = \{(a, Q)h \mid h \in H\} = \left\{ \begin{pmatrix} Q & a + Qb \\ 0 & 1 \end{pmatrix} \mid b \in \mathbb{R}^n \right\} = \left\{ \begin{pmatrix} Q & c \\ 0 & 1 \end{pmatrix} \mid c \in \mathbb{R}^n \right\},$$

we have an isomorphism between $\mathbf{SO}(n)$ and $X = \mathbf{SE}(n)/H$ given by

$$Q \mapsto (a, Q)H = \left\{ \begin{pmatrix} Q & c \\ 0 & 1 \end{pmatrix} \mid c \in \mathbb{R}^n \right\}.$$

Since each matrix in the coset $(a, Q)H$ can be written uniquely as

$$\begin{pmatrix} Q & c \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & Q^\top c \\ 0 & 1 \end{pmatrix}$$

it is very easy to pick a coset representative in $\mathbf{SE}(n)$, namely

$$r_Q = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}, \quad Q \in \mathbf{SO}(n).$$

The coset H as point in $X = \mathbf{SE}(n)/H \approx \mathbf{SO}(n)$ is $x_0 = I_n$. Since the action of $\mathbf{SE}(n)$ on $X = \mathbf{SE}(n)/H \approx \mathbf{SO}(n)$ is given by

$$s_1(sH) = (s_1 s)H, \quad s_1, s \in \mathbf{SE}(n),$$

we have

$$s_1(sH) = (s_1 s)H = (a_1, Q_1)(a, Q)H = (a_1 + Q_1 a, Q_1 Q)H = \left\{ \begin{pmatrix} Q_1 Q & c \\ 0 & 1 \end{pmatrix} \mid c \in \mathbb{R}^n \right\},$$

and using our isomorphism between $\mathbf{SO}(n)$ and $X = \mathbf{SE}(n)/H$, the above equation becomes

$$s_1 \cdot Q = (a_1, Q_1) \cdot Q = Q_1 Q, \quad Q, Q_1 \in \mathbf{SO}(n),\ a_1 \in \mathbb{R}^n.$$

Then since

$$s \cdot R = (a, Q) \cdot R = QR,$$

$u(s,R) = (r_{s\cdot R})^{-1} s r_R$ is given by

$$\begin{aligned} u(s,R) = (r_{s\cdot R})^{-1} s r_R &= \begin{pmatrix} R^\top Q^\top & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} R^\top & R^\top Q^\top a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} I_n & R^\top Q^\top a \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Technically we prefer dealing with representations $\sigma\colon \mathbb{R}^n \to \mathbf{U}(1)$ rather than $\sigma\colon H \to \mathbf{U}(1)$, so using the isomorphism $\mathbb{R}^n \approx H$ we have

$$u((a,Q),R) = R^\top Q^\top a.$$

Consequently, for every irreducible representation $\sigma = \chi_y\colon \mathbb{R}^n \to \mathbf{U}(1)$, we have the cocycle $\alpha\colon \mathbf{SE}(n) \times \mathbf{SO}(n) \to \mathbf{U}(1)$ given by

$$\alpha((a,Q),R) = \sigma(u((a,Q),R)) = \sigma(R^\top Q^\top a) = \chi_y(R^\top Q^\top a).$$

Observe that if $(a,Q) \in H \approx \mathbb{R}^n$, that is, $Q = I$, and $R = x_0 = I$, we have $\alpha((a,I),I) = \sigma(a) = \chi_y(a)$. Since

$$s^{-1} \cdot R = (a,Q)^{-1} \cdot R = Q^\top R,$$

the representation $\Pi\colon \mathbf{SE}(n) \to \mathbf{GL}(\mathbb{C}^{\mathbf{SO}(n)})$ of $\mathbf{SE}(n)$ in $\mathbb{C}^{\mathbf{SO}(n)}$ induced by the representation $\sigma = \chi_y\colon \mathbb{R}^n \to \mathbf{U}(1)$ is defined such that for all $s = (a,Q) \in \mathbf{SE}(n)$, $R \in \mathbf{SO}(n)$ and all functions $f\colon \mathbf{SO}(n) \to \mathbb{C}$,

$$\begin{aligned} (\Pi_{(a,Q)}(f))(R) &= \alpha(s, s^{-1}\cdot R)(f(s^{-1}\cdot R)) \\ &= \alpha((a,Q), Q^\top R)(f((a,Q)^{-1}\cdot R)) \\ &= \sigma((Q^\top R)^\top Q^\top a)(f((a,Q)^{-1}\cdot R)) \\ &= \sigma(R^\top a) f((a,Q)^{-1}\cdot R) \\ &= \chi_y(R^\top a) f((a,Q)^{-1}\cdot R) \\ &= e^{i(y\cdot(R^\top a))} f((a,Q)^{-1}\cdot R) \\ &= e^{i((Ry)\cdot a)} f(Q^\top R). \end{aligned}$$

Since the action of $\mathbf{SE}(n)$ on $\mathbf{SO}(n)$ is identical to the action of $\mathbf{SO}(n)$ on $\mathbf{SO}(n)$, the homogeneous space $X = \mathbf{SO}(n)$ has an $\mathbf{SE}(n)$-invariant Radon measure, namely the Haar measure μ on $\mathbf{SO}(n)$. We already checked that the cocycle

$$\alpha((a,Q),R) = \chi_y(R^\top Q^\top a) = e^{i(y\cdot(R^\top Q^\top a))} = e^{i((Ry)\cdot(Q^\top a))}$$

satisfies Condition (1) of Theorem 15.9, and we leave it as an exercise to prove that Conditions (2) and (3) are also satisfied. As a consequence, we obtain a unitary representation $\Pi\colon \mathbf{SE}(n) \to \mathbf{U}(\mathrm{L}^2_\mu(\mathbf{SO}(n);\mathbb{C}))$ of $\mathbf{SE}(n)$ in the Hilbert space $\mathrm{L}^2_\mu(\mathbf{SO}(n);\mathbb{C})$ given by

$$\begin{aligned} (\Pi_{(a,Q)}(f))(R) = e^{i((Ry)\cdot a)} f(Q^\top R), \quad & (a,Q) \in \mathbf{SE}(n),\ R \in \mathbf{SO}(n), \\ & f \in \mathrm{L}^2_\mu(\mathbf{SO}(n);\mathbb{C}),\ y \in \mathbb{R}^n. \end{aligned}$$

The above formula suggests that it might be possible to define a representation of $\mathbf{SE}(n)$ in the smaller Hilbert space $\mathrm{L}^2_\lambda(S^{n-1};\mathbb{C})$, where λ is an $\mathbf{SO}(n)$-invariant Radon measure on S^{n-1}, which exists since S^{n-1} is a homogeneous space obtained by making $\mathbf{SO}(n)$ act on S^{n-1} by the action $R\cdot x = Rx$ where $R \in \mathbf{SO}(n)$ and $x \in S^{n-1}$, so $S^{n-1} \approx \mathbf{SO}(n)/\mathbf{SO}(n-1)$ with $\mathbf{SO}(n-1)$ compact. Before proceeding any further, the reader may want to review Section C.2 and Section C.3. We may assume that $y \neq 0$, because when $y = 0$ we have

$$(\Pi_{(a,Q)}(f))(R) = f(Q^\top R),$$

a reducible representation called a *quasi-regular representation* of $\mathbf{SE}(n)$. Here we pick the base point to be

$$x_0 = (1/r)y \in S^{n-1}, \text{ with } r = \|y\|.$$

The stabilizer $\mathbf{SO}(n)_{x_0} \approx \mathbf{SO}(n-1)$ of x_0 is given by

$$\mathbf{SO}(n)_{x_0} = \{R \in \mathbf{SO}(n) \mid Rx_0 = x_0\},$$

and so, for any $R_1, R_2 \in \mathbf{SO}(n)$, the two cosets $R_1\mathbf{SO}(n)_{x_0}$ and $R_2\mathbf{SO}(n)_{x_0}$ are identical iff $R_2^\top R_1 \in \mathbf{SO}(n)_{x_0}$ iff $R_2^\top R_1 x_0 = x_0$ iff $R_1 x_0 = R_2 x_0$. The isomorphism between $\mathbf{SO}(n)/\mathbf{SO}(n)_{x_0}$ and the orbit $\mathbf{SO}(n)x_0 = S^{n-1}$ is given by $R\,\mathbf{SO}(n)_{x_0} \mapsto Rx_0 = (1/r)(Ry)$, where $R \in \mathbf{SO}(n)$. Consider the map $\widetilde{\Pi}\colon \mathbf{SE}(n) \to \mathbf{U}(\mathrm{L}^2_\lambda(S^{n-1};\mathbb{C}))$ given by

$$(\widetilde{\Pi}_{(a,Q)}(f))([R]) = e^{i((Ry)\cdot a)} f(Q^\top [R]), \quad (a,Q) \in \mathbf{SE}(n),\ y \in \mathbb{R}^n \qquad (*_1)$$

with $[R] \in \mathbf{SO}(n)/\mathbf{SO}_{x_0}$ and $f \in \mathrm{L}^2_\lambda(S^{n-1};\mathbb{C})$, and where $[R]$ denotes the coset $R\,\mathbf{SO}_{x_0}$. Since by definition of the stabilizer $\mathbf{SO}_{x_0}$, if $[R_1] = [R_2]$, then $R_1 y = R_2 y$, the right-hand side of $(*_1)$ does not depend on the representative chosen in the coset $[R]$, so $\widetilde{\Pi}_{(a,Q)}$ is well-defined, and if we write $x = (1/r)(Ry) \in S^{n-1}$, since $\mathbf{SO}(n)/\mathbf{SO}_{x_0} \approx S^{n-1}$ under the map $[R] \mapsto Rx_0 = (1/r)Ry = x$, we have

$$(\widetilde{\Pi}_{(a,Q)}(f))(x) = e^{ir(x\cdot a)} f(Q^\top x), \quad (a,Q) \in \mathbf{SE}(n),\ x \in S^{n-1},\ f \in \mathrm{L}^2_\lambda(S^{n-1};\mathbb{C}),\ r > 0. \quad (*_2)$$

The above also shows that the representation $\Pi\colon \mathbf{SE}(n) \to \mathbf{U}(\mathrm{L}^2_\mu(\mathbf{SO}(n);\mathbb{C}))$ of $\mathbf{SE}(n)$ in the Hilbert space $\mathrm{L}^2_\mu(\mathbf{SO}(n);\mathbb{C})$ is *reducible* because the subspace of $\mathrm{L}^2_\mu(\mathbf{SO}(n);\mathbb{C})$ consisting of the functions $f \in \mathrm{L}^2_\mu(\mathbf{SO}(n);\mathbb{C})$ such that

$$f(RT) = f(R) \quad \text{for all } R \in \mathbf{SO}(n) \text{ and all } T \in \mathbf{SO}(n)_{x_0}$$

is invariant under $\Pi_{(a,Q)}$, because for all $Q, R \in \mathbf{SO}(n)$ and all $T \in \mathbf{SO}(n)_{x_0}$ we have,

$$e^{i((RTy)\cdot a)} f(Q^\top RT) = e^{ir((RTx_0)\cdot a)} f(Q^\top R) = e^{ir((Rx_0)\cdot a)} f(Q^\top R) = e^{i((Ry)\cdot a)} f(Q^\top R),$$

since $Tx_0 = x_0$ and $f(Q^\top RT) = f(Q^\top R)$.

The representations given by $(*_2)$ are half of the representations of $\mathbf{SE}(n)$ discussed in Vilenkin [101] (Chapter XI, Section 2), the other half corresponding to $r < 0$. However, it is easy to see that each representation given by

$$(\widetilde{\Pi}_{(a,Q)}(f))(x) = e^{-ir(x\cdot a)} f(Q^\top x), \quad (a,Q) \in \mathbf{SE}(n),\ x \in S^{n-1},\ f \in \mathrm{L}^2_\lambda(S^{n-1};\mathbb{C}),\ r > 0 \quad (*_3)$$

is equivalent to the corresponding representation given by $(*_2)$ (with no negative sign in front of $ir > 0$) using the isometry T of $\mathrm{L}^2_\lambda(S^{n-1};\mathbb{C})$ given by

$$T(f)(x) = f(-x), \quad x \in S^{n-1},$$

in other words, $T(f) = \check{f}$ (see Definition 8.11). It is proven in Vilenkin [101] (Chapter XI, Section 2) that the representations given by $(*_2)$ (and thus by $(*_3)$) are irreducible.

Actually, if we allow ir to be *any* nonzero complex number $z = ir$, then Vilenkin proves that

$$(\widetilde{\Pi}_{(a,Q)}(f))(x) = e^{z(x\cdot a)} f(Q^\top x), \quad (a, Q) \in \mathbf{SE}(n),\ x \in S^{n-1},\ f \in \mathrm{L}^2_\lambda(S^{n-1};\mathbb{C}),\ z \in \mathbb{C}^* \quad (*_4)$$

still defines an irreducible representation, but it is not unitary unless $z = ir$ with $r \in \mathbb{R}$ and $r \neq 0$.

The representations of $\mathbf{SE}(n)$ given by $(*_2)$ have the following interesting property. If we consider their restriction to $\mathbf{SO}(n)$, so that $s = (0, Q)$, then we see that they are given by

$$(\widetilde{\Pi}_{(0,Q)}(f))(x) = f(Q^\top x), \quad Q \in \mathbf{SO}(n),\ x \in S^{n-1},\ f \in \mathrm{L}^2_\lambda(S^{n-1};\mathbb{C}). \qquad (*_5)$$

The constant function $f_0\colon S^{n-1} \to \mathbb{C}$ with value 1 is invariant under $\mathbf{SO}(n)$, in the sense that

$$(\widetilde{\Pi}_{(0,Q)}(f_0))(x) = f_0(Q^\top x) = 1 \quad \text{for all } Q \in \mathbf{SO}(n) \text{ and all } x \in S^{n-1},$$

which means that

$$\widetilde{\Pi}_{(0,Q)}(f_0) = f_0 \quad \text{for all } Q \in \mathbf{SO}(n).$$

This is an instance of what is called a representation of class 1 relative to $\mathbf{SO}(n)$.

Definition 15.12. Let G be a locally compact group and let H be a closed subgroup of G. A unitary representation $U\colon G \to \mathbf{U}(E)$ of G in a Hilbert space E is a *representation of class* 1 *relative to* H if there is some nonzero vector $x \in E$ invariant relative to H, which means that

$$U_h(x) = x \quad \text{for all } h \in H.$$

Remark: Vilenkin [101] (Chapter I, Section 2) allows U to be nonunitary, but in this case the restriction of U to H must be unitary.

The representations of Example 15.3 given by $(*_2)$ are of class 1 relative to $\mathbf{SO}(n)$.

One of the reasons why representations of class 1 are interesting is the following. Suppose $a \in E$ is a nonzero vector invariant under H as above. For every $x \in E$ we define the function $f_x\colon G \to \mathbb{C}$ given by

$$f_x(s) = \langle U_s(x), a\rangle, \quad s \in G.$$

The functions f_x are called *spherical functions of* U *relative to* H. We claim that the functions f_x are constant on right cosets Hs.

Indeed, for all $s \in G$ and all $h \in H$ we have

$$\begin{aligned} f_x(hs) &= \langle U_{hs}(x), a\rangle \\ &= \langle (U_h(U_s(x)), a\rangle \\ &= \langle U_s(x), U_h^*(a)\rangle \\ &= \langle U_s(x), U_{h^{-1}}(a)\rangle \\ &= \langle U_s(x), a\rangle = f_x(s), \end{aligned}$$

so

$$f_x(hs) = f(s) \quad \text{for all } s \in G \text{ and all } h \in H.$$

In particular, for $x = a$, we claim that the function f_a, called a *zonal spherical function*, is constant on the two-sided cosets HsH $(s \in G)$.

Since we already know that $f_a(h_1 s) = f_a(s)$ for all $h_1 \in H$, it suffices to show that $f_a(sh_2) = f_a(s)$ for all $h_2 \in H$. We have

$$\begin{aligned} f_a(sh_2) &= \langle U_{sh_2}(a), a\rangle \\ &= \langle U_s(U_{h_2}(a)), a\rangle \\ &= \langle U_s(a), a\rangle = f_a(s). \end{aligned}$$

Thus we proved that

$$f_a(h_1 s h_2) = f_a(s) \quad \text{for all } h_1, h_2 \in H \text{ and all } s \in G,$$

which means that f_a is constant on the double cosets HsH. Geometrically, this means that f_a is constant on "spheres." In particular, if $G = \mathbf{SO}(3)$ and $H = \mathbf{SO}(2)$, then the spherical functions are the well-known spherical harmonics $Y_l^m(\theta, \varphi)$ and the zonal spherical functions are the Legendre polynomials $P_l(\cos\theta)$. If $G = \mathbf{SO}(n)$ and $H = \mathbf{SO}(n-1)$, then the zonal spherical functions are given in terms of Gegenbaur polynomials; see Gallier and Quaintance [39] (Chapter 7, Sections 3, 5, 6, 7).

Under some mild additional conditions, induced unitary representations of G in $\mathrm{L}^2_\mu(X; E)$ can be converted to unitary representations of G in a closed subspace of $\mathrm{L}^2_\lambda(G; E)$ (where λ is a left Haar measure on G).

Suppose that the unitary cocycle α has the property that the map

$$s \mapsto f^\alpha(s) = \alpha(s^{-1}, s \cdot x_0)(f(s \cdot x_0))$$

from G to E is λ-measurable for every $f \in \mathcal{L}^2_\mu(X; E)$. If so, using Proposition 15.5 we have

$$\left\| f^\alpha(sh) \right\|_E = \left\| f^\alpha(s) \right\|_E = \left\| f(s \cdot x_0) \right\|_E$$

for all $s \in G$ and all $h \in H$, and since by Proposition 8.43 and Theorem 7.10, for any $g \in \mathrm{L}^2(G/H; \mathbb{C})$, we have

$$\int_{G/H} g \, d\mu = \int_G (g \circ \pi) \, d\lambda,$$

so we obtain

$$N_2(f^\alpha)^2 = \int_G \|f^\alpha(s)\|_E^2 \, d\lambda(s) = \int_G \|f(s\cdot x_0)\|_E^2 \, d\lambda(s) = N_2(f\circ\pi)^2$$
$$N_2(f\circ\pi)^2 = \int_G \|f(s\cdot x_0)\|_E^2 \, d\lambda(s) = \int_{G/H} \|f(x)\|_E \, d\mu(x) = N_2(f)^2,$$

that is, $N_2(f^\alpha) = N_2(f)$, and we conclude that $f^\alpha \in \mathcal{L}^2_\lambda(G; E)$.

Conversely, if $g \in \mathcal{L}^2_\lambda(G; E)$ satisfies the property

$$g(sh) = U(h^{-1})(g(s)) \qquad \text{for all } s \in G \text{ and all } h \in H, \tag{$*_U$}$$

and if the map $s \mapsto \alpha(s, x_0)(g(s))$ from G to E is λ-measurable, then as in Proposition 15.6 we can write this map as $f\circ\pi$ for some $f \in \mathrm{L}^2_\mu(X; E)$, and we have $g = f^\alpha$.

In this case, up to equivalence, we can consider the unitary representation $\mathrm{Ind}^G_{H,F}\,\alpha$ induced by α as a unitary representation of G in the closed subspace F of $\mathrm{L}^2_\lambda(G; E)$ spanned by the functions $g \in \mathcal{L}^2_\lambda(G; E)$ satisfying property $(*_U)$. Then for all $s \in G$,

$$(\mathrm{Ind}^G_{H,F}\,\alpha)_s(g) = \lambda_s g, \qquad \text{for all } g \in F, \tag{Ind_G}$$

equivalently, for all $s, t \in G$,

$$((\mathrm{Ind}^G_{H,F}\,\alpha)_s(g))(t) = g(s^{-1}t), \qquad \text{for all } g \in F.$$

Notice the analogy with Proposition 15.7.

Note that $\mathrm{Ind}^G_{H,F}\,\alpha$ depends only on U, so we usually write $\mathrm{Ind}^G_{H,F}\,U$ instead of $\mathrm{Ind}^G_{H,F}\,\alpha$.

If $E = \mathbb{C}$, then $\mathrm{Ind}^G_{H,F}\,U$ is a subrepresentation of the regular representation of G in $\mathrm{L}^2(G)$.

Definition 15.13. If we choose U to be the trivial representation of H in E, then the functions $g \in \mathrm{L}^2_\lambda(G; E)$ satisfying Condition $(*_U)$ are constant on the classes sH, so we can identify F with $\mathrm{L}^2_\mu(X; E)$. In this case we say that the induced representation $\mathrm{Ind}^G_H\,U$ of G in $\mathrm{L}^2_\mu(X; E)$ is the *canonical representation* of G corresponding to the compact subgroup H and to its trivial representation in E.

If $H = (e)$ and $E = \mathbb{C}$, then the induced representation is the regular representation of G in $\mathrm{L}^2(G)$.

Going back to the case where H is an arbitrary closed subgroup of G, and where there is a G-invariant measure on G/H, there is another method, not using cocyles, for defining a unitary induced representation of G from a unitary representation $U\colon H \to \mathbf{U}(E)$. We can define a Hilbert space $\mathcal{H}$ such that formula (Ind_G) defines a unitary induced representation $\mathrm{Ind}^G_{H,\mathcal{H}}\,U$ of G in $\mathcal{H}$. This method is described in Folland [33] (Chapter 6, Section 1), and we briefly describe it.

Given a unitary representation $U\colon H \to \mathbf{U}(E)$, let $\mathcal{H}_0$ be the following set of functions:

$$\begin{aligned}\mathcal{H}_0 = \{f \in \mathcal{C}(G, E) \mid \pi(\mathrm{supp}(f)) \text{ is compact and}\\ f(sh) = U(h^{-1})(f(s)) \quad \text{for all } s \in G \text{ and all } h \in H\}.\end{aligned}$$

The problem is that it is not obvious that $\mathcal{H}_0$ is nonempty! However, the following result proven in Folland [33] (Chapter 6, Proposition 6.1) shows that this is not the case.

Proposition 15.10. *If $\varphi\colon G \to E$ is a continuous function with compact support, then the function f_φ from G to E given by*

$$f_\varphi(s) = \int_H U(h)(\varphi(hs))\, d\lambda_H(h)$$

belongs to $\mathcal{H}_0$ and is uniformly continuous on G. Moreover, every element of $\mathcal{H}_0$ is of the form f_φ for some $\varphi \in \mathcal{K}(G, E)$.

The group G acts on the left on $\mathcal{H}_0$ by $f \mapsto \lambda_s f$. In order to act by unitary maps, we need to define an inner product on $\mathcal{H}_0$ with respect to which these left translations are isometries. Since G/H has G-invariant measures, this is easy to achieve. If $f, g \in \mathcal{H}_0$, then the map $s \mapsto \langle f(s), g(s)\rangle_E$ depends only on the coset sH, so we can define the inner product $\langle f, g\rangle$ by

$$\langle f, g\rangle = \int_{G/H} \langle f(s), g(s)\rangle_E \, d\mu(sH).$$

This is clearly a positive hermitian form, and it is positive definite because $\mu(A) > 0$ for every nonempty open set A. This inner product is invariant under the left translations λ_s because μ is G-invariant. Therefore, with respect to this inner product, the maps $f \mapsto \lambda_s f$ are unitary. If $\mathcal{H}$ is the Hilbert space which is the completion of $\mathcal{H}_0$, then the maps $f \mapsto \lambda_s f$ extend to unitary operators on $\mathcal{H}$. It follows from Proposition 15.10 that the map $s \mapsto \lambda_s f$ from G to $\mathcal{H}$ are continuous for every $f \in \mathcal{H}_0$. Therefore, they define a unitary representation of G in $\mathcal{H}$ given by

$$(\mathrm{Ind}_{H,\mathcal{H}}^G U)_s(f) = \lambda_s(f), \qquad f \in \mathcal{H}.$$

This unitary representation has the advantage that it depends only on U, but one should not neglect the fact that the construction involving cocycles allows more flexibility. The Hilbert space $\mathcal{H}$ is also more complicated than the Hilbert space $\mathrm{L}^2_\mu(X; E)$.

When G/H admits no G-invariant measure, then we need to use a weaker notion of invariance. It turns out that the notion of (strong) quasi-invariance does the job.

15.6 Quasi-Invariant Measures on G/H

The notion of quasi-invariance was first introduced by Mackey and Bruhat in the early 1950's. It also occurs in Bourbaki [7] (Chapter VII, §2, No. 5). We follow the exposition in Folland [33] (Chapter 2, Section 2.6, and Chapter 6, Section 1).

As we said in Section 15.5, given any measure μ on $X = G/H$, for any $s \in G$, the measure $\lambda_s(\mu)$ is given by

$$(\lambda_s(\mu))(A) = \mu(s^{-1} \cdot A),$$

for every Borel subset A of $X = G/H$. We say μ is G-invariant if for every Borel subset A of X,

$$\mu(s^{-1} \cdot A) = \mu(A) \qquad \text{for all } s \in G.$$

In this case,

$$\int_{G/H} g(s \cdot x)\, d\mu(x) = \int_{G/H} g(x)\, d\mu(x), \qquad \text{for all } s \in G$$

and for all $g \in \mathrm{L}^1_\mu(G/H)$. It is not hard to prove an analog of Proposition 8.16(3), namely

$$\int_{G/H} g(s \cdot x)\, d\mu(x) = \int_{G/H} g(x)\, d\lambda_s(\mu)(x)$$

for all $g \in \mathrm{L}^1_\mu(G/H)$ and all $s \in G$. A weaker requirement than G-invariance is that

$$\int_{G/H} g(s \cdot x)\, d\mu(x) = \int_{G/H} g(x)\, d\lambda_s(\mu)(x) = \int_{G/H} \varrho(s, x) g(x)\, d\mu(x),$$

for some continuous function $\varrho\colon G \times (G/H) \to (0, \infty)$, for all $g \in \mathcal{K}_{\mathbb{C}}(G/H)$ and all $s \in G$. The above discussion suggests the following definition.

Definition 15.14. A measure μ on G/H is *(strongly) quasi-invariant* if there is a continuous function $\varrho\colon G \times (G/H) \to (0, \infty)$ such that

$$\int_{G/H} g(s \cdot x)\, d\mu(x) = \int_{G/H} \varrho(s, x) g(x)\, d\mu(x), \quad \text{for all } g \in \mathcal{K}_{\mathbb{C}}(G/H) \text{ and all } s \in G. \quad (\mathrm{qi}_\varrho)$$

The key to quasi-invariance is the existence of certain functions from G to $(0, \infty)$ called *rho-functions*.

Definition 15.15. A function $\rho\colon G \to (0, \infty)$ is a *rho-function* for the pair (G, H) if it is a continuous function such that

$$\rho(sh) = \frac{\Delta_H(h)}{\Delta_G(h)}\, \rho(s), \qquad s \in G, h \in H, \qquad (*_\rho)$$

where Δ_G is the modular function on G and Δ_H is the modular function on H.

Proposition 15.11. *If G is any locally compact group and H is any closed subgroup of G, then (G, H) admits rho-functions.*

Proposition 15.11 is proven in Folland [33] (Chapter 2, Proposition 2.54). One first proves ([33] (Chapter 2, Lemma 2.53) that there is a continuous function $\varphi\colon G \to (0, \infty)$ such that the following properties hold:

(i) $\{y \in G \mid \varphi(y) > 0\} \cap sH \neq \emptyset$ for all $s \in G$.

(ii) $\mathrm{supp}(\varphi) \cap KH$ is compact for every compact subset K of G.

Then define ρ by
$$\rho(s) = \int_H \frac{\Delta_G(h)}{\Delta_H(h)}\, \varphi(sh)\, d\lambda_H(h).$$
It is not hard to check that the above function is a rho-function.

Recall from Definition 8.20 the definition of the projection map $P\colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G/H)$ defined as follow: for every $f \in \mathcal{K}_{\mathbb{C}}(G)$, for every $s \in G$, let
$$(P(f))(sH) = \int_H f(sh)\, d\lambda_H(h).$$
By Proposition 8.40, the map P is surjective.

The next proposition is proven in Folland [33] (Chapter 2, Lemma 2.55).

Proposition 15.12. *For any function $f \in \mathcal{K}_{\mathbb{C}}(G)$, if $P(f) = 0$, then $\int f\rho\, d\lambda = 0$, for any rho-function ρ.*

The proof of Proposition 15.12 is very similar to the argument given in the proof of Theorem 8.42. Then we have our first main theorem. Recall that $\pi\colon G \to G/H$ denotes the quotient map.

Theorem 15.13. *Let G be any locally compact group and H be any closed subgroup of G. For every rho-function ρ for (G, H), there is a unique σ-Radon measure μ on G/H such that*
$$\int_{G/H} P(f)(x)\, d\mu(x) = \int_G f(s)\rho(s)\, d\lambda(s), \qquad \text{for all } f \in \mathcal{K}_{\mathbb{C}}(G). \tag{qi}$$
Furthermore, if we let $\varrho\colon G \times (G/H) \to (0, \infty)$ be the continuous function given by
$$\varrho(s, \pi(t)) = \frac{\rho(s^{-1}t)}{\rho(t)} \qquad s, t \in G,$$
then for every $g \in \mathcal{K}_{\mathbb{C}}(G/H)$, we have
$$\int_{G/H} g(s \cdot x)\, d\mu(x) = \int_{G/H} \varrho(s, x) g(x)\, d\mu(x), \qquad \text{for all } s \in G, \tag{qi$_\varrho$}$$
which means that μ is strongly quasi-invariant.

Proof. Theorem 15.13 is proven in Folland [33] (Chapter 2, Theorem 2.56). For any $f \in \mathcal{K}_{\mathbb{C}}(G)$, since P is surjective and since by Proposition 15.12, if $P(f) = P(g)$, then $\int_G f\rho\, d\lambda = \int_G g\rho\, d\lambda$, the map Φ given by $\Phi(P(f)) = \int_G f\rho\, d\lambda$ is a well-defined positive linear functional

on $\mathcal{K}_{\mathbb{C}}(G/H)$. By Radon–Riesz I, it defines a unique σ-Radon measure μ on G/H satisfying (qi).

The equation

$$\rho(sh) = \frac{\Delta_H(h)}{\Delta_G(h)}\,\rho(s), \qquad s \in G, h \in H,$$

satisfied by a rho-function shows that the ratio $\rho(st)/\rho(t)$ depends only on the coset $\pi(t) = tH$, because

$$\frac{\rho(sth)}{\rho(th)} = \frac{\Delta_H(h)}{\Delta_G(h)}\,\frac{\rho(st)}{\rho(th)} = \frac{\Delta_H(h)}{\Delta_G(h)}\,\frac{\Delta_G(h)}{\Delta_H(h)}\,\frac{\rho(st)}{\rho(t)} = \frac{\rho(st)}{\rho(t)},$$

so we obtain a continuous function $\varrho\colon G \times (G/H) \to (0, \infty)$ given by

$$\varrho(s, \pi(t)) = \frac{\rho(s^{-1}t)}{\rho(t)} \qquad s, t \in G.$$

First by expanding both integrals as double integrals it is easy to show that

$$\int_{G/H} P(f)(s \cdot x)\, d\mu(x) = \int_{G/H} P(\lambda_{s^{-1}} f)(x)\, d\mu(x).$$

Then we have

$$\begin{aligned}
\int_{G/H} P(f)(s \cdot x)\, d\mu(x) &= \int_{G/H} P(\lambda_{s^{-1}} f)(x)\, d\mu(x) \\
&= \int_G f(st)\rho(t)\, d\lambda(t) \\
&= \int_G f(t)\rho(s^{-1}t)\, d\lambda(t) \\
&= \int_G \varrho(s, \pi(t)) f(t)\rho(t)\, d\lambda(t) \\
&= \int_{G/H} P(\varrho(s, \pi(-)) f)(x)\, d\mu(x) \\
&= \int_{G/H} \varrho(s, x) P(f)(x)\, d\mu(x),
\end{aligned}$$

where we used Proposition 8.38(3) to prove the last step, which concludes the proof. □

Remark: The map $x \mapsto \varrho(s, x)$ is the Radon–Nikodym derivative of $\lambda_s(\mu)$ with respect to μ.

The following converse of Theorem 15.13 is proven in Folland [33] (Chapter 2, Theorem 2.59).

Theorem 15.14. *Let G be any locally compact group and H be any closed subgroup of G. Every quasi-invariant measure μ on G/H arises from a rho-function as in* (qi) *and* (qi_ϱ). *Furthermore, any two such measures μ and μ' are strongly equivalent, which means that there is a continuous function $\varphi\colon G/H \to (0, \infty)$ such that $\int_{G/H} g(x)\, d\mu'(x) = \int_{G/H} \varphi(x) g(x)\, d\mu(x)$ for all $g \in \mathcal{K}_{\mathbb{C}}(G/H)$.*

The following proposition shows that ϱ behaves like a cocycle.

Proposition 15.15. *Let G be any locally compact group and H be any closed subgroup of G. For any quasi-invariant measure μ on G/H associated with the continuous function $\varrho\colon G \times (G/H) \to (0, \infty)$, we have*

$$\varrho(st, x) = \varrho(s, t \cdot x)\varrho(t, x), \qquad \text{for all } s, t \in G \text{ and all } x \in G/H \tag{$*_\varrho$}$$

Proof. Using (qi_ϱ), for every function $g \in \mathcal{K}_{\mathbb{C}}(G/H)$, we have

$$\begin{aligned}
\int_{G/H} g(x)\varrho(st, x)\, d\mu(x) &= \int_{G/H} g((st) \cdot x)\, d\mu(x) = \int_{G/H} g(s \cdot (t \cdot x))\, d\mu(x) \\
&= \int_{G/H} \varrho(s, t \cdot x) g(t \cdot x)\, d\mu(x) \\
&= \int_{G/H} g(x)\varrho(s, t \cdot x)\varrho(t, x)\, d\mu(x),
\end{aligned}$$

which proves that

$$\varrho(st, x) = \varrho(s, t \cdot x)\varrho(t, x),$$

as claimed. ∎

Remark: Dieudonné denotes $\varrho(s, x)$ by $J_s(x)$; see Dieudonné [22] (Chapter XXII, Section 3, No. 22.3.8.1-22.3.8.2).

We now use quasi-invariant measures to generalize the construction of Section 15.5.

15.7 Induced Representations, II; G/H has a Quasi-Invariant Measure

If μ is a quasi-invariant measure on G/H, then by making a simple modification to Condition (1) of Theorem 15.9 we obtain the following result.

Theorem 15.16. *Let G be a locally compact group, H be a closed subgroup of G, E be a separable Hilbert space, and $U\colon H \to \mathbf{U}(E)$ be a unitary representation of H. For any quasi-invariant measure μ on $X = G/H$ associated with the continuous function $\varrho\colon G \times (G/H) \to (0, \infty)$, for any cocycle $\alpha\colon G \times X \to \mathbf{U}(E)$, if the following conditions hold*

(1) *The map* $\varrho(s^{-1}, x)^{1/2}\alpha(s, x)$ *is a unitary map of* E *for all* $s \in G$ *and all* $x \in X$, *such that* $\varrho(h^{-1}, x_0)^{1/2}\alpha(h, x_0) = U(h)$ *for all* $h \in H$;

(2) *For every* $s \in G$, *for every* $f \in \mathrm{L}^2_\mu(X; E)$, *the map* $x \mapsto \alpha(x, s)(f(x))$ *from* X *to* E *is* μ-*measurable;*

(3) *For every* $f \in \mathrm{L}^2_\mu(X; E)$, *the map* $s \mapsto \Pi_s(f)$ *is a continuous map from* G *to* $\mathrm{L}^2_\mu(X; E)$, *where* Π *is the homomorphism* $\Pi\colon G \to \mathbf{GL}(E^X)$ *induced by the cocycle* α;

then the homomorphism $\Pi\colon G \to \mathbf{U}(\mathrm{L}^2_\mu(X; E))$ *induced by the cocycle* α *given by*

$$(\Pi_s(f))(x) = (\alpha(s^{-1}, x))^{-1}(f(s^{-1} \cdot x)), \qquad f \in \mathrm{L}^2_\mu(X; E),\, x \in X,$$

(see Definition 15.3) is a unitary representation of G.

Proof. We simply have to prove that

$$N_2(\Pi_s(f)) = N_2(f), \qquad \text{for all } f \in \mathcal{L}^2_\mu(X; E) \text{ and all } s \in G,$$

which implies that $\Pi_s(f) \in \mathcal{L}^2_\mu(X; E)$, and the other conditions imply that the homomorphism $\Pi\colon G \to \mathbf{GL}(\mathrm{L}^2_\mu(X; E))$ induced by α is a unitary representation of G. Since by hypothesis $\varrho(s^{-1}, x)^{1/2}\alpha(s, x)$ is a unitary operator, using (qi_ϱ), we have

$$\begin{aligned}
N_2(\Pi_s(f))^2 &= \int_{G/H} \left\|\alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x))\right\|^2 d\mu(x) \\
&= \int_{G/H} \varrho(s^{-1}, x) \left\|\alpha(s, x)(f(x))\right\|^2 d\mu(x) \\
&= \int_{G/H} \left\|\varrho(s^{-1}, x)^{1/2}\alpha(s, x)(f(x))\right\|^2 d\mu(x) \\
&= \int_{G/H} \left\|f(x)\right\|^2 d\mu(x),
\end{aligned}$$

and so $N_2(\Pi_s(f)) = N_2(f)$. $\square$

As an application of Theorem 15.16, we can pick

$$\alpha(s, x) = (\varrho(s^{-1}, x))^{-1/2+ri}\,\mathrm{id}_E,$$

with $r \in \mathbb{R}$. By Proposition 15.15, the function α is a cocycle. Condition (2) is satisfied because ϱ is measurable (in fact, continuous). The maps $\varrho(s^{-1}, x)^{1/2}\alpha(s, x) = (\varrho(s^{-1}, x))^{ri}\,\mathrm{id}_E$ are unitary, since they are multiplication by a complex number of modulus 1. It remains to check Condition (3). This verification is performed in Dieudonné [22] (Chapter XXII, Section 3, No. 22.3.8.3).

15.8 Induced Representations, III; Blattner's Method

It is possible to modify the construction of the Hilbert space $\mathcal{H}$ and the inner product described at the end of Section 15.5 to deal with the situation where G/H has no G-invariant measure. This can be done in two ways as explained in Folland [33] (Chapter 6, Section 6.1). These constructions yield induced unitary representations of G from a unitary representation $U\colon H \to \mathbf{U}(E)$ of H and do not involve cocycles.

First method. In the first construction, the space $\mathcal{H}_0$ and the inner product are defined as before, namely

$$\begin{aligned}\mathcal{H}_0 = \{f \in \mathcal{C}(G,E) \mid \pi(\mathrm{supp}(f)) &\text{ is compact and}\\ f(sh) = U(h^{-1})(f(s)) &\quad \text{for all } s \in G \text{ and all } h \in H\},\end{aligned}$$

and

$$\langle f, g\rangle = \int_{G/H} \langle f(s), g(s)\rangle_E \, d\mu(sH).$$

Recall that $\pi\colon G \to G/H$ denotes the quotient map. The Hilbert space $\mathcal{H}$ is the completion of $\mathcal{H}_0$. The new ingredient is that to make the operators Π_s unitary, we use a quasi-invariant measure μ associated with a continuous function $\varrho\colon G \times (G/H) \to (0,\infty)$. We define Π^μ_s by

$$(\Pi^\mu_s(f))(t) = \varrho(s, tH)^{1/2} f(s^{-1}t), \qquad f \in \mathcal{H},\ s,t \in G. \tag{indv1}$$

Then, as in the proof of Theorem 15.16, we check that the operators Π^μ_s are unitary with respect to the inner product on $\mathcal{H}$ defined above, and we obtain a unitary representation $\Pi^\mu\colon G \to \mathbf{U}(\mathcal{H})$, also denoted $\mathrm{Ind}^{G,\mu}_{H,\mathcal{H}}\, U$ (for short $\mathrm{Ind}^G_H\, U$). This representation depends on μ, but it can be shown that if μ' is another quasi-invariant measure on G/H associated with $\varrho'\colon G \times (G/H) \to (0,\infty)$, then the automorphism $f \mapsto (\rho'/\rho)^{1/2} f$ is a unitary equivalence of ρ^μ and $\rho^{\mu'}$, where ρ and ρ' are the rho-functions associated with μ and μ' (see Theorem 15.14).

Blattner's Method. The second construction, due to Blattner, does not make use of quasi-invariant measures, but instead modifies the definition of the space $\mathcal{H}_0$. In this sense, it is more intrinsic. Define the space $\mathcal{H}^0$ as

$$\begin{aligned}\mathcal{H}^0 = \Bigg\{ f \in \mathcal{C}(G,E) \mid \pi(\mathrm{supp}(f)) &\text{ is compact and}\\ f(sh) = \left(\frac{\Delta_H(h)}{\Delta_G(h)}\right)^{1/2} U(h^{-1})(f(s)) &\quad \text{for all } s \in G \text{ and all } h \in H\Bigg\}.\end{aligned}$$

Again, it is not obvious that $\mathcal{H}^0$ is not empty, but Proposition 15.10 can be modified (by adding the factor $\left(\frac{\Delta_G(h)}{\Delta_H(h)}\right)^{1/2}$ under the integral) to show that $\mathcal{H}^0$ is nonempty.

Next we need to define an inner product on $\mathcal{H}^0$ so that the operators Π_s become unitary. The construction of such an inner product is a bit eccentric. For every $f \in \mathcal{H}^0$, the map

$s \mapsto \|f(s)\|_E^2$ is almost a rho-function, except that it is not strictly positive. However, it is still possible to prove that the map

$$P\colon \varphi \mapsto \int_G \varphi(s)\, \|f(s)\|_E^2\, d\lambda(s), \qquad \varphi \in \mathcal{K}_{\mathbb{C}}(G)$$

is a positive Radon functional on $\mathcal{K}_{\mathbb{C}}(G/H)$, so by Radon–Riesz I, there is σ-Radon measure μ_f on G/H such that

$$\int_{G/H} P(\varphi)\, d\mu_f = \int_G \varphi(s)\, \|f(s)\|_E^2\, d\lambda(s),$$

for all $\varphi \in \mathcal{K}_{\mathbb{C}}(G)$. Furthermore, the support of μ_f is contained in $\pi(\mathrm{supp}(f))$, hence compact. Therefore, $\mu_f(G/H)$ is finite. Then, given $f, g \in \mathcal{H}^0$, define the complex measure $\mu_{f,g}$ by polarization as

$$\mu_{f,g} = \frac{1}{4}(\mu_{f+g} - \mu_{f-g} + i\mu_{f+ig} - i\mu_{f-ig}),$$

so that

$$\int_{G/H} P(\varphi)\, d\mu_{f,g} = \int_G \varphi(s) \langle f(s), g(s)\rangle_E\, d\lambda(s), \qquad \varphi \in \mathcal{K}_{\mathbb{C}}(G).$$

The inner product on $\mathcal{H}^0$ is defined as

$$\langle f, g\rangle = \mu_{f,g}(G/H).$$

It can be verified that we obtain a hermitian inner product, and we let $\mathcal{H}'$ be the Hilbert space completion of $\mathcal{H}^0$. Finally, it can be verified that the operators Π_s are unitary with respect to the inner product on $\mathcal{H}'$ defined above, where

$$(\Pi_s(f))(t) = f(s^{-1}t), \qquad f \in \mathcal{H}',\ s, t \in G. \tag{indv2}$$

Therefore we obtain a unitary representation $\Pi'\colon G \to \mathbf{U}(\mathcal{H}')$, also denoted $\mathrm{Ind}_{H,\mathcal{H}'}^G U$ (for short $\mathrm{Ind}_H^G U$).

It can also be shown that for any quasi-invariant measure μ on G/H, the representations $\Pi^\mu\colon G \to \mathbf{U}(\mathcal{H})$ and $\Pi'\colon G \to \mathbf{U}(\mathcal{H}')$ are equivalent (the isomorphism between $\mathcal{H}$ and $\mathcal{H}'$ is given by $f \mapsto \rho^{1/2} f$).

There is also an interpretation of the representations Π^μ and Π' as sections of homogeneous hermitian vector bundles over G/H, but this would lead us to far; see Folland [33] (Chapter 6) or Kirillov [53] (Section 13).

15.9 Examples of Induced Representations Via Method II

We will now give several examples of the application of Theorem 15.16 to the group $\mathbf{SL}(2,\mathbb{R})$. It turns out that the group $\mathbf{SL}(2,\mathbb{R})$ has no finite-dimensional unitary representations except the trivial one, and Theorem 15.16 can be used to produce nontrivial unitary representations.

Example 15.4. Let $G = \mathbf{SL}(2,\mathbb{R})$ and $H = S_1$ be the subgroup

$$S_1 = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \;\middle|\; a, b \in \mathbb{R}, a \neq 0 \right\},$$

and let $E = \mathbb{C}$. We claim that the homogeneous space $\mathbf{SL}(2,\mathbb{R})/S_1$ is homeomorphic to $\mathbb{P}^1(\mathbb{R}) = \mathbb{RP}^1$, the real projective line. Indeed, there is an action of $\mathbf{SL}(2,\mathbb{R})$ on $\mathbb{RP}^1$ viewed as $\mathbb{R} \cup \{\infty\}$ given by

$$s \cdot z = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}, \qquad z \in \mathbb{RP}^1,$$

with the convention that when $z = -d/c$, then the result is ∞, and when $z = \infty$, then the result if a/c. It is easy to check that this action is transitive and that the stabilizer of ∞ is the subgroup S_1. We give $\mathbb{RP}^1$ the measure μ which is the Lebesgue measure extended so that $\{\infty\}$ has measure zero. Since

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

(recall that $ad - bc = 1$), and since the derivative of the function

$$x \mapsto \frac{dx - b}{-cx + a}$$

is

$$\frac{d(-cx+a) - (dx-b)(-c)}{(-cx+a)^2} = \frac{1}{(cx-a)^2},$$

we see that for any function $f \in \mathrm{L}^2_\mu(\mathbb{RP}^1;\mathbb{C})$, using the change of variable $x = \frac{az+b}{cz+d}$,

$$\int_{-\infty}^{+\infty} f(s\cdot z)\, d\mu(z) = \int_{-\infty}^{+\infty} f\left(\frac{az+b}{cz+d}\right) d\mu(z) = \int_{-\infty}^{+\infty} \frac{1}{(cx-a)^2} f(x)\, d\mu(x).$$

It follows that μ is quasi-invariant with

$$\varrho(s,x) = \frac{1}{(cx-a)^2}, \quad \text{where} \quad s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The method of Theorem 15.16 with

$$\alpha(s,x) = (\varrho(s^{-1},x))^{-1/2+(r/2)i}\,\mathrm{id}_{\mathbb{C}}$$

where $r \in \mathbb{R}$, and with

$$(\Pi_s(f))(x) = (\alpha(s^{-1},x))^{-1}(f(s^{-1}\cdot x)),$$

yields the unitary representations of $\mathbf{SL}(2,\mathbb{R})$ in $\mathbf{L}^2_\mu(\mathbb{RP}^1;\mathbb{C})$ given by

$$\Pi_s(f)(x) = |cx-a|^{-1+ri} f\left(\frac{b-dx}{cx-a}\right), \qquad f \in \mathbf{L}^2_\mu(\mathbb{RP}^1;\mathbb{C}), \quad \text{where} \quad s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

It is also easy to check that the cocycles

$$\alpha(s,x) = \left(\frac{1}{(cx-a)^2}\right)^{-1/2+(r/2)i} \mathrm{sign}(cx-a)\,\mathrm{id}_{\mathbb{C}}$$

with $r \in \mathbb{R}$ also work, and we get the representations of $\mathbf{SL}(2,\mathbb{R})$ in $\mathbf{L}^2_\mu(\mathbb{RP}^1;\mathbb{C})$ given by

$$\Pi_s(f)(x) = |cx-a|^{-1+ri}\,\mathrm{sign}(cx-a) f\left(\frac{b-dx}{cx-a}\right), \quad f \in \mathbf{L}^2_\mu(\mathbb{RP}^1;\mathbb{C}), \quad \text{where} \quad s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

It can be shown that all these representations are irreducible and pairwise inequivalent for $r > 0$. These representations constitute the *principal series* of irreducible unitary representations of $\mathbf{SL}(2,\mathbb{R})$.

Example 15.5. Let $G = \mathbf{SL}(2,\mathbb{R})$ and $H = \mathbf{SO}(2)$ be the subgroup

$$\mathbf{SO}(2) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \;\middle|\; 0 \leq \theta \leq 2\pi, \right\},$$

and let $E = \mathbb{C}$. We claim that the homogeneous space $\mathbf{SL}(2,\mathbb{R})/\mathbf{SO}(2)$ is homeomorphic to the upper half plane $P = \{z = x + iy \in \mathbb{C} \mid y > 0\}$. Indeed, there is an action of $\mathbf{SL}(2,\mathbb{R})$ on P given by

$$s \cdot z = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}, \qquad z = x + iy \in P,$$

It is easy to check that this action is transitive and that the stabilizer of $z = i$ is $\mathbf{SO}(2)$. Since the group $\mathbf{SO}(2)$ is compact, the space $P = \mathbf{SL}(2,\mathbb{R})/\mathbf{SO}(2)$ admits $\mathbf{SL}(2,\mathbb{R})$-invariant measures. In fact, the measure μ corresponding to the positive Radon functional

$$h \mapsto \int_P h(x+iy)\frac{dx\,dy}{y^2} = \int_{y>0}\int_{x=-\infty}^{+\infty} h(x+iy)\frac{dx\,dy}{y^2}, \qquad h \in \mathcal{K}_{\mathbb{C}}(P)$$

is such a measure.

We will need a method for picking a representative in every coset of $\mathbf{SL}(2,\mathbb{R})/\mathbf{SO}(2)$ that corresponds in a one-to-one fashion to an element $z = x + iy \in P$. For this, we use the fact that every matrix $s \in \mathbf{SL}(2,\mathbb{R})$ can be uniquely factored as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda & \mu \\ 0 & \lambda^{-1} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \lambda\cos\theta + \mu\sin\theta & -\lambda\sin\theta + \mu\cos\theta \\ \lambda^{-1}\sin\theta & \lambda^{-1}\cos\theta \end{pmatrix},$$

with $\lambda, \mu \in \mathbb{R}, \lambda > 0$, and $0 \leq \theta < 2\pi$.

Indeed, if there is such a decomposition, then

$$c = \lambda^{-1}\sin\theta, \qquad d = \lambda^{-1}\cos\theta,$$

so

$$\sin\theta = \lambda c, \qquad \cos\theta = \lambda d,$$

and since

$$\begin{pmatrix} \lambda & \mu \\ 0 & \lambda^{-1} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} a\cos\theta - b\sin\theta & a\sin\theta + b\cos\theta \\ c\cos\theta - d\sin\theta & c\sin\theta + d\cos\theta \end{pmatrix},$$

we see that

$$\lambda^{-1} = c\sin\theta + d\cos\theta = \lambda(c^2 + d^2),$$

and since $ad - bc = 1$, we have $c^2 + d^2 \neq 0$, so we can pick

$$\lambda = \frac{1}{\sqrt{c^2 + d^2}},$$

and then $\theta \in [0, 2\pi)$ is uniquely determined by

$$\cos\theta = \frac{d}{\sqrt{c^2 + d^2}}, \qquad \sin\theta = \frac{c}{\sqrt{c^2 + d^2}},$$

and μ is determined by

$$\mu = a\sin\theta + b\cos\theta = \lambda(ac + bd) = \frac{ac + db}{\sqrt{c^2 + d^2}}.$$

Observe that the group

$$S_0 = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \;\middle|\; a, b \in \mathbb{R}, a > 0 \right\}$$

is a subgroup of the group S_1 of Example 15.4.

Given any $z = x + iy \in P$, there is a unique coset $s\mathbf{SO}(2) \subseteq \mathbf{SL}(2,\mathbb{R})$ (where $s \in \mathbf{SL}(2,\mathbb{R})$) that maps i to z, and in view of the above factorization of matrices in $\mathbf{SL}(2,\mathbb{R})$, we can pick as a representative of this coset $s\mathbf{SO}(2)$ the matrix $r_z \in S_0$ such that

$$r_z \cdot i = z = x + iy,$$

namely

$$r_z = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix}.$$

We now determine $u(s,z)$ such that $sr_z = r_{s\cdot z}u(s,z)$ (see Definition 15.5), with $u(s,z) \in \mathbf{SO}(2)$ and

$$s = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

as in Section 15.2 (see Propositions 15.3 and 15.4). Since the imaginary part of $s \cdot z = (az+b)/(cz+d)$ is $y/|cz+d|^2$, we have

$$r_{s\cdot z} = \begin{pmatrix} \sqrt{y}/|cz+d| & * \\ 0 & |cz+d|/\sqrt{y} \end{pmatrix},$$

so the equation $sr_z = r_{s\cdot z}u(s,z)$ with

$$u(s,z) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

namely

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix} = \begin{pmatrix} \sqrt{y}/|cz+d| & * \\ 0 & |cz+d|/\sqrt{y} \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

yields in particular

$$c\sqrt{y} = \frac{|cz+d|}{\sqrt{y}}\sin\theta, \qquad \frac{cx+d}{\sqrt{y}} = \frac{|cz+d|}{\sqrt{y}}\cos\theta.$$

Therefore,

$$\cos\theta + i\sin\theta = \frac{cx+d+ciy}{|cz+d|} = \frac{cz+d}{|cz+d|},$$

equivalently

$$e^{i\theta} = \frac{cz+d}{|cz+d|}.$$

The group $\mathbf{SO}(2)$ is abelian, and since its unitary representations in $\mathbb{C}$ are characters, by Proposition 10.9, they are of the form

$$h \mapsto \sigma_n(h) = e^{ni\theta}, \qquad n \in \mathbb{Z},$$

with

$$h = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

By the method of Section 15.2, since $e^{i\theta} = \frac{cz+d}{|cz+d|}$, we can pick the cocycle α to be

$$\alpha(s, z) = \sigma_n(u(s, z)) = \frac{(cz+d)^n}{|cz+d|^n} \, \mathrm{id}_{\mathbb{C}},$$

and then

$$(\Pi_s(f))(z) = (\alpha(s^{-1}, x))^{-1}(f(s^{-1} \cdot z)), \qquad f \in \mathrm{L}^2_\mu(P; \mathbb{C}),\ z \in P.$$

Since

$$s^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

we get

$$\alpha(s^{-1}, z) = \frac{(cz-a)^n}{|cz-a|^n} \, \mathrm{id}_{\mathbb{C}},$$

and thus

$$(\Pi_s(f))(z) = \frac{(cz-a)^{-n}}{|cz-a|^{-n}} f\left(\frac{b-dz}{cz-a}\right), \qquad f \in \mathrm{L}^2_\mu(P; \mathbb{C}),\ z \in P,\ s = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

and for $n \in \mathbb{Z}$.

These representations are not always irreducible. A way to see this is to construct an equivalent representation by using the cocycles $\alpha'(s, z) = c(s \cdot z) \circ \alpha(s, z) \circ c(z)^{-1}$, as in Section 15.2, with

$$c(z) = c(x + iy) = y^{-n/2}.$$

The image of $\mathrm{L}^2_\mu(P; \mathbb{C})$ under the map $f \mapsto cf$ is the space E_n of functions $g \colon P \to \mathbb{C}$ such that the map $z \mapsto y^n g(z)^2$ is μ-integrable. One can then show that the equivalent unitary representation is given by

$$(\Pi_s(f))(z) = (cz-a)^{-n} g\left(\frac{b-dz}{cz-a}\right), \qquad g \in E_n,\ z \in P,\ s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

It can be shown that for $n > 1$, the space H^2_n of holomorphic functions on P is nonempty and invariant under Π. Furthermore, these representations in H^2_n are irreducible. The complex conjugates of these representations are also irreducible and not equivalent to the previous ones; see Dieudonné [22] (Chapter XXII, Section 3).

These irreducible unitary representations of $\mathbf{SL}(2, \mathbb{R})$ constitute the *discrete series*. There are other irreducible unitary representations of $\mathbf{SL}(2, \mathbb{R})$ called the *complementary series*.

For comprehensive treatments of the irreducible representations of $\mathbf{SL}(2, \mathbb{R})$ and other semisimple Lie groups, see Knapp [56], Vilenkin [101], and Taylor [96].

15.10 Partial Traces, Induced Representations of Compact Groups

In this section we consider a compact (metrizable) group G and a closed subgroup H of G, and our goal is to determine the canonical (unitary) representation of G in $\mathrm{L}^2_\mu(G/H;\mathbb{C})$ induced by the trivial representation of H in $E = \mathbb{C}$ (see Definition 15.13), where μ is the G-invariant measure on G/H induced by a Haar measure λ on G. For simplicity of notation we write $\mathrm{L}^2_\mu(G/H)$ instead of $\mathrm{L}^2_\mu(G/H;\mathbb{C})$. To do this it is necessary to understand what is the restriction of the representation $M_\rho\colon G \to \mathbf{U}(\mathbb{C}^{n_\rho})$ to H, with $\rho \in R(G)$.

We will denote the complete set of the irreducible representations of G given by the Peter-Weyl theorem I (Theorem 13.2) by $\rho \in R(G)$, the corresponding representations by $M_\rho\colon G \to \mathbf{U}(\mathbb{C}^{n_\rho})$, and the identity element of $\mathfrak{a}_\rho$ by $u_\rho = \frac{1}{n_\rho}\chi_\rho$, where χ_ρ is the character associated with ρ. Similarly, we will denote the complete set of irreducible representations of H given by the Peter-Weyl theorem I by $\sigma \in R(H)$, the corresponding representations by $M_\sigma\colon H \to \mathbf{U}(\mathbb{C}^{n_\sigma})$, and the identity element of $\mathfrak{a}_\sigma$ by $u_\sigma = \frac{1}{n_\sigma}\chi_\sigma$, where χ_σ is the character associated with σ. The Haar measure on G is denoted by λ_G, and the Haar measure on H is denoted by λ_H.

Consider the restriction $V\colon H \to \mathbf{U}(\mathbb{C}^{n_\rho})$ of the representation $M_\rho\colon G \to \mathbf{U}(\mathbb{C}^{n_\rho})$ to H. Recall that for any function $f \in \mathrm{L}^2(H)$, $V_{\mathrm{ext}}(f)$ is the weak integral of the function $t \mapsto V(t)(x)$ with respect to $f d\lambda_H$ $(t \in H)$. We will write $M_\rho(f)$ for $V_{\mathrm{ext}}(f)$. By the Peter–Weyl theorem II (Theorem 13.16), for every $\sigma \in R(H)$, the map

$$M_\rho(u_{\overline{\sigma}}) = \frac{1}{n_\sigma}\int_H M_\rho(t)\overline{\chi_\sigma(t)}\, d\lambda_H(t)$$

is the orthogonal projection of $\mathbb{C}^{n_\rho}$ onto a closed subspace E_σ of $\mathbb{C}^{n_\rho}$, and we have a Hilbert sum

$$\mathbb{C}^{n_\rho} = \bigoplus_{\sigma\in R(H)} E_\sigma.$$

Recall from Section 13.4 that the integral defining $M_\rho(u_{\overline{\sigma}})$ can be computed by integrating the matrix $M_\rho(t)\overline{\chi_\sigma(t)}$ term by term. Furthermore, for each subspace $E_\sigma \neq (0)$, each irreducible representation M_σ of H is contained a certain number of times in the restriction of M_ρ to H, which we denote $d_\sigma = (\rho : \sigma)$, so E_σ is a finite Hilbert sum

$$E_\sigma = \bigoplus_{k=1}^{d_\sigma} F^\sigma_k,$$

of subspaces $F^\sigma_1, F^\sigma_2, \ldots, F^\sigma_{d_\sigma}$ of dimension n_σ, invariant under $M_\rho(t)$ for every $t \in H$, and such that the restriction of M_ρ to H and to each F^σ_k is equivalent to the irreducible representation M_σ. Thus E_σ has dimension $p_\sigma = d_\sigma n_\sigma$.

We can pick an orthonormal basis of $\mathbb{C}^{n_\rho}$ consisting of the union of orthonormal bases of each of the F^σ_j and of a basis of the orthogonal complement F' of E_σ in $\mathbb{C}^{n_\rho}$. Let P

be the change of basis matrix, which is unitary. For the basis of E_σ consisting of the first $p_\sigma = d_\sigma n_\sigma$ vectors of this basis, the matrix $M_{\rho,\sigma}(t)$ of the restriction of $P^* M_\rho(t) P$ to E_σ is a block diagonal matrix (consisting of d_σ blocks) of the form

$$M_{\rho,\sigma}(t) = \begin{pmatrix} M_\sigma(t) & 0 & \cdots & 0 \\ 0 & M_\sigma(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_\sigma(t) \end{pmatrix}$$

for every $t \in H$.

The automorphism $M_\rho(s)$ of $\mathbb{C}^{n_\rho} = E_\sigma \oplus F'$ is defined by four linear maps $P_{\rho,\sigma}(s)\colon E_\sigma \to E_\sigma$, $M_2(s)\colon F' \to E_\sigma$, $M_3(s)\colon E_\sigma \to F'$, and $M_4(s)\colon F' \to F'$, such that for any $(u, v) \in E_\sigma \times F'$ we have

$$M_\rho(s) \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} P_{\rho,\sigma}(s) & M_2(s) \\ M_3(s) & M_4(s) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} P_{\rho,\sigma}(s)u + M_2(s)v \\ M_3(s)u + M_4(s)v \end{pmatrix}. \tag{M1}$$

Since $M_\rho(u_{\overline{\sigma}})$ is the orthogonal projection of $\mathbb{C}^{n_\rho} = E_\sigma \oplus F'$ onto E_σ, the endomorphism $M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}})$ of $\mathbb{C}^{n_\rho}$ $(s \in G)$ is defined by

$$M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}}) \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} P_{\rho,\sigma}(s)u \\ 0 \end{pmatrix},$$

and so we can write

$$M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}}) = \begin{pmatrix} P_{\rho,\sigma}(s) & 0 \\ 0 & 0 \end{pmatrix}. \tag{M2}$$

In the new orthonormal basis of $E_\sigma \oplus F' = \mathbb{C}^{n_\rho}$, the matrix of the endomorphism $M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}})$ is the block matrix

$$P^* M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}}) P = \begin{pmatrix} Q(s) & 0 \\ 0 & 0 \end{pmatrix},$$

with $Q(s)$ the $p_\sigma \times p_\sigma$ matrix

$$Q(s) = \frac{1}{n_\rho} \begin{pmatrix} m_{11}^{(\rho,\sigma)}(s) & m_{12}^{(\rho,\sigma)}(s) & \ldots & m_{1p_\sigma}^{(\rho,\sigma)}(s) \\ m_{21}^{(\rho,\sigma)}(s) & m_{22}^{(\rho,\sigma)}(s) & \ldots & m_{2p_\sigma}^{(\rho,\sigma)}(s) \\ \vdots & \vdots & \ddots & \vdots \\ m_{p_\sigma 1}^{(\rho,\sigma)}(s) & m_{p_\sigma 2}^{(\rho,\sigma)}(s) & \ldots & m_{p_\sigma p_\sigma}^{(\rho,\sigma)}(s) \end{pmatrix}.$$

Note that since we have made a change of basis, the entries $m_{ij}^{(\rho,\sigma)}(s)$ are *not equal* to the original entries $m_{ij}^{(\rho)}(s)$ occurring in $M_\rho(s)$.

For any $t \in H$, the matrix $Q(t)$ of $P_{\rho,\sigma}(t)$ is the block diagonal matrix $M_{\rho,\sigma}(t)$, because the subspaces F_k^σ are invariant under $M_\rho(t)$ for $t \in H$. Since E_σ is invariant under $M_\rho(t)$ for $t \in H$, by (M1), for any $(u, v) \in E_\sigma \times F'$,

$$M_\rho(t) M_\rho(u_{\overline{\sigma}}) \begin{pmatrix} u \\ v \end{pmatrix} = M_\rho(t) \begin{pmatrix} u \\ 0 \end{pmatrix} = \begin{pmatrix} P_{\rho,\sigma}(t) u \\ 0 \end{pmatrix}$$

and

$$M_\rho(u_{\overline{\sigma}}) M_\rho(t) \begin{pmatrix} u \\ v \end{pmatrix} = M_\rho(u_{\overline{\sigma}}) \begin{pmatrix} P_{\rho,\sigma}(t) u \\ M_3(t) u + M_4(t) v \end{pmatrix} = \begin{pmatrix} P_{\rho,\sigma}(t) u \\ 0 \end{pmatrix},$$

so

$$M_\rho(t) M_\rho(u_{\overline{\sigma}}) = M_\rho(u_{\overline{\sigma}}) M_\rho(t) = \begin{pmatrix} P_{\rho,\sigma}(t) & 0 \\ 0 & 0 \end{pmatrix}, \tag{M3}$$

and since

$$M_\rho(tst') = M_\rho(t) M_\rho(s) M_\rho(t'),$$

by (M3) and (M2) we obtain the equation

$$\begin{aligned} M_\rho(u_{\overline{\sigma}}) M_\rho(tst') M_\rho(u_{\overline{\sigma}}) &= M_\rho(u_{\overline{\sigma}}) M_\rho(t) M_\rho(s) M_\rho(t') M_\rho(u_{\overline{\sigma}}) \\ &= M_\rho(t) M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}}) M_\rho(t') \\ &= M_\rho(t) M_\rho(u_{\overline{\sigma}}) M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}}) M_\rho(u_{\overline{\sigma}}) M_\rho(t') \\ &= \begin{pmatrix} P_{\rho,\sigma}(t) & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} P_{\rho,\sigma}(s) & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} P_{\rho,\sigma}(t') & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} P_{\rho,\sigma}(t) P_{\rho,\sigma}(s) P_{\rho,\sigma}(t') & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

and since by (M2) we have

$$M_\rho(u_{\overline{\sigma}}) M_\rho(tst') M_\rho(u_{\overline{\sigma}}) = \begin{pmatrix} P_{\rho,\sigma}(tst') & 0 \\ 0 & 0 \end{pmatrix},$$

we obtain the equation

$$P_{\rho,\sigma}(tst') = P_{\rho,\sigma}(t) P_{\rho,\sigma}(s) P_{\rho,\sigma}(t'), \qquad \text{for all } s \in G \text{ and all } t, t' \in H. \tag{$*$}$$

Since $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ and $M_\rho(u_{\overline{\sigma}}) M_\rho(u_{\overline{\sigma}}) = M_\rho(u_{\overline{\sigma}})$ since $M_\rho(u_{\overline{\sigma}})$ is a projection, we have

$$\mathrm{tr}(M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}})) = \mathrm{tr}(M_\rho(u_{\overline{\sigma}}) M_\rho(u_{\overline{\sigma}}) M_\rho(s)) = \mathrm{tr}(M_\rho(u_{\overline{\sigma}}) M_\rho(s)).$$

Definition 15.16. The *partial trace* of ρ with respect to σ is the function

$$s \mapsto \theta_{\rho,\sigma}(s) = \mathrm{tr}(M_\rho(u_{\overline{\sigma}}) M_\rho(s) M_\rho(u_{\overline{\sigma}})) = \mathrm{tr}(M_\rho(u_{\overline{\sigma}}) M_\rho(s)),$$

which can also be expressed as

$$\theta_{\rho,\sigma}(s) = \frac{1}{n_\rho} (m_{11}^{(\rho,\sigma)}(s) + \cdots + m_{p_\sigma p_\sigma}^{(\rho,\sigma)}(s)).$$

The function $\theta_{\rho,\sigma}$ is continuous, no longer central, and not identically zero if σ is contained in the restriction of M_ρ to H. This function depends on ρ and σ, and we have

$$\begin{aligned} \theta_{\rho,\sigma}(t) &= (\rho : \sigma)\chi_\sigma(t) && \text{for all } t \in H \\ \chi_\rho(s) &= \sum_\sigma \theta_{\rho,\sigma}(s) && \text{for all } s \in G. \end{aligned}$$

It will be shown in Section 17.1 (see Example 17.5) that the partial traces for which $p = 1$ are the spherical functions when (G, H) is a Gelfand pair.

We have the following proposition.

Proposition 15.17. *The following properties hold.*

(1) We have

$$\theta_{\rho,\sigma}(tst^{-1}) = \theta_{\rho,\sigma}(s), \qquad \textit{for all } s \in G \textit{ and all } t \in H.$$

(2) When ρ ranges over $R(G)$ and σ ranges over $R(H)$, the partial traces $\theta_{\rho,\sigma}$ are pairwise orthogonal. In particular, $\theta_{\rho,\sigma}$ and $\theta_{\rho',\sigma'}$ can only be proportional if $\rho' = \rho$ and $\sigma' = \sigma$.

(3) The partial traces $\theta_{\rho,\sigma}$ are continuous functions of positive type.

Proof. (1) This equation follows immediately from $(*)$ and the commutativity of the trace.

(2) This follows from the equation

$$\theta_{\rho,\sigma}(s) = \frac{1}{n_\rho}(m_{11}^{(\rho,\sigma)}(s) + \cdots + m_{p_\sigma p_\sigma}^{(\rho,\sigma)}(s))$$

and the orthogonality properties of the $m_{ij}^{(\rho,\sigma)}$; see Proposition 13.9.

(3) This follows from the properties $m_{ii}^{(\rho,\sigma)} = \overline{\check{m}}_{ii}^{(\rho,\sigma)} = m_{ii}^{(\rho,\sigma)} * m_{ii}^{(\rho,\sigma)}$ of Proposition 13.9, the fact that $f * \overline{\check{f}}$ is of positive type for every $f \in \mathcal{L}^2(G)$, and the equation

$$\theta_{\rho,\sigma}(s) = \frac{1}{n_\rho}(m_{11}^{(\rho,\sigma)}(s) + \cdots + m_{p_\sigma p_\sigma}^{(\rho,\sigma)}(s)). \qquad \square$$

Since G and H are compact, G/H has a G-invariant measure μ induced by a Haar measure on G. We now try to understand what the canonical unitary representation of G in $\mathrm{L}^2_\mu(G/H)$ induced by the trivial representation of H in $E = \mathbb{C}$ looks like. With the notations as above, we have $n_{\sigma_0} = 1$, and $p_{\sigma_0} = d$.

First, let us observe that a function $g \in \mathcal{L}^2_\mu(G/H)$ can be viewed as a function $g \in \mathcal{L}^2(G)$ such that

$$g(st) = g(s) \qquad \text{for all } t \in H \text{ and all } s \in G. \tag{$*_{G/H}$}$$

Since $(g * \delta_t)(s) = g(st)$, the above condition is equivalent to

$$g * \delta_t = g \qquad \text{for all } t \in H, \tag{$*'_{G/H}$}$$

and thus for any measure $\nu \in \mathcal{M}^1(G)$, the function $\nu * g \in \mathcal{L}^2_\mu(G/H)$ also satisfies the equation

$$(\nu * g) * \delta_t = \nu * g,$$

so we deduce that $\mathrm{L}^2_\mu(G/H)$ is a closed left ideal in $\mathcal{M}^1(G)$, which implies that $\mathrm{L}^2_\mu(G/H)$ is a closed left ideal in $\mathrm{L}^2(G)$. In particular, for every $\rho \in R(G)$, the projection $g \mapsto u_\rho * g$ of $\mathrm{L}^2(G)$ onto the ideal $\mathfrak{a}_\rho$ maps $\mathrm{L}^2_\mu(G/H)$ onto itself, so $\mathrm{L}^2_\mu(G/H)$ is the Hilbert sum of the subspaces

$$L_\rho = \mathrm{L}^2_\mu(G/H) \cap \mathfrak{a}_\rho.$$

It remains to determine what the L_ρ are. We explained that by applying Peter–Weyl II (Theorem 13.16) to the restriction of the representation $M_\rho \colon G \to \mathbf{U}(C^{n_\rho})$ to H we obtain a decomposition of $\mathbb{C}^{n_\rho}$ as a finite Hilbert sum

$$\mathbb{C}^{n_\rho} = E_{\sigma_1} \oplus \cdots \oplus E_{\sigma_q},$$

with each E_{σ_i} a direct sum

$$E_{\sigma_i} = \bigoplus_{k=1}^{d_{\sigma_i}} F_k^{\sigma_i}$$

of subspaces $F_1^{\sigma_i}, F_2^{\sigma_i}, \ldots, F_{d_{\sigma_i}}^{\sigma_i}$ of dimension n_{σ_i}, invariant under $M_\rho(t)$ for every $t \in H$, and such that the restriction of M_ρ to each $F_k^{\sigma_i}$ is equivalent to the irreducible representation M_{σ_i}. Let us pick for an orthonormal basis of $\mathbb{C}^{n_\rho}$ the union of orthonormal bases of the $F_k^{\sigma_i}$, and let P be the change of basis matrix, which is unitary. Then for any $t \in H$ we have

$$P^* M_\rho(t) P = \begin{pmatrix} M_{\rho,\sigma_1}(t) & 0 & \cdots & 0 \\ 0 & M_{\rho,\sigma_2}(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{\rho,\sigma_q}(t) \end{pmatrix},$$

where $M_{\rho,\sigma_i}(t)$ is the block matrix

$$M_{\rho,\sigma_i}(t) = \begin{pmatrix} M_{\sigma_i}(t) & 0 & \cdots & 0 \\ 0 & M_{\sigma_i}(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{\sigma_i}(t) \end{pmatrix}$$

(consisting of d_{σ_i} blocks) defined earlier. Thus the matrix $M_\rho^{(H)}(t) = P^* M_\rho(t) P$ (with $t \in H$) is the block matrix consisting of the blocks $M_{\sigma_i}(t)$, each one repeated d_{σ_i} times. We also define the matrices $M_\rho^{(H)}(s) = (m_{ij}^{(\rho,H)}(s))$ for *all* $s \in G$ by

$$M_\rho^{(H)}(s) = P^* M_\rho(s) P, \quad s \in G.$$

The representations of G in $\mathbb{C}^{n_\rho}$ defined by the matrices $M_\rho(s)$ and $M_\rho^{(H)}(s)$ $(s \in G)$ are equivalent. The matrix $M_\rho^{(H)}$ denotes the matrix of n_ρ^2 functions $m_{ij}^{(\rho,H)}$ given by $s \mapsto$

$m_{ij}^{(\rho,H)}(s)$ and we also write $M_\rho^{(H)} = P^* M_\rho P$. By Proposition 13.9, the matrix $M_\rho^{(H)}$ defines n_ρ^2 functions $m_{ij}^{(\rho,H)}$ that form an orthonormal basis of $\mathfrak{a}_\rho$ and satisfy the same properties as the functions $m_{i,j}^{(\rho)}$ defined by the matrix M_ρ.

Proposition 15.18. *The space $\mathrm{L}^2_\mu(G/H)$ is the Hilbert sum of subspaces $L_\rho \subseteq \mathfrak{a}_\rho$. If the trivial representation σ_0 of H is contained $d = (\rho : \sigma_0) \geq 1$ times in the restriction of M_ρ to H, then L_ρ is the direct sum of the first d columns of $M_\rho^{(H)} = P^* M_\rho P$,*

$$L_\rho = \bigoplus_{j=1}^{d} \mathfrak{l}_j^{(\rho,H)} \quad \text{and} \quad \mathfrak{l}_j^{(\rho,H)} = \bigoplus_{k=1}^{n_\rho} \mathbb{C} m_{kj}^{(\rho,H)}.$$

If $d = 0$, then $L_\rho = (0)$. The subrepresentation $\Pi\colon G \to \mathbf{U}(L_\rho)$ in L_ρ of the canonical representation $\Pi\colon G \to \mathbf{U}(\mathrm{L}^2_\mu(G/H))$ of G in $\mathrm{L}^2_\mu(G/H)$ induced by the trivial representation of H in $\mathbb{C}$ is the Hilbert sum of $d = (\rho : \sigma_0)$ irreducible representations equivalent to $M_{\overline{\rho}}$.

Proof. Since any function $g \in \mathrm{L}^2(G/H) \cap \mathfrak{a}_\rho$ can be written as $g = \sum_{ij} c_{ij} m_{ij}^{(\rho,H)}$ (for some $c_{ij} \in \mathbb{C}$), the equation $(*_{G/H})$ and Proposition 13.9(4) yields

$$\sum_{i,j,k} c_{ij} m_{ik}^{(\rho,H)}(s) m_{kj}^{(\rho,H)}(t) = \sum_{i,j} c_{ij} m_{ij}^{(\rho,H)}(s),$$

with $1 \leq i, j, k \leq n_\rho$, all $s \in G$ and all $t \in H$, and since the n_ρ^2 functions $m_{ij}^{(\rho,H)}$ are linearly independent, we get

$$\sum_{j=1}^{n_\rho} c_{ij} m_{kj}^{(\rho,H)}(t) = c_{ik}, \tag{$*_1$}$$

for all i, k with $1 \leq i, k \leq n_\rho$ and for all $t \in H$.

Suppose that the trivial representation σ_0 of H is contained $d \geq 1$ times in the restriction of M_ρ to H, which means that the first d matrices $M_{\sigma_h}(t)$ are just one-dimensional matrices equal to 1, the other matrices being at least two-dimensional. We then have $m_{kj}^{(\rho,H)}(t) = \delta_{kj}$ for $k \leq d$ or $j \leq d$, hence $(*_1)$ is trivially verified for $k \leq d$, and we are left with the equations

$$\sum_{j=d+1}^{n_\rho} c_{ij} m_{kj}^{(\rho,H)}(t) = c_{ik}, \qquad k > d \text{ and } t \in H. \tag{$*_2$}$$

Consider one of the matrices $M_{\sigma_h}(t)$ and assume it corresponds to the lines of index k such that $k_h' \leq k \leq k_h''$. Then we have $m_{kj}^{(\rho,H)}(t) = 0$ for $k_h' \leq k \leq k_h''$ and all j except those for which $k_h' \leq j \leq k_h''$; in addition, since $\sigma \neq \sigma_0$, by the fact stated just after Definition 13.3, we have

$$\int_H m_{kj}^{(\rho,H)}(t)\, d\lambda_H(t) = 0$$

for all these indices. Integrating both sides of $(*_2)$, we see that $c_{ik} = 0$ for all indices i and all $k > d$.

Therefore L_ρ is the subspace of $\mathfrak{a}_\rho$, of dimension dn_ρ, spanned by the $m_{ij}^{(\rho,H)}$ such that $j \leq d$, equivalently, the *direct sum of the first d columns of $M_\rho^{(H)}$*,

$$L_\rho = \bigoplus_{j=1}^{d} \mathfrak{l}_j^{(\rho,H)} \quad \text{and} \quad \mathfrak{l}_j^{(\rho,H)} = \bigoplus_{k=1}^{n_\rho} \mathbb{C} m_{kj}^{(\rho,H)}.$$

If $d = 0$, then the above reasoning shows that $L_\rho = (0)$.

The canonical representation $\Pi\colon G \to \mathbf{U}(\mathrm{L}^2_\mu(G/H))$ of G in $\mathrm{L}^2_\mu(G/H)$ induced by the trivial representation of H in $\mathbb{C}$ is a subrepresentation of the regular representation of G. We know from the discussion just after Definition 13.7 that on $\mathfrak{a}_\rho$, the regular representation $\mathbf{R}$ splits into n_ρ irreducible representations all equivalent to $M_{\overline{\rho}}$, and we can view these representation as acting on the columns of M_ρ, the left ideals $\mathfrak{l}_j^{(\rho)}$. Therefore, the subrepresentation in L_ρ of the canonical representation Π of G induced by the trivial representation of H in $\mathbb{C}$ is the Hilbert sum of $(\rho : \sigma_0)$ irreducible representations equivalent to $M_{\overline{\rho}}$. $\square$

Remark: It is possible to describe the unitary representations of G induced by the nontrivial irreducible representations M_{σ_i} of H; see Dieudonné's [22], Chapter XXII, Section 5, Problem 1.

We can also consider the space $H\backslash G$ of right cosets Hs of G ($s \in G$). If $\pi\colon G \to H\backslash G$ is the quotient map $\pi(s) = Hs$, the fact that the Haar measure λ on a compact group is left and right invariant implies immediately that there is a G-invariant measure μ' on $H\backslash G$ such that

$$\int_{G/H} g(x)\, d\mu'(x) = \int_G (g \circ \pi)\, d\lambda,$$

and

$$\int_{G/H} g(x \cdot s)\, d\mu'(x) = \int_{G/H} g(x)\, d\mu'(x) \qquad \text{for all } s \in G,$$

with

$$(Ht) \cdot s = Hts, \qquad s, t \in G.$$

Every function $g \in \mathcal{L}^2_{\mu'}(H\backslash G)$ can be viewed as a function $g \in \mathcal{L}^2(G)$ such that

$$g(ts) = g(s) \qquad \text{for all } t \in H \text{ and all } s \in G. \qquad (*_{H\backslash G})$$

Since $(\delta_t * g)(s) = g(t^{-1}s)$, the above condition is equivalent to

$$\delta_t * g = g \qquad \text{for all } t \in H. \qquad (*'_{H\backslash G})$$

The space $\mathrm{L}^2_{\mu'}(H\backslash G)$ is the image of the space $\mathrm{L}^2_\mu(G/H)$ under the isomorphism $g \mapsto \check{\overline{g}}$ (here we use the fact that G is unimodular). Therefore $\mathrm{L}^2_{\mu'}(H\backslash G)$ is a closed right ideal

in $\mathrm{L}^2(G)$, and it is the Hilbert sum of the images $\check{\overline{L}}_\rho$ of the L_ρ; since by Theorem 13.6(2) we have $m_{ji} = \check{\overline{m}}_{ij}$, we deduce that $\check{\overline{L}}_\rho$ is the direct sum of the first d rows of $M_\rho^{(H)}$ (with $d = (\rho : \sigma_0)$).

Let us record this fact.

Proposition 15.19. *The space $\mathrm{L}^2_{\mu'}(H\backslash G)$ is the Hilbert sum of subspaces $\check{\overline{L}}_\rho \subseteq \mathfrak{a}_\rho$. If the trivial representation σ_0 of H is contained $d = (\rho : \sigma_0) \geq 1$ times in the restriction of M_ρ to H, then $\check{\overline{L}}_\rho$ is the direct sum of the first d rows of $M_\rho^{(H)}$; that is,*

$$\check{\overline{L}}_\rho = \bigoplus_{i=1}^{d} \bigoplus_{j=1}^{n_\rho} \mathbb{C} m_{ij}^{(\rho,H)}.$$

Let us now consider the intersection $\mathrm{L}^2_\mu(G/H) \cap \mathrm{L}^2_{\mu'}(H\backslash G)$. This is a closed involutive subalgebra of $\mathrm{L}^2(G)$, thus a complete Hilbert algebra. We can view a function $g \in \mathrm{L}^2_\mu(G/H) \cap \mathrm{L}^2_{\mu'}(H\backslash G)$ as a function $g \in \mathrm{L}^2(G)$ such that

$$g(tst') = g(s) \qquad \text{for all } t, t' \in H \text{ and all } s \in G, \tag{$*_{H\backslash G/H}$}$$

or equivalently

$$\delta_t * g * \delta_{t'} = g \qquad \text{for all } t, t' \in H. \tag{$*'_{H\backslash G/H}$}$$

We can also think of the functions $g \in \mathrm{L}^2_\mu(G/H) \cap \mathrm{L}^2_{\mu'}(H\backslash G)$ as functions defined on the *double classes (or double cosets) HsH* of G with respect to H. In this case, If $\pi\colon G \to H\backslash G/H$ is the quotient map $\pi(s) = HsH$, the fact that the Haar measure λ on a compact group is left and right invariant implies that there is a G-invariant measure μ on $H\backslash G/H$ such that

$$\int_{H\backslash G/H} g(x)\, d\mu(x) = \int_G (g \circ \pi)\, d\lambda.$$

We denote the algebra of functions in $\mathrm{L}^2(G)$ satisfying $(*_{H\backslash G/H})$ as $\mathrm{L}^2_\mu(H\backslash G/H)$, or simply as $\mathrm{L}^2(H\backslash G/H)$. The following proposition follows immediately from the previous two propositions.

Proposition 15.20. *The algebra $\mathrm{L}^2(H\backslash G/H)$ is the Hilbert sum of the minimal two-sided ideals*

$$\mathfrak{a}_{\rho,\sigma_0} = L_\rho \cap \check{\overline{L}}_\rho = \bigoplus_{i=1}^{d} \bigoplus_{j=1}^{d} \mathbb{C} m_{ij}^{(\rho,H)}.$$

Each $\mathfrak{a}_{\rho,\sigma_0}$ is a matrix algebra of dimension d^2 having the family $(m_{ij}^{(\rho,H)})_{1\leq i,j\leq d}$ as a basis. The center of $\mathfrak{a}_{\rho,\sigma_0}$ is the one-dimensional subspace

$$\mathbb{C}(m_{11}^{(\rho,H)} + \cdots + m_{dd}^{(\rho,H)}) = \mathbb{C} n_\rho \theta_{\rho,\sigma_0},$$

and $u_{\rho,\sigma_0} = m_{11}^{(\rho,H)} + \cdots + m_{dd}^{(\rho,H)}$ is the unit of $\mathfrak{a}_{\rho,\sigma_0}$. The map $g \mapsto u_{\rho,\sigma_0} * g = g * u_{\rho,\sigma_0}$ is the orthogonal projection of $\mathrm{L}^2(H\backslash G/H)$ onto $\mathfrak{a}_{\rho,\sigma_0}$.

The subspace

$$\mathfrak{l}_{\sigma_0,1}^{(\rho,H)} = \mathfrak{l}_1^{(\rho,H)} \cap \mathfrak{a}_{\rho,\sigma_0} = \mathbb{C}m_{11}^{(\rho,H)} \oplus \cdots \oplus \mathbb{C}m_{d1}^{(\rho,H)}$$

is a minimal left ideal of $\mathrm{L}^2(H\backslash G/H)$. By Theorem 11.34, this ideal defines the irreducible representation $W_\rho \colon \mathrm{L}^2(H\backslash G/H) \to \mathcal{L}(\mathfrak{l}_{\sigma_0,1}^{(\rho,H)})$ of the algebra $\mathrm{L}^2(H\backslash G/H)$ in $\mathfrak{l}_{\sigma_0,1}^{(\rho,H)}$ (of dimension d), given by

$$(W_\rho(g))(f) = g * f, \qquad g \in \mathrm{L}^2(H\backslash G/H),\ f \in \mathfrak{l}_{\sigma_0,1}^{(\rho,H)}.$$

From Theorem 11.34(2), up to equivalence, we obtain all irreducible representations of $\mathrm{L}^2(H\backslash G/H)$ in $\mathfrak{l}_{\sigma_0,1}^{(\rho,H)}$ in this fashion. We can describe the representation W_ρ explicitly as follows. For every $g \in \mathrm{L}^2(H\backslash G/H)$, by Proposition 13.9 we can write

$$g * u_{\rho,\sigma_0} = \sum_{1\leq i,k\leq d} c_{ik}(g) m_{ik}^{(\rho,H)} \in \mathfrak{a}_{\rho,\sigma_0}$$

and the jth column of the matrix $W_\rho(g)$ consists of the coordinates of $W_\rho(g)(m_{j1}^{(\rho,H)}) = g*m_{j1}^{(\rho,H)}$ over the basis $(m_{11}^{(\rho,H)}, \ldots, m_{d1}^{(\rho,H)})$, and since u_{ρ,σ_0} is the unit of $\mathfrak{a}_{\rho,\sigma_0}$, by Proposition 13.9,

$$g * m_{1j}^{(\rho,H)} = g * u_{\rho,\sigma_0} * m_{1j}^{(\rho,H)} = \left(\sum_{1\leq i,k\leq d} c_{ik}(g) m_{ik}^{(\rho,H)}\right) * m_{j1}^{(\rho,H)} = \sum_{1\leq i\leq d} c_{ij}(g) m_{i1}^{(\rho,H)},$$

so $W_\rho(g) = (c_{ij}(g))$, a $d \times d$ matrix.

The above facts imply the following proposition.

Proposition 15.21. *The algebra $\mathrm{L}^2(H\backslash G/H)$ is commutative if and only if $(\rho : \sigma_0) \leq 1$ for all $\rho \in R(G)$. If so, then for every $\rho \in R(G)$ such that $(\rho : \sigma_0) = 1$, the ideal $\mathfrak{a}_{\rho,\sigma_0}$ is one-dimensional and is spanned by the function*

$$\omega_\rho(s) = \theta_{\rho,\sigma_0} = \frac{1}{n_\rho} m_{11}^{(\rho,H)}(s),$$

which is continuous and of positive type. Thus

$$\mathrm{L}^2(H\backslash G/H) = \bigoplus_{\rho \mid (\rho:\sigma_0)=1} \mathbb{C}\omega_\rho.$$

The orthogonal projection of $\mathrm{L}^2(H\backslash G/H)$ onto $\mathbb{C}\omega_\rho$ is given by

$$g \mapsto \omega_\rho * g = g * \omega_\rho.$$

The function ω_ρ also satisfies the following equations:

$$\begin{aligned} \omega_\rho(tst') &= \omega_\rho(s), \qquad \textit{for all } s \in G \textit{ and all } t, t' \in H \\ \omega_\rho(e) &= 1. \end{aligned}$$

The function ω_ρ is called a *(zonal) spherical function.*

The irreducible unitary representation W_ρ is one-dimensional, which implies that for every $g \in \mathcal{L}^2(H\backslash G/H)$ (since $g * \omega_\rho$ is continuous), we have

$$g * \omega_\rho = \zeta(g)\omega_\rho,$$

where ζ is must be a character of $\mathrm{L}^2(H\backslash G/H)$ with values in $\mathbb{T}$ (because $\omega_\rho * \omega_\rho = \omega_\rho$ and $g * \omega_\rho = \omega_\rho * g$). Since ζ is an algebra homomorphism and $\zeta(g) \in \mathbb{T}$, we conclude that ζ is a hermitian character of $\mathrm{L}^2(H\backslash G/H)$.

Since $(\rho : \sigma_0) = 1$, the left ideal L_ρ is equal to the ideal $\mathfrak{l}_1^{(\rho,H)}$, which by Proposition 13.9(5) is a minimal ideal in $\mathfrak{a}_\rho$, and by Proposition 13.4, it is spanned by the elements of the form $\lambda_s\omega_\rho = \delta_s * \omega_\rho$, for all $s \in G$.

15.11 Spherical Harmonics on S^n and $\mathrm{L}^2(S^n)$

A nice example of the above situation arises if $G = \mathbf{SO}(n+1)$ and $H = \mathbf{SO}(n)$. In this case, $G/H = \mathbf{SO}(n+1)/\mathbf{SO}(n) \simeq S^n$. By Proposition 15.18, the space $\mathrm{L}^2(\mathbf{SO}(n+1)/\mathbf{SO}(n)) \simeq \mathrm{L}^2(S^n)$ is the Hilbert sum of the subspaces $L_\rho \subseteq \mathfrak{a}_\rho$ for which $(\rho : \sigma_0) \geq 1$, where

$$\mathrm{L}^2(\mathbf{SO}(n+1)) = \bigoplus_\rho \mathfrak{a}_\rho$$

is the Hilbert sum given by Peter–Weyl I and where $d = (\rho : \sigma_0) \geq 1$ is the number of times that the trivial representation σ_0 of $\mathbf{SO}(n)$ is contained in the restriction of M_ρ to $\mathbf{SO}(n)$. Then L_ρ is the direct sum of the first d columns of $M_\rho^{(H)}$,

$$L_\rho = \bigoplus_{j=1}^{d} \mathfrak{l}_j^{(\rho,H)} \quad \text{and} \quad \mathfrak{l}_j^{(\rho,H)} = \bigoplus_{k=1}^{n_\rho} \mathbb{C} m_{kj}^{(\rho,H)}.$$

The subrepresentation $\Pi\colon \mathbf{SO}(n+1) \to \mathbf{U}(L_\rho)$ of the canonical representation (see Definition 15.13) $\Pi\colon \mathbf{SO}(n+1) \to \mathbf{U}(\mathrm{L}^2(S^n))$ of $\mathbf{SO}(n+1)$ in $\mathrm{L}^2(S^n)$ induced by the trivial representation of $\mathbf{SO}(n)$ in $\mathbb{C}$ is the Hilbert sum of $d = (\rho : \sigma_0)$ irreducible representations equivalent to $M_{\overline{\rho}}$. Recall (see (Ind_G) before Definition 15.13) that

$$(\Pi_Q(f))(x) = f(Q^{-1}x) = f(Q^\top x), \quad Q \in \mathbf{SO}(n+1),\ f \in \mathrm{L}^2(S^n),\ x \in S^n.$$

However, $(\mathbf{SO}(n+1), \mathbf{SO}(n))$ is one of examples of a Gelfand pair given in Section 17.6, Case 1. We need to exhibit $\mathbf{SO}(n)$ as a subgroup of the fixed points of an involution σ of $\mathbf{SO}(n+1)$. To do this, let $s\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be the reflection about the hyperplane $x_1 = 0$, which is given by

$$s(x_1, x_2, \ldots, x_{n+1}) = (-x_1, x_2, \ldots, x_{n+1}).$$

Obviously $s^{-1} = s$. Then let $\sigma\colon \mathbf{SO}(n+1) \to \mathbf{SO}(n+1)$ be the automorphism given by

$$\sigma(Q) = sQs, \quad Q \in \mathbf{SO}(n+1).$$

Since $s^2 = I$, we also have $\sigma^2 = \mathrm{id}$. In matrix form

$$\sigma(Q) = \begin{pmatrix} -1 & 0 \\ 0 & I_n \end{pmatrix} Q \begin{pmatrix} -1 & 0 \\ 0 & I_n \end{pmatrix}.$$

The groups $\mathbf{SO}(n+1)^\sigma$ of fixed points of σ are the rotations $Q \in \mathbf{SO}(n+1)$ such that

$$Q = \begin{pmatrix} -1 & 0 \\ 0 & I_n \end{pmatrix} Q \begin{pmatrix} -1 & 0 \\ 0 & I_n \end{pmatrix},$$

and if we write

$$Q = \begin{pmatrix} q_{11} & u \\ v & Q_1 \end{pmatrix},$$

we must have

$$\begin{aligned}\begin{pmatrix} q_{11} & u \\ v & Q_1 \end{pmatrix} &= \begin{pmatrix} -1 & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} q_{11} & u \\ v & Q_1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & I_n \end{pmatrix} \\ &= \begin{pmatrix} q_{11} & -u \\ -v & Q_1 \end{pmatrix},\end{aligned}$$

and so $u = v = 0$. Consequently, $\mathbf{SO}(n+1)^\sigma = S(\mathbf{O}(1) \times \mathbf{O}(n))$, with

$$S(\mathbf{O}(1) \times \mathbf{O}(n)) = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & Q_1 \end{pmatrix} \;\middle|\; Q_1 \in \mathbf{O}(n),\ \lambda \det(Q_1) = 1 \right\}.$$

The stabilizer of $e_1 = (1, 0, \ldots, 0)$ corresponds to $\lambda = +1$, and it is indeed isomorphic to $\mathbf{SO}(n)$.

Since $(\mathbf{SO}(n+1), \mathbf{SO}(n))$ is a Gelfand pair, we have $d = (\rho : \sigma_0) \leq 1$ for all ρ.

It can be shown that the L_ρ for which $(\rho : \sigma_0) = 1$ are exactly the spaces $\mathcal{H}_k^{\mathbb{C}}(S^n)$ of spherical harmonics on S^n; see Definition 14.1. Thus we have a Hilbert sum

$$\mathrm{L}^2(S^n) = \bigoplus_{k \geq 0} \mathcal{H}_k^{\mathbb{C}}(S^n).$$

We also obtain a decomposition of the regular representation $\Pi\colon \mathbf{SO}(n+1) \to \mathbf{U}(\mathrm{L}^2(S^n))$ into irreducible representations $\Pi_k\colon \mathbf{SO}(n+1) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(S^n))$ of $\mathbf{SO}(n+1)$ in the spaces $\mathcal{H}_k^{\mathbb{C}}(S^n)$ of spherical harmonics on S^n. The above facts are proven in Dieudonné [23] (Chapter XXIII, Section 38). A different proof is given in Gallier and Quaintance [39] (Chapter 7). One of the technical results used in these proofs is that

$$\mathcal{P}_k^{\mathbb{C}}(n) = \mathcal{H}_k^{\mathbb{C}}(n) \oplus \|x\|^2 \mathcal{H}_{k-2}^{\mathbb{C}}(n) \oplus \cdots \oplus \|x\|^{2j} \mathcal{H}_{k-2j}^{\mathbb{C}}(n) \oplus \cdots \oplus \|x\|^{2[k/2]} \mathcal{H}_{[k/2]}^{\mathbb{C}}(n),$$

with the understanding that only the first term occurs on the right-hand side when $k < 2$ (the spaces $\mathcal{P}_k^{\mathbb{C}}(n)$ and $\mathcal{H}_k^{\mathbb{C}}(n)$ are described in Definition 14.1).

It is shown in Vilenkin [101] (Chapter IX, Sections 2.10, 2.11) that the irreducible representations $\Pi_k \colon \mathbf{SO}(n+1) \to \mathbf{U}(\mathcal{H}_k^{\mathbb{C}}(S^n))$ are irreducible representations of class 1 relative to $\mathbf{SO}(n)$ (see Definition 15.12) and that they form a complete set of representations of class 1 of $\mathbf{SO}(n+1)$ relative to the subgroup $\mathbf{SO}(n)$; For $n = 2$, these are actually all the irreducible representations of $\mathbf{SO}(3)$ (see Proposition 14.3).

The space $\mathcal{H}_k^{\mathbb{C}}(S^n)$ is also the eigenspace associated to the eigenvalue $-k(n+k-1)$ of the Laplacian Δ_{S^n} on S^n. The unique zonal spherical function $\omega_\rho = \frac{1}{n_\rho} m_{11}^{(\rho,H)}$ in $\mathcal{H}_k^{\mathbb{C}}(S^n)$ is given in terms of Gegenbaur polynomials; see Gallier and Quaintance [39] (Chapter 7, Sections 3, 5, 6, 7).

Chapter 16

Constructing Induced Representations a la Mackey

One of the most important contributions to the theory of unitary representations is a method due to Mackey for constructing all irreducible representations of a locally compact group as induced irreducible representations from "small" subgroups H of G. This method is often referred to as the "Mackey machine." In its most general form the method is very complicated but in the case where G has an abelian normal subgroup N, it is tractable. The basic reason is that because N is abelian, its irreducible representations are given by the characters of N. There is also a natural action $\cdot\colon G \times \widehat{N} \to \widehat{N}$ of G on the dual group $\widehat{N}$ (the group of characters of N). The key to the construction is that because N is an abelian locally compact group, by Theorem 12.17, for any unitary representation $U\colon G \to \mathbf{U}(H_U)$ of G, since the restriction of U to N is a unitary representation, there is a unique regular projection-valued measure P on the dual group $\widehat{N}$ such that

$$U(n) = \int_{\widehat{N}} \chi(n)\, dP(\chi), \quad n \in N.$$

Observe that since H is a subgroup of G, we denote the space of the representation U by H_U to avoid confusion; see the comment following Definition 12.5. Moreover, the projection-valued measure P on $\widehat{N}$ satisfies two properties (see Proposition 16.1):

(1) We have

$$U(s)P(E)U(s)^{-1} = P(s \cdot E), \quad \text{for all Borel subsets } E \subseteq \widehat{N} \text{ and all } s \in G. \tag{imp}$$

(2) If U is irreducible, then for every G-invariant Borel set $E \subseteq \widehat{N}$ (which means that $\{s \cdot \chi \mid \chi \in E\} = s \cdot E = E$ for every $s \in G$), either $P(E) = I$ or $P(E) = 0$.

If the action of G on $\widehat{N}$ is nice enough (the space of orbits of this action is countably separated, see Definition 16.2), then P is identically zero except on a single orbit $\mathcal{O}_\nu$ so we can consider P as living on G/G_ν (where G_ν is the stabilizer of ν), and G acts transitively on this

space. Then the data (G, U, X, P) consisting of the unitary representation $U: G \to \mathbf{U}(H_U)$, of a transitive action of G on the homogeneous space $X = G/G_\nu$ (for some fixed $\nu \in \widehat{N}$), and of a regular projection-valued measure P on G/G_ν such that

$$U(s)P(E)U(s)^{-1} = P(s \cdot E), \quad \text{for all Borel sets } E \subseteq G/G_\nu \text{ and all } s \in G,$$

constitute a *transitive system of imprimitivity* (see Definition 16.3). Technically there are two equivalent ways of defining systems of imprimitivity but in this introduction we can ignore the second definition. The relevance of systems of imprimitivity is Mackey's *imprimitivity theorem* (Theorem 16.3), which implies that U is equivalent to a representation obtained by the induction method from some irreducible representation of G_ν. Technically, Mackey's *imprimitivity theorem* says more, namely that any transitive system of imprimitivity is equivalent to a system of imprimitivity arising by induction from the subgroup defining the homogeneous space X. If the action of G on $\widehat{N}$ is regular (see Definition 16.6), then Mackey's imprimitivity theorem implies Theorem 16.4, which is the result that shows that for every irreducible representation $U: G \to \mathbf{U}(H_U)$ of G, there is a unique orbit $\mathcal{O}$ such that for any $\nu \in \mathcal{O}$ (so that $\mathcal{O} = \mathcal{O}_\nu$), there is an irreducible unitary representation $\sigma: G_\nu \to \mathbf{U}(H_\sigma)$ such that U is equivalent to $\mathrm{Ind}_{G_\nu}^G \sigma$, the induced representation obtained from σ.

Unfortunately, the subgroups G_ν may still not be small enough. However, if for some $\nu \in \widehat{N}$ there is a continuous homomorphism $\widetilde{\nu}: G_\nu \to \mathbf{U}(1)$ extending ν, then for every irreducible representation $\rho: G_\nu/N \to \mathbf{U}(H_\rho)$ of G_ν/N, the map $\sigma: G_\nu \to \mathbf{U}(H_\rho)$ given by

$$\sigma(s) = \widetilde{\nu}(s)\rho(sN), \quad s \in G_\nu$$

defines an irreducible representation of G_ν in H_ρ (see Proposition 16.6). In this case we can use irreducible representations of the "little groups" $H_\nu = G_\nu/N$ in the inducing process of Theorem 16.4.

The above extension condition is satisfied by semi-direct products $G = N \rtimes H$, where N is a normal abelian subgroup of G. Then every irreducible representation of G is obtained in terms of the characters of N and of the irreducible representations of the little groups H_ν associated with the characters $\nu \in \widehat{N}$; see Theorem 16.7. Using this method we describe all irreducible representations of $\mathbf{SE}(n)$; see Example 16.1.

Historically the little group method was first used by Wigner in a famous paper (1939) on the representations of the *Poincaré group* $\mathbb{R}^4 \rtimes \mathbf{SO}_0(3,1)$, where $\mathbf{SO}_0(3,1)$ is the so-called *restricted Lorentz group*.

A thorough exposition of Mackey's method is given in Folland [33] (Chapter 6). A concise but very clear description of Mackey's method is also provided in Warner [103] (Chapter 5, Section 5.4). The reader interested in the history and the applications to physics (in particular quantum mechanics) of harmonic analysis should consult Mackey [66].

16.1 Introduction to the Mackey Machine

The reader may want to review the notion of projection-valued measure discussed in Section 11.11. Let G be a locally compact group and let N be a nontrivial closed *abelian normal* subgroup of G. The group G acts by conjugation on the normal subgroup N, namely for every $s \in G$, we let C_s be the automorphism of N given by $C_s(t) = sts^{-1}$ for all $t \in N$. Then the map $s \mapsto C_s$ is a homomorphism from G to $\mathrm{Aut}(N)$. Since G is a locally compact group and N is a closed abelian subgroup of G, the dual group $\widehat{N}$, namely the group of characters $\chi\colon N \to \mathbb{C}$ of N, is well-defined. But then we can define an action of G on $\widehat{N}$ as follows.

Definition 16.1. With G, N and $\widehat{N}$ as above we define an action $\cdot\colon G \times \widehat{N} \to \widehat{N}$ such that for all $s \in G$, $n \in N$, and $\chi \in \widehat{N}$,

$$(s \cdot \chi)(n) = \chi(s^{-1}ns). \tag{act}$$

To simplify notation we often denote $s \cdot \chi$ as $s\chi$.

Note that

$$\begin{aligned}((st) \cdot \chi)(n) &= \chi((st)^{-1}nst) \\ &= \chi(t^{-1}s^{-1}nst) \\ &= (t \cdot \chi)(s^{-1}ns) \\ &= (s \cdot (t \cdot \chi))(n),\end{aligned}$$

and obviously

$$(e \cdot \chi)(n) = \chi(e^{-1}ne) = \chi(n),$$

so $\cdot\colon G \times \widehat{N} \to \widehat{N}$ is indeed an action of G on $\widehat{N}$. Then, as usual, for every $\chi \in \widehat{N}$, we define the stabilizer G_χ of χ and the orbit $\mathcal{O}_\chi \subseteq \widehat{N}$ of χ as

$$\begin{aligned}G_\chi &= \{s \in G \mid s \cdot \chi = \chi\} \\ \mathcal{O}_\chi &= \{s \cdot \chi \mid s \in G\}.\end{aligned}$$

The subgroup G_χ is closed in G. Since N is abelian, we have $N \subseteq G_\chi$, and recall that there is a bijection between $\mathcal{O}_\chi$ and G/G_χ. The action $\cdot\colon G \times \widehat{N} \to \widehat{N}$ is never transitive (for instance, $\mathcal{O}_1 = \{1\}$) and the orbits can be complicated. What is remarkable is the fact that under certain conditions on the action of G on $\widehat{N}$, an irreducible unitary representation U of G arises from some irreducible representation ρ of G_ν for some $\nu \in \widehat{N}$ as an induced representation from G_ν to G.

The key to the construction is that because N is an abelian locally compact group, by Theorem 12.17, for any unitary representation $U\colon G \to \mathbf{U}(H)$ of G, since the restriction of U to N is a unitary representation, there is a unique regular projection-valued measure P on the dual group $\widehat{N}$ such that

$$U(n) = \int_{\widehat{N}} \chi(n)\, dP(\chi), \quad n \in N.$$

The crucial step is to figure out for every fixed $s \in G$ what are the projection-valued measures associated with the representations

$$n \mapsto U(s)U(n)U(s^{-1})$$

and

$$n \mapsto U(sns^{-1}).$$

Since U is a representation, $U(s)U(n)U(s^{-1}) = U(sns^{-1})$, so we obtain a condition on P.

Proposition 16.1. *Let G be a locally compact group and let N be a closed abelian normal subgroup of G. For any unitary representation $U\colon G \to \mathbf{U}(H)$ of G, let P be the unique regular projection-valued measure on $\widehat{N}$ such that for the restriction $U\colon N \to \mathbf{U}(H)$ of U to N we have*

$$U(n) = \int_{\widehat{N}} \chi(n)\, dP(\chi), \quad n \in N.$$

The following properties hold.

(1) The projection-valued measure P on $\widehat{N}$ satisfies the equation

$$U(s)P(E)U(s)^{-1} = P(s \cdot E), \quad \text{for all Borel subsets } E \subseteq \widehat{N} \text{ and all } s \in G. \qquad \text{(imp)}$$

(2) If U is irreducible, then for every G-invariant Borel set $E \subseteq \widehat{N}$ (which means that $\{s \cdot \chi \mid \chi \in E\} = s \cdot E = E$ for every $s \in G$), either $P(E) = I$ or $P(E) = 0$. We say that P is ergodic.

Proof. Since $U(s)$ is a unitary map for every $s \in G$ and since each $P(E)$ is a self-adjoint idempotent linear map it is immediately verified that $U(s)P(E)U(s)^{-1} = U(s)P(E)U(s)^*$ is also a self-adjoint idempotent linear map. It is not hard to check that for any fixed $s \in G$, the map Q defined on the Borel subsets of $\widehat{N}$ by

$$Q(E) = U(s)P(E)U(s)^{-1}$$

is a regular projection-valued measure (see Definition 11.20). For all $u, v \in H$, for all $n \in N$, since $U(s)$ is unitary we have

$$\langle U(s)U(n)U(s)^{-1}(u), v\rangle = \langle U(n)U(s)^{-1}(u), U(s)^{-1}(v)\rangle$$

so by definition

$$\langle U(n)U(s)^{-1}(u), U(s)^{-1}(v)\rangle = \int_{\widehat{N}} \chi(n)\, dP_{U(s)^{-1}(u),U(s)^{-1}(v)}(\chi).$$

But by definition, for any Borel set E in $\widehat{N}$,

$$\begin{aligned} P_{U(s)^{-1}(u),U(s)^{-1}(v)}(E) &= \langle P(E)U(s)^{-1}(u), U(s)^{-1}(v)\rangle \\ &= \langle U(s)P(E)U(s)^{-1}(u), v\rangle \\ &= \langle Q(E)(u), v\rangle = Q_{u,v}(E). \end{aligned}$$

so we deduce that

$$\langle U(s)U(n)U(s)^{-1}(u), v\rangle = \int_{\widehat{N}} \chi(n)\, dQ_{u,v}(\chi).$$

and thus

$$U(s)U(n)U(s)^{-1} = \int_{\widehat{N}} \chi(n)\, dQ(\chi). \tag{$*_1$}$$

We also have

$$\langle U(sns^{-1})(u), v\rangle = \int_{\widehat{N}} \chi(sns^{-1})\, dP_{u,v}(\chi).$$

Since by (act),

$$(s\cdot\chi)(n) = \chi(s^{-1}ns),$$

we obtain

$$\langle U(sns^{-1})(u), v\rangle = \int_{\widehat{N}} (s^{-1}\cdot\chi)(n)\, dP_{u,v}(\chi).$$

We now need to go back to Section 8.10. We have an action $\cdot\colon G\times\widehat{N}\to\widehat{N}$ and a σ-Radon measure μ on $\widehat{N}$ (which is locally compact). Recall from Definition 8.18 that for any $s\in G$ and any Borel subset E of $\widehat{N}$, we define $s\cdot E$ as

$$s\cdot E = \{s\cdot\chi \mid \chi\in E\},$$

for any function $f\colon\widehat{N}\to\mathbb{C}$, the function $\lambda_s(f)$ by

$$(\lambda_s(f))(\chi) = f(s^{-1}\cdot\chi),$$

and the measure $\lambda_s(\mu)$ by

$$(\lambda_s(\mu))(E) = \mu(s^{-1}\cdot E).$$

The proof of Proposition 8.16 is immediately adapted to show that for any $f\in \mathrm{L}^1(\widehat{N})$, we have

$$\int_{\widehat{N}} \lambda_s(f)\, d\mu = \int_{\widehat{N}} f\, d\lambda_{s^{-1}}(\mu),$$

which can also be written as

$$\int_{\widehat{N}} f(s^{-1}\cdot\chi)\, d\mu(\chi) = \int_{\widehat{N}} f(\chi)\, d(\lambda_{s^{-1}}(\mu))(\chi).$$

If we apply the above equation to the function f given by

$$f(\chi) = \chi(n)$$

for some fixed $n\in N$ and to the positive measure $P_{u,u}$, we obtain

$$\int_{\widehat{N}} (s^{-1}\cdot\chi)(n)\, dP_{u,u}(\chi) = \int_{\widehat{N}} \chi(n)\, d(\lambda_{s^{-1}}(P_{u,u}))(\chi).$$

Using the polarization method of Section 11.11, since

$$\int f\,dP_{u,v} = \frac{1}{4}\left(\int f\,dP_{u+v,u+v} - \int f\,dP_{u-v,u-v} + i\left(\int f\,dP_{u+iv,u+iv} - \int f\,dP_{u-iv,u-iv}\right)\right),$$

we obtain

$$\begin{aligned}\langle U(sns^{-1})(u), v\rangle &= \int_{\widehat{N}} (s^{-1}\cdot\chi)(n)\,dP_{u,v}(\chi)\\ &= \int_{\widehat{N}} \chi(n)\,d(\lambda_{s^{-1}}(P_{u,v}))(\chi)\end{aligned}$$

where $(\lambda_{s^{-1}}(P_{u,v}))(E) = P_{u,v}(s\cdot E)$, so

$$\langle U(sns^{-1})(u), v\rangle = \int_{\widehat{N}} \chi(n)\,d(\lambda_{s^{-1}}(P_{u,v}))(\chi). \tag{$*_2$}$$

Now by definition, for any Borel subset E of $\widehat{N}$,

$$P_{u,v}(s\cdot E) = \langle P(s\cdot E)(u), v\rangle,$$

so

$$(\lambda_{s^{-1}}(P_{u,v}))(E) = P_{u,v}(s\cdot E) = \langle P(s\cdot E)(u), v\rangle.$$

We can check quickly that the map $E \mapsto P(s\cdot E)$ is a regular projection-valued measure, since the map $E \mapsto s\cdot E$ is a bijection on Borel sets such that $s\cdot\widehat{N} = \widehat{N}$ and $s\cdot\emptyset = \emptyset$. Consequently, the map $\lambda_{s^{-1}}(P)$ given by $(\lambda_{s^{-1}}(P))(E) = P(s\cdot E)$ is a projection-valued measure, and by $(*_2)$, we have

$$U(sns^{-1}) = \int_{\widehat{N}} \chi(n)\,d(\lambda_{s^{-1}}(P))(\chi). \tag{$*_3$}$$

Since U is a representation, $U(s)U(n)U(s)^{-1} = U(sns^{-1})$, so by $(*_1)$ and $(*_3)$ we obtain

$$U(sns^{-1}) = \int_{\widehat{N}} \chi(n)\,dQ(\chi) = \int_{\widehat{N}} \chi(n)\,d(\lambda_{s^{-1}}(P))(\chi). \tag{$*_4$}$$

By uniqueness of the projection-valued measure defining a unitary representation, we conclude that

$$Q = \lambda_{s^{-1}}(P),$$

which more explicitly means that

$$U(s)P(E)U(s)^{-1} = P(s\cdot E), \quad \text{for all Borel sets } E \subseteq \widehat{N}.$$

If U is irreducible and if the Borel set E is G-invariant, that is, $s\cdot E = E$, then

$$U(s)P(E)U(s)^{-1} = P(E) \quad \text{for all } s \in G,$$

so $U(s)P(E) = P(E)U(s)$ for all $s \in G$, which means that $P(E) \in \mathcal{C}(U)$, where $\mathcal{C}(U)$ is the commutant of U (see Definition 12.6). By Schur's lemma, $P(E)$ is a scalar multiple of the identity, and since it is a projection, either $P(E) = I$ or $P(E) = 0$. $\square$

In summary, for any unitary representation $U\colon G \to \mathbf{U}(H)$ of G, there is some regular projection-valued measure P on $\widehat{N}$ such that

$$U(s)P(E)U(s)^{-1} = P(s \cdot E), \quad \text{for all Borel sets } E \subseteq \widehat{N} \text{ and all } s \in G,$$

and there is an action of the group G on $\widehat{N}$. These are just the ingredients that constitute Mackey's systems of imprimitivity! However, before defining systems of imprimitivity, we note that if the representation U is irreducible, it would be nice if P was identically zero except on a single orbit $\mathcal{O}_\nu$ (for some $\nu \in \widehat{N}$) because then we could consider P as living on G/G_ν, and G acts transitively on this space. Furthermore, in this case, Mackey's *imprimitivity theorem* applies, which implies that U is equivalent to a representation obtained by the induction method from some irreducible representation of G_ν. This is the essence of the Mackey machine for constructing induced representations.

Definition 16.2. Let G be a locally compact group and let N be a closed normal abelian subgroup of G. Consider the action of G on $\widehat{N}$ as in Definition 16.1. The space of orbits of this action is *countably separated* if there is a countable family (E_j) of G-invariant Borel subsets of $\widehat{N}$ such that for each orbit $\mathcal{O}$, we have

$$\mathcal{O} = \bigcap \{E_j \mid \mathcal{O} \subseteq E_j\};$$

in other words, each orbit is the intersection of the E_j that contain it.

Proposition 16.2. *If U is an irreducible unitary representation $U\colon G \to \mathbf{U}(H)$ and if the space of orbits of the action of Definition 16.1 is countably separated, then there is a single orbit $\mathcal{O} = \mathcal{O}_\nu$ in $\widehat{N}$ such that $P(\mathcal{O}_\nu) = I$.*

Proof. Let (E_j) be a countable family of G-invariant Borel subsets of $\widehat{N}$ with the property of Definition 16.2, so that for every orbit $\mathcal{O}$, there is some countable index set J such that

$$\mathcal{O} = \bigcap_{j \in J} E_j.$$

It follows that $P(\mathcal{O})$ is the projection onto the intersection of the ranges of the $P(E_j)$, with $j \in J$. Since U is irreducible, by Proposition 16.1, either $P(E_j) = I$ or $P(E_j) = 0$. Consequently, if $P(E_j) = I$ for all $j \in J$, then $P(\mathcal{O}) = I$, or else $P(\mathcal{O}) = 0$ if $P(E_j) = 0$ for some $j \in J$. We claim that there is some orbit $\mathcal{O}$ such that $P(\mathcal{O}) = I$. Otherwise, for every orbit $\mathcal{O}$ there is some index $j_{\mathcal{O}}$ such that $\mathcal{O} \subseteq E_{j_{\mathcal{O}}}$ and $P(E_{j_{\mathcal{O}}}) = 0$. Since $\widehat{N}$ is the union of the orbits, by Property (4) of Definition 11.20 we obtain $P(\widehat{N}) = 0$, which is absurd since $P(\widehat{N}) = 1$. Finally suppose that there are two disjoint orbits $\mathcal{O}_1$ and $\mathcal{O}_2$ such that $P(\mathcal{O}_1) = P(\mathcal{O}_2) = I$. But then by Property (3) of Definition 11.20,

$$I = P(\mathcal{O}_1) \circ P(\mathcal{O}_2) = P(\mathcal{O}_1 \cap \mathcal{O}_2) = P(\emptyset) = 0,$$

a contradiction. $\square$

If the action of G on $\widehat{N}$ is nice enough so that the space of orbits of this action is countably separated and if G/G_χ is homeomorphic to $\mathcal{O}_\chi$ for all $\chi \in \widehat{N}$, then the data consisting of the unitary representation $U\colon G \to \mathbf{U}(H)$, of a transitive action of G on the homogeneous space $X = G/G_\nu$ (for some fixed $\nu \in \widehat{N}$), and of a regular projection-valued measure P on G/G_ν such that

$$U(s)P(E)U(s)^{-1} = P(s \cdot E), \quad \text{for all Borel sets } E \subseteq G/G_\nu \text{ and all } s \in G,$$

constitute a *transitive system of imprimitivity*. Mackey's *imprimitivity theorem* applies to such a system, and this theorem is the key to defining irreducible representations obtained by the induced representation method. We now define (transitive) systems of imprimitivity and state Mackey's famous imprimitivity theorem.

16.2 Systems of Imprimitivity and the Imprimitivity Theorem

There are two equivalent ways of defining systems of imprimitivity. The first definition makes explicit use of a projection-valued measure and the second one uses a representation of the algebra $\mathcal{C}_0(S;\mathbb{C})$. The second definition is often technically easier to work with.

Definition 16.3. A *system of imprimitivity, version 1,* is a quadruple $\Sigma = (G, U, X, P)$, where

(1) G is a locally compact group.

(2) $U\colon G \to \mathbf{U}(H)$ is a unitary representation of G in a Hilbert space H.

(3) X is a G-space, which means that X is a locally compact Hausdorff space and there is a continuous action $\cdot\colon G \times X \to X$.

(4) P is projection-valued measure on X with values in $\mathcal{L}(H)$ satisfying the equation

$$U(s)P(E)U(s)^{-1} = P(s \cdot E), \quad \text{for all Borel subsets } E \subseteq X \text{ and all } s \in G. \qquad \text{(imp1)}$$

The system of imprimitivity $\Sigma = (G, U, X, P)$ is *transitive* if X is a homogeneous G-space. This means that $X = G/K$ for some closed subgroup K of G (with the action $g \cdot (sK) = (gs)K$, for all $g, s \in G$).

The projection-valued measure P on X determines a non-degenerate representation $V\colon \mathcal{C}_0(X;\mathbb{C}) \to \mathcal{L}(H)$ of the algebra $\mathcal{C}_0(X;\mathbb{C})$ defined by

$$V(f) = \int_X f\, dP, \quad f \in \mathcal{C}_0(X;\mathbb{C}).$$

As in the proof of Proposition 16.1, for all $u, v \in H$ and all $f \in \mathcal{C}_0(X;\mathbb{C})$, since $U(s)$ is unitary we have

$$\langle U(s)V(f)U(s)^{-1}(u), v\rangle = \langle V(f)U(s)^{-1}(u), U(s)^{-1}(v)\rangle$$

so by definition

$$\langle V(f)U(s)^{-1}(u), U(s)^{-1}(v)\rangle = \int_X f\, dP_{U(s)^{-1}(u),U(s)^{-1}(v)}.$$

But by definition, for any Borel set E in X,

$$\begin{aligned} P_{U(s)^{-1}(u),U(s)^{-1}(v)}(E) &= \langle P(E)U(s)^{-1}(u), U(s)^{-1}(v)\rangle \\ &= \langle U(s)P(E)U(s)^{-1}(u), v\rangle \\ &= \langle P(s\cdot E)(u), v\rangle && \text{by (imp1)} \\ &= P_{u,v}(s\cdot E) = \lambda_{s^{-1}}(P_{u,v})(E). \end{aligned}$$

Consequently

$$\langle U(s)V(f)U(s)^{-1}(u), v\rangle = \int_X f\, d\lambda_{s^{-1}}(P_{u,v}) = \int_X \lambda_s(f)\, dP_{u,v}.$$

The above equation says that

$$U(s)V(f)U(s)^{-1} = \int_X \lambda_s(f)\, dP,$$

which, by definition of V, means that

$$U(s)V(f)U(s)^{-1} = V(\lambda_s(f)), \quad f \in \mathcal{C}_0(X;\mathbb{C}),\ s \in G.$$

Conversely, if we have a nondegenerate representation $V\colon \mathcal{C}_0(X;\mathbb{C}) \to \mathcal{L}(H)$ satisfying the above equation, then by Theorem 11.59, there is a projection-valued measure P on X such that

$$V(f) = \int_X f\, dP, \quad f \in \mathcal{C}_0(X;\mathbb{C}).$$

Since the equation

$$U(s)V(f)U(s)^{-1} = V(\lambda_s(f)), \quad f \in \mathcal{C}_0(X;\mathbb{C}),\ s \in G$$

holds, a reasoning similar to the one used in the proof of Proposition 16.1 shows that Equation (imp1) holds. We are led to the following definition, which, by the above reasoning, is equivalent to Definition 16.3.

Definition 16.4. A *system of imprimitivity, version 2,* is a quadruple $\Sigma = (G, U, X, V)$, where

(1) G is a locally compact group.

(2) $U\colon G \to \mathbf{U}(H)$ is a unitary representation of G in a Hilbert space H.

(3) X is a G-space, which means that X is a locally compact Hausdorff space and there is a continuous action $\cdot\colon G \times X \to X$.

(4) V is a nondegenerate representation $V\colon \mathcal{C}_0(X;\mathbb{C}) \to \mathcal{L}(H)$ satisfying the equation

$$U(s)V(f)U(s)^{-1} = V(\lambda_s(f)), \quad f \in \mathcal{C}_0(X;\mathbb{C}),\ s \in G. \tag{imp2}$$

As before, the system of imprimitivity $\Sigma = (G, U, X, V)$ is *transitive* if X is a homogeneous G-space, $X = G/K$, for some closed subgroup K of G.

One of the main sources of systems of imprimitivity is from induced representations. In fact, we obtain transitive systems of imprimitivity.

Technically it is better to use Blattner's method for constructing an induced unitary representation $\Pi'\colon G \to \mathbf{U}(\mathcal{H}')$ of G from a unitary representation U of H, where H is a closed subgroup of G, as described in Section 15.8, because the definition of Π' in Formula (indv2) is simpler than Formula (indv1). But here we run into a notational clash. Since we denote the subgroup of G by H and since we use E to denote Borel sets, we should not denote the Hilbert space involved in the representation U of H by either H or E! Although this involves using more subscripts, we propose denoting the Hilbert space involved in the unitary representation U of H by H_U, so that our representation of H is written $U\colon H \to \mathbf{U}(H_U)$. Also, since we are using Blattner's construction instead of the first method from Section 15.8, we will drop the prime superscript and write Π instead of Π' and $\mathcal{H}$ instead of $\mathcal{H}'$. The Hilbert space $\mathcal{H}$ associated with the induced unitary representation $\Pi\colon G \to \mathbf{U}(\mathcal{H})$ of the representation $U\colon H \to \mathbf{U}(H_U)$ is the completion of a space $\mathcal{H}^0$ defined as

$$\begin{aligned}\mathcal{H}^0 = \Bigg\{ & f \in \mathcal{C}(G, H_U) \mid \pi(\mathrm{supp}(f)) \text{ is compact and} \\ & f(sh) = \left(\frac{\Delta_H(h)}{\Delta_G(h)}\right)^{1/2} U(h^{-1})(f(s)) \quad \text{for all } s \in G \text{ and all } h \in H \Bigg\}.\end{aligned}$$

Given a unitary representation $U\colon H \to \mathbf{U}(H_U)$, the induced unitary representation $\Pi\colon G \to \mathbf{U}(\mathcal{H})$, also denoted $\mathrm{Ind}_{H,\mathcal{H}}^G U$ or even $\mathrm{Ind}_H^G U$, is given by

$$(\Pi_s(f))(t) = f(s^{-1}t), \qquad f \in \mathcal{H},\ s, t \in G.$$

As usual, let $\pi\colon G \to G/H$ be the quotient map. A natural candidate for a projection-valued measure P^U on $X = G/H$ is to set

$$P^U(E)(f) = (\chi_E \circ \pi)(f), \quad E \subseteq G/H,\ f \in \mathcal{H}.$$

Here $P^U(E) \in \mathcal{L}(\mathcal{H})$ and $(\chi_E \circ \pi)(f)$ is the pointwise-multiplication of the functions $\chi_E \circ \pi$ and f, both defined on G. However it is not obvious that this definition makes sense and that Condition (imp1) is satisfied, so we circumvent these difficulties by using the definition of a system of imprimitivity given by Definition 16.4. We need to define a representation $V\colon \mathcal{C}_0(G/H, \mathbb{C}) \to \mathcal{L}(\mathcal{H})$ satisfying Condition (imp2).

If we take a close look at the definition of $\mathcal{H}^0$ in Section 15.8, we can check that for any $\varphi \in \mathcal{C}_0(G/H; \mathbb{C})$ and any $f \in \mathcal{H}^0$, since $f\colon G \to H_U$ and $\varphi \circ \pi\colon G \to \mathbb{C}$, the function $(\varphi \circ \pi)f$ from G to H_U given by

$$((\varphi \circ \pi)f)(s) = (\varphi \circ \pi)(s)f(s)$$

belongs to $\mathcal{H}^0$ and that

$$\|(\varphi \circ \pi)f\|_{\mathcal{H}} \leq \|\varphi\|_\infty \|f\|_{\mathcal{H}}.$$

As a consequence, since $\mathcal{H}^0$ is dense in $\mathcal{H}$, if we set

$$V(\varphi)(f) = (\varphi \circ \pi)f, \quad f \in \mathcal{H},$$

we obtain a representation $V\colon \mathcal{C}_0(G/H, \mathbb{C}) \to \mathcal{L}(\mathcal{H})$, and it is easy to see that V is nondegenerate. It remains to that prove that (indv2) hold (with respect to Π). For all $f \in \mathcal{H}^0$ and all $s, t \in G$ (recall that functions in $\mathcal{H}^0$ have domain G), we have

$$\begin{aligned}
((\Pi_s V(\varphi)\Pi_s^{-1})(f))(t) &= \Pi_s(V(\varphi)(\Pi_{s^{-1}}(f(t)))) \\
&= V(\varphi)(\Pi_{s^{-1}}(f(s^{-1}t))) && \text{by definition of } \Pi \\
&= \varphi(\pi(s^{-1}t))\Pi_{s^{-1}}(f(s^{-1}t)) && \text{by definition of } V \\
&= \varphi(\pi(s^{-1}t))f(t) && \text{by definition of } \Pi \\
&= \varphi(s^{-1} \cdot \pi(t))f(t) && \text{by definition of the action on } G/H \\
&= \lambda_s(\varphi)(\pi(t))f(t) \\
&= (V(\lambda_s(\varphi))(f))(t), && \text{by definition of } V
\end{aligned}$$

proving that

$$(\Pi_s V(\varphi)\Pi_s^{-1})(f) = V(\lambda_s(\varphi))(f)$$

for all $f \in \mathcal{H}^0$. Since $\mathcal{H}^0$ is dense in $\mathcal{H}$, we deduce that (ind2) holds, and so $(G, \Pi, G/H, V)$ is a transitive system of imprimitivity, version 2.

By Theorem 11.59, there is a unique projection-valued measure P^U on X such that

$$V(f) = \int_X f \, dP^U, \quad f \in \mathcal{C}_0(X; \mathbb{C}),$$

and $(G, \mathrm{Ind}_H^G U, G/H, P^U)$ is called the *canonical system of imprimitivity* associated to $\Pi = \mathrm{Ind}_H^G U$. It can be verified that

$$P^U(E)(f) = (\chi_E \circ \pi)(f), \quad E \subseteq G/H,\ f \in \mathcal{H},$$

as we said earlier, but using the representation V we verified that such a definition is legitimate.

Remark: The projection-valued measure arising from V is denoted by P^U. We prefer the notation P^U to the notation P^V even though P^U arises from V, because V is a representation in a Hilbert space $\mathcal{H}$ obtained as the completion of a space $\mathcal{H}^0$ which consists of certain functions from G to H_U, where H_U is the Hilbert space of the representation U.

The *raison d'être* for all this is that every transitive system of imprimitivity is equivalent to the canonical system of imprimitivity arising from some induced representation. This theorem originally due to Mackey is one of the greatest results in the theory of unitary representations. First we define the notion of equivalence of systems of imprimivity.

Definition 16.5. Two systems of imprimitivity $\Sigma = (G, U, X, P)$ and $\Sigma' = (G, U', X, P')$ (with the same group G and the same space X), where $U\colon G \to \mathbf{U}(H_U)$ and $U'\colon G \to \mathbf{U}(H_{U'})$ are two unitary representations, are *equivalent* if there is a unitary map $T\colon H_U \to H_{U'}$ such that

$$\begin{aligned} TU(s)T^{-1} &= U'(s) \quad \text{for all} \in G \\ TP(E)T^{-1} &= P'(E) \quad \text{for all Borel sets } E \subseteq X. \end{aligned}$$

We now state the celebrated imprimitivity theorem.

Theorem 16.3. *(Mackey's Imprimitivity Theorem, 1949-1953) Let G be a locally compact group and let H be a closed subgroup of G. Every transitive system of imprimitivity $\Sigma = (G, U, G/H, P)$, where $U\colon G \to \mathbf{U}(H_U)$ is a unitary representation of G, is equivalent to a transitive system of imprimitivity of the form $(G, \Pi, G/H, P^\sigma)$, where $\Pi = \mathrm{Ind}_H^G \sigma\colon G \to \mathbf{U}(\mathcal{H})$ is the representation induced by some some unitary representation $\sigma\colon H \to \mathbf{U}(H_\sigma)$ of H. Thus there is a unitary map $T\colon H_U \to \mathcal{H}$ such that*

$$\begin{aligned} TU(s)T^{-1} &= (\mathrm{Ind}_H^G \sigma)(s) \quad \textit{for all } s \in G \\ TP(E)T^{-1} &= P^\sigma(E) \quad \textit{for all Borel sets } E \subseteq G/X. \end{aligned}$$

Moreover, the unitary representation $\sigma\colon H \to \mathbf{U}(H_\sigma)$ is determined by Σ up to equivalence.

The proof of Theorem 16.3 is long and very technical. The version of the proof given in Folland [33] requires two sections (Sections 6.4 and 6.5) and stretches from Page 167 to Page 182. A key idea due to Blatter is to use an algebra $L(X \times G)$ and to extend a unitary representation of G to this algebra, by analogy with the method of extending a representation of G to a representation of $\mathrm{L}^1(G)$. If G is a Lie group, the proof is significantly simpler. A version of the imprimitivity theorem for Lie groups called the *Mackey Inducibility Criterion* by Kirillov is proven in Kirillov [54]; see Appendix V, Section 2.4. A sketch of proof for Lie groups is also given Taylor [96]; see Chapter V, Section 1. The good news is that we now have all the machinery needed to tacke the problem introduced in Section 16.1.

16.3 The Mackey Machine

Let G be a locally compact group and let N be a nontrivial closed *abelian normal* subgroup of G. As introduced in Definition 16.1, there is an action $\cdot\colon G \times \widehat{N} \to \widehat{N}$ such that for all $s \in G$, $n \in N$, and $\chi \in \widehat{N}$,

$$(s \cdot \chi)(n) = \chi(s^{-1}ns). \tag{act}$$

Recall that for every $\chi \in \widehat{N}$, we define the stabilizer G_χ of χ and the orbit $\mathcal{O}_\chi$ of χ as

$$\begin{aligned} G_\chi &= \{s \in G \mid s \cdot \chi = \chi\} \\ \mathcal{O}_\chi &= \{s \cdot \chi \mid s \in G\}. \end{aligned}$$

Our goal is to show that if the action of G on $\widehat{N}$ is nice enough, then every irreducible representation $U\colon G \to \mathbf{U}(H_U)$ of G arises as an induced representation of some irreducible representation σ of G_ν for some $\nu \in \widehat{N}$. The notion of nice action is formalized as follows.

Definition 16.6. We say the action $\cdot\colon G \times \widehat{N} \to \widehat{N}$ is *regular*, or that G *acts regularly on* $\widehat{N}$, if the following two conditions hold:

(1) The orbit space of our action is countably separated, as in Definition 16.2. Recall that this means that there is a countable family (E_j) of G-invariant Borel subsets of $\widehat{N}$ such that for each orbit $\mathcal{O}$, we have

$$\mathcal{O} = \bigcap \{E_j \mid \mathcal{O} \subseteq E_j\},$$

(2) For every $\nu \in \widehat{N}$, the map from G/G_ν to $\mathcal{O}_\nu$ given by $gG_\nu \mapsto g \cdot \nu$ is a homeomorphism.

Remark: When Condition (1) of Definition 16.6 holds, Kirillov called the orbit space *tame*; see Kirillov [54], Appendix V, Section 2.4.

Now given an irreducible representation $U\colon G \to \mathbf{U}(H_U)$ of G, the "miracle" is that $(G, U, G/G_\nu, P)$ is a transitive system of imprimitivity for some $\nu \in \widehat{N}$, where P is the projection-valued measure arising from Proposition 16.1 and where the unique orbit $\mathcal{O}_\nu$ exists by Proposition 16.2. Then the imprimitivity theorem applies and yields a unitary representation $\sigma\colon G_\nu \to H_\sigma$ such that U is equivalent to $\mathrm{Ind}_{G_\nu}^G \sigma\colon G \to \mathbf{U}(\mathcal{H})$. This yields most of the first main theorem of this section.

Theorem 16.4. *Let G be a locally compact group and let N be a nontrivial closed abelian normal subgroup of G. Suppose the action $\cdot\colon G \times \widehat{N} \to \widehat{N}$ is regular. For every irreducible representation $U\colon G \to \mathbf{U}(H_U)$ of G, there is a unique orbit $\mathcal{O}$ such that for any $\nu \in \mathcal{O}$ (so that $\mathcal{O} = \mathcal{O}_\nu$), there is an irreducible unitary representation $\sigma\colon G_\nu \to \mathbf{U}(H_\sigma)$ such that U is equivalent to $\mathrm{Ind}_{G_\nu}^G \sigma$. Moreover, we have*

$$\sigma(n) = \nu(n)\mathrm{id}_{H_\sigma}$$

for all $n \in N$.

Proof sketch. Since the unitary representation $U\colon G \to \mathbf{U}(H_U)$ is irreducible, by Proposition 16.1, the projection-valued measure P induced by the restriction of U to N has Property (indv1). Since the action of G on $\widehat{N}$ is regular, the orbit space is countably separated, so by Proposition 16.2, there is a single orbit $\mathcal{O}$ such that $P(\mathcal{O}) = I$, and so P is identically zero on $\widehat{N} - \mathcal{O}$. Pick any $\nu \in \widehat{N}$ such that $\mathcal{O} = \mathcal{O}_\nu$. The fact that the action of G on $\widehat{N}$ is regular implies that G/G_ν is homeomorphic to $\mathcal{O}_\nu$ and we may pull P back to G/G_ν. Now $(G, U, G/G_\nu, P)$ is a transitive system of imprimitivity so we can apply the imprimitivity theorem (Theorem 16.3) which tells us that $(G, U, G/G_\nu, P)$ is equivalent to a transitive system of imprimitivity of the form $(G, \mathrm{Ind}_{G_\nu}^G \sigma, G/G_\nu, P^\sigma)$, where $\mathrm{Ind}_{G_\nu}^G \sigma\colon G \to \mathbf{U}(\mathcal{H})$ is the representation induced by some some unitary representation $\sigma\colon G_\nu \to \mathbf{U}(H_\sigma)$ of G_ν. In particular, the unitary representations $U\colon G \to \mathbf{U}(H_U)$ and $\mathrm{Ind}_{G_\nu}^G \sigma\colon G \to \mathbf{U}(\mathcal{H})$ are equivalent, and since U is irreducible, so is $\mathrm{Ind}_{G_\nu}^G \sigma$. This also implies that σ is irreducible. The last part of the theorem is proven in Folland [33]; see Proposition 6.37. $\square$

The orbit $\mathcal{O}$ in Theorem 16.4 is unique, but ν may be chosen arbitrarily in $\mathcal{O}$. If ν' is another element of $\mathcal{O} = \mathcal{O}_\nu$, then $\nu' = s \cdot \nu$ for some $s \in G$ and the stabilizers G_ν and $G_{\nu'}$ are isomorphic; in fact, $G_{\nu'} = s \cdot G_\nu \cdot s^{-1}$; see Section C.3. But then, given any unitary representation $\sigma\colon G_\nu \to \mathbf{U}(H_\sigma)$ of G_ν we obtain the representation $\sigma'\colon G_{\nu'} \to \mathbf{U}(H_\sigma)$ of $G_{\nu'}$ given by $\sigma'(t) = \sigma(s^{-1}ts)$, for all $t \in G_\nu$, and this map is obviously bijective. We can also check that the unitary transformation $T\colon \mathcal{H} \to \mathcal{H}'$ given by $(Tf)(t) = f(s^{-1}ts)$ (where $f \in \mathcal{H}$) is an equivalence of the unitary representations $\mathrm{Ind}_{G_\nu}^G \sigma\colon G \to \mathbf{U}(\mathcal{H})$ and $\mathrm{Ind}_{G_{\nu'}}^G \sigma'\colon G \to \mathbf{U}(\mathcal{H}')$. Hence the choice of ν in $\mathcal{O}$ is not essential.

Theorem 16.4 has the following converse.

Theorem 16.5. *Let G be a locally compact group and let N be a nontrivial closed abelian normal subgroup of G. Suppose the action $\cdot\colon G \times \widehat{N} \to \widehat{N}$ is regular. For any $\nu \in \widehat{N}$ and for any irreducible unitary representation $\sigma\colon G_\nu \to \mathbf{U}(H_\sigma)$ of G_ν such that $\sigma(n) = \nu(n)\mathrm{id}_{H_\sigma}$ for all $n \in N$, the unitary representation $\mathrm{Ind}_{G_\nu}^G \sigma\colon G \to \mathbf{U}(\mathcal{H})$ is irreducible. If $\sigma'\colon G_\nu \to \mathbf{U}(H_{\sigma'})$ is another unitary representation of G_ν such that $\mathrm{Ind}_{G_\nu}^G \sigma\colon G \to \mathbf{U}(\mathcal{H})$ and $\mathrm{Ind}_{G_\nu}^G \sigma'\colon G \to \mathbf{U}(\mathcal{H}')$ are equivalent, then σ and σ' are equivalent.*

Theorem 16.5 is proven in Folland [33]; see Theorem 6.39.

Theoretically, Theorem 16.4 and Theorem 16.5 settle our problem, but in many cases these results are not useful because the groups G_ν may be rather large and their representations may not be much easier to analyze than the representations of G itself. For example, if $\nu = 1$ (the constant character with value 1), then $\mathcal{O}_\nu = \{1\}$ and $G_\nu = G$. In this case, Theorem 16.4 yields an irreducible representation $\sigma\colon G \to \mathbf{U}(H_\sigma)$ such that $\sigma(n) = \mathrm{id}_{H_\sigma}$ for all $n \in N$ equivalent to the orginal representation $U\colon G \to \mathbf{U}(H_U)$. Since σ is trivial on$N$ it follows that σ yields an irreducible representation $\rho\colon G/N \to \mathbf{U}(H_\sigma)$ of G/N, and σ is a lift of the representation ρ of the smaller group G/N to G, in the sense that $\rho = q \circ \sigma$ where $q\colon G \to G/N$ is the quotient map.

Recall that $N \subseteq G_\nu$. There are a number of examples where the character $\nu \in \widehat{N}$ can be extended "nicely" to a representation of G_ν. In this case we can lift an irreducible representation of the smaller group G_ν/N to G_ν.

Proposition 16.6. *Let G be a locally compact group and let N be a nontrivial closed abelian normal subgroup of G. Suppose that for some $\nu \in \widehat{N}$ there is a continuous homomorphism $\widetilde{\nu}\colon G_\nu \to \mathbb{T}$ extending ν. For every irreducible representation $\rho\colon G_\nu/N \to \mathbf{U}(H_\rho)$ of G_ν/N, the map $\sigma\colon G_\nu \to \mathbf{U}(H_\rho)$ given by*

$$\sigma(s) = \widetilde{\nu}(s)\rho(sN), \quad s \in G_\nu$$

defines an irreducible representation of G_ν in H_ρ such that $\sigma(n) = \nu(n)\mathrm{id}_{H_\rho}$ for all $n \in N$. Furthermore, every irreducible unitary representation σ of G_ν as above arises in this way.

Proposition 16.6 is proven in Folland [33]; see Proposition 6.40. The proof is simple but not illuminating. An orbit $\mathcal{O}$ such that that some character $\nu \in \mathcal{O}$ can be extended to a continuous homomorphism $\widetilde{\nu}\colon G_\nu \to \mathbb{T}$ is called *accommodating*; see Warner [103] (Section 5.4).

It turns out that an interesting class of groups to which Proposition 16.6 applies is the class of semi-direct products $N \rtimes H$ in which N is an abelian group. In this case, the groups G_ν/N are isomorphic to the groups $H_\nu = G_\nu \cap H$, called *little groups*. The Mackey machine yields *all* irreducible representations of $G = N \rtimes H$ as induced representations obtained by combining characters of N and irreducible representations of the little groups H_ν. The little group method was first used by Wigner in a famous paper (1939) on the representations of the *Poincaré group* $\mathbb{R}^4 \rtimes \mathbf{SO}_0(3,1)$, where $\mathbf{SO}_0(3,1)$ is the so-called *restricted Lorentz group*.

16.4 Irreducible Representations of Semi-Direct Products

For our purposes it is more convenient to adopt the "internal" view of a semi-direct product where a group G is already given as well as two subgroups N and H such that

(1) N is a normal subgroup of G.

(2) $G = NH$.

(3) $N \cap H = \{e\}$.

Then (2) and (3) imply that the map $N \times H \mapsto G$ given by $(n, h) \mapsto nh$ is a bijection. The multiplication operation in G is given by

$$(n_1h_1)(n_2h_2) = (n_1[h_1n_2h_1^{-1}])(h_1h_2), \quad n_1, n_2 \in N,\ h_1, h_2 \in H.$$

So H acts on N by conjugation on the left.

Since we also assume that G is locally compact, we require N and H to be closed, and that the map $N \times H \mapsto G$ given by $(n, h) \mapsto nh$ is a homeomorphism. The standard notation for a semi-direct product is $G = N \rtimes H$.[1] The multiplication operation in $G = N \rtimes H$ makes it clear that the map $q\colon N \rtimes H \to H$ given by $q(nh) = h$ $(n \in N, h \in H)$ is a surjective homomorphism with kernel N.

Now if N is also abelian, as before the group of characters $\widehat{N}$ (the dual group) makes sense. For any $\nu \in \widehat{N}$, since $N \subseteq G_\nu$, the quotient group G_ν/N is well-defined. Since $G = NH$, if we let $H_\nu = G_\nu \cap H$, we check immediately that

$$G_\nu = N \rtimes H_\nu.$$

Since G_ν/N is the group of cosets $sN = Ns$ (since N is normal) with $s \in G_\nu$, the map $Ns \mapsto s$ (with $s \in H_\nu$) is an isomorphism from G_ν/N to H_ν.

Definition 16.7. Given any semi-direct product $g = N \rtimes H$ (with N normal and abelian) as above, for any $\nu \in \widehat{N}$, the group H_ν given by

$$H_\nu = G_\nu \cap H$$

is called the *little group* associated with ν. As observed above,

$$G_\nu = N \rtimes H_\nu, \quad H_\nu \approx G_\nu/N.$$

The reason why little groups are interesting is that Proposition 16.6 applies. Indeed, given any character $\nu\colon N \to \mathbb{T}$, we can extend ν to a homomorphism $\widetilde{\nu}\colon G_\nu \to \mathbb{T}$ as follows:

$$\widetilde{\nu}(nh) = \nu(n), \quad h \in H_\nu,\, n \in N. \tag{$\widetilde{\nu}$}$$

We need to check that $\widetilde{v}$ is a homomorphism. First we have

$$\begin{aligned}
\widetilde{\nu}((n_1h_1)(n_2h_2)) &= \widetilde{\nu}((n_1[h_1n_2h_1^{-1}])(h_1h_2)) \\
&= \nu(n_1[h_1n_2h_1^{-1}]) \\
&= \nu(n_1)\nu(h_1n_2h_1^{-1}),
\end{aligned}$$

that is

$$\widetilde{\nu}((n_1h_1)(n_2h_2)) = \nu(n_1)\nu(h_1n_2h_1^{-1}). \tag{1}$$

However, by definition of the action of G on $\widehat{N}$ (see Definition 16.1), for any $\chi \in \widehat{N}$ and any $s \in G$ we have

$$(s \cdot \chi)(n) = \chi(s^{-1}ns),$$

so

$$\nu(h_1n_2h_1^{-1}) = (h_1^{-1} \cdot \nu)(n_2). \tag{2}$$

[1]Curiously Folland uses the notation $N \ltimes H$; see Folland [33].

But $h_1 \in H_\nu = G_\nu \cap H$ and since H_ν is a group, $h_1^{-1} \in H_\nu$, so as $h_1^{-1} \in H_\nu$ is a stabilizer of ν, we have

$$h_1^{-1} \cdot \nu = \nu, \tag{3}$$

and thus by (2) and (3),

$$\nu(h_1 n_2 h_1^{-1}) = \nu(n_2). \tag{4}$$

Finally by (1) and (4) and by definition of $\widetilde{\nu}$, we have

$$\widetilde{\nu}((n_1h_1)(n_2h_2)) = \nu(n_1)\nu(n_2) = \widetilde{\nu}(n_1h_1)\widetilde{\nu}(n_2h_2),$$

that is,

$$\widetilde{\nu}((n_1h_1)(n_2h_2)) = \widetilde{\nu}(n_1h_1)\widetilde{\nu}(n_2h_2), \tag{5}$$

which shows that $\widetilde{\nu}\colon G_\nu \to \mathbb{T}$ is a homomorphism extending ν.

We can now apply Proposition 16.6. Since $H_\nu \approx G_\nu/N$, for every irreducible representation $\rho\colon H_\nu \to \mathbf{U}(H_\rho)$ of H_ν, the map $\sigma\colon G_\nu \to \mathbf{U}(H_\rho)$ given by

$$\sigma(nh) = \widetilde{\nu}(nh)\rho(h) = \nu(n)\rho(h), \quad n \in N,\, h \in H_\nu,$$

defines an irreducible representation of G_ν in H_ρ such that $\sigma(n) = \nu(n)\mathrm{id}_{H_\rho}$ for all $n \in N$.

Definition 16.8. For any $\nu \in \widehat{N}$ and any irreducible representation $\rho\colon H_\nu \to \mathbf{U}(H_\rho)$ of H_ν, the irreducible representation $\sigma\colon G_\nu \to \mathbf{U}(H_\rho)$ given by

$$\sigma(nh) = \nu(n)\rho(h), \quad n \in N,\, h \in H_\nu$$

is denoted by $\nu\rho$.

Since the restriction of $\nu\rho$ to H_ν is equal to σ, it is easy to see that $\nu\rho$ is equivalent to $\nu\rho'$ iff ρ is equivalent to ρ'.

Remark: (Serre) Since $H_\nu \approx G_\nu/N$, there is a surjective quotient map $q_\nu\colon G_\nu \to H_\nu$, so any representation $\rho\colon H_\nu \to \mathbf{U}(H_\rho)$ of H_ν lifts to the representation $q_\nu \circ \rho\colon G_\nu \to \mathbf{U}(H_\rho)$ of G_ν. It is also clear that if ρ is irreducible, then so is $q_\nu \circ \rho$. Write $\widetilde{\rho} = q_\nu \circ \rho$. Since $\widetilde{\nu}\colon G_\nu \to \mathbf{U}(1)$ is also an irreducible representation of G_ν, we deduce that $\nu\rho$ is equivalent to the tensor product representation $\widetilde{\nu} \otimes \widetilde{\rho}$ (recall that $\mathbb{C} \otimes_{\mathbb{C}} W \approx W$ for any complex vector space W).

We can now apply Theorem 16.4 and Theorem 16.5 to the above situation to obtain a complete characterization of the irreducible representations of a semi-direct product $G = N \rtimes H$ (with N normal an abelian) in terms of the characters of N and of the irreducible representations of the little groups H_ν associated with the characters $\nu \in \widehat{N}$.

Theorem 16.7. *Let G be locally compact group which is a semi-direct product $G = N \rtimes H$ with N normal and abelian. Suppose that G acts regularly on $\widehat{N}$.*

(1) *For any $\nu \in \widehat{N}$, if $\rho\colon H_\nu \to \mathbf{U}(H_\rho)$ is any irreducible representation of the little group H_ν, then the induced representation $\mathrm{Ind}_{G_\nu}^G \nu\rho$ of G (with $\nu\rho$ as in Definition 16.8) is irreducible.*

(2) *Every irreducible representation $U\colon G \to \mathbf{U}(H_U)$ of G is equivalent to some irreducible induced representation $\mathrm{Ind}_{G_\nu}^G \nu\rho$ as in (1).*

(3) *Two induced representations $\mathrm{Ind}_{G_\nu}^G \nu\rho$ and $\mathrm{Ind}_{G_{\nu'}}^G \nu'\rho'$ are equivalent iff $\nu' = s \cdot \nu$ for some $s \in G$ (ν and ν' belong to the same orbit), and the representation ρ and $h \mapsto \rho'(s^{-1}hs)$ are equivalent.*

We are now ready for some examples.

Example 16.1. Consider the group $\mathbf{SE}(n)$ of rigid motions of $\mathbb{R}^n$ defined as the group of $(n+1) \times (n+1)$ matrices

$$\mathbf{SE}(n) = \left\{ \begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} \mid Q \in \mathbf{SO}(n),\, u \in \mathbb{R}^n \right\}.$$

We assume that $n \geq 2$, since $\mathbf{SE}(1) \approx \mathbb{R}$ is abelian so its irreducible unitary representations are one-dimensional, and thus are of the form $z \mapsto \chi(x)z$ for all $x \in \mathbb{R}$ and all $z \in \mathbb{C}$, where χ is any character of $\mathbb{R}$. We know from Proposition 10.9 the characters of $\mathbb{R}$ are of the form $x \mapsto e^{iyx}$, for any fixed $y \in \mathbb{R}$. The subgroups N and H are defined as follows:

$$N = \left\{ \begin{pmatrix} I_n & u \\ 0 & 1 \end{pmatrix} \mid u \in \mathbb{R}^n \right\}, \quad H = \left\{ \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \mid Q \in \mathbf{SO}(n) \right\}.$$

We have

$$\begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} I_n & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix},$$

$$\begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R & v \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} QR & u + Qv \\ 0 & 1 \end{pmatrix},$$

and so

$$\begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} Q^\top & -Q^\top u \\ 0 & 1 \end{pmatrix}.$$

Clearly, $N \cap H = \{I_{n+1}\}$ and N is abelian. We also have

$$\begin{aligned}
\begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix}^{-1} &= \begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q^\top & -Q^\top u \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} Q & u + Qv \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q^\top & -Q^\top u \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} I_n & Qv \\ 0 & 1 \end{pmatrix},
\end{aligned}$$

so N is a normal subgroup. Consequently, $\mathbf{SE}(n)$ is the semidirect product $\mathbf{SE}(n) = N \rtimes H$. It is also clear that H is isomorphic to $\mathbf{SO}(n)$ and that N is isomorphic to $\mathbb{R}^n$, so we may write $\mathbf{SE}(n) = \mathbb{R}^n \rtimes \mathbf{SO}(n)$. It is often convenient to use a more concise notation for the element of $\mathbf{SE}(n) = \mathbb{R}^n \rtimes \mathbf{SO}(n)$, namely we denote the matrix

$$\begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix}$$

by (u, Q). Multiplication in $\mathbf{SE}(n)$ is then given by

$$(u, Q)(v, R) = (u + Qv, QR).$$

The action of $\mathbf{SE}(n)$ on $\mathbb{R}^n$ is given by

$$(u, Q)x = Qx + u, \quad x \in \mathbb{R}^n,$$

namely rotate x by Q and then translate by u. This is equivalent to the usual trick of embedding $\mathbb{R}^n$ in $\mathbb{R}^{n+1}$ by mapping x to $\binom{x}{1}$, and then

$$\begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} Qx + u \\ 1 \end{pmatrix}.$$

We have to figure out how $\mathbf{SE}(n)$ acts on $\widehat{N}$ to determine its orbits. Observe that

$$\begin{aligned} \begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} I_n & v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} Q^\top & -Q^\top u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} Q^\top & -Q^\top u + Q^\top v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q & u \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} I_n & Q^\top v \\ 0 & 1 \end{pmatrix}, \end{aligned}$$

that is

$$(u, Q)^{-1}(v, I_n)(u, Q) = (Q^\top v, I_n). \tag{$\dagger$}$$

To describe the action of $G = \mathbf{SE}(n)$ on $\widehat{N}$ we introduce the isomorphism $t \colon N \to \mathbb{R}^n$ given by

$$t(x, I_n) = x, \quad x \in \mathbb{R}^n.$$

Then we have an isomorphism between $\widehat{\mathbb{R}^n}$ and $\widehat{N}$ given by $\chi \mapsto \chi \circ t$, with $\chi \in \widehat{\mathbb{R}^n}$. By Corollary 10.11, the characters in $\widehat{\mathbb{R}^n}$ are the homomorphims χ_y from $\mathbb{R}^n$ to $\mathbb{T}$ (with $y \in \mathbb{R}^n$) given by

$$\chi_y(x) = e^{iy\cdot x}, \qquad x \in \mathbb{R}^n,$$

where $y\cdot x$ is the Euclidean product in $\mathbb{R}^n$ ($y\cdot x = \sum_{k=1}^n y_k x_k$). By composing the isomorphism from $\mathbb{R}^n$ to $\widehat{\mathbb{R}^n}$ given by $y \mapsto \chi_y$ and the isomorphism between $\widehat{\mathbb{R}^n}$ and $\widehat{N}$ given by $\chi \mapsto \chi \circ t$, we obtain the isomorphism between $\mathbb{R}^n$ and $\widehat{N}$ given by $y \mapsto \chi_y \circ t$ (with $y \in \mathbb{R}^n$).

Thus the characters in $\widehat{N}$ are of the form $(x, I_n) \mapsto \chi_y(t(x, I_n)) = \chi_y(x)$. By Definition 16.1 the action of $G = \mathbf{SE}(n)$ on $\widehat{N}$ is given by

$$((u, Q) \cdot \chi_y)(x, I_n) = \chi_y(t((u, Q)^{-1}(x, I_n)(u, Q))), \quad x \in \mathbb{R}^n,$$

which by (†) yields

$$((u, Q) \cdot \chi_y)(x, I_n) = \chi_y(t(Q^\top x, I_n)) = \chi_y(Q^\top x) = e^{iy\cdot(Q^\top x)} = e^{i(Qy)\cdot x)} = \chi_{Qy}(t(x, I_n)).$$

Therefore, under the isomorphism between $\mathbb{R}^n$ and $\widehat{N}$ given by $y \mapsto \chi_y \circ t$ (with $y \in \mathbb{R}^n$), we see that the action of $G = \mathbf{SE}(n)$ on $\widehat{N}$ is the action of $G = \mathbf{SE}(n)$ on $\mathbb{R}^n$ given by

$$(u, Q)(y) = Qy, \quad y \in \mathbb{R}^n; \tag{††}$$

in other words, only the rotation Q is applied. This is the usual action of $\mathbf{SO}(n)$ on $\mathbb{R}^n$.

Remark: Note the subtle point that the action of $G = \mathbf{SE}(n)$ on $\widehat{N}$ uses a *right conjugation* of (x, I_n) by (u, Q), namely $(u, Q)^{-1}(x, I_n)(u, Q))$, and this yields $(Q^\top x, I_n)$. The appearance of $Q^\top$ seems wrong, but it is compensated by the fact that in the argument of χ_y, we now have the inner product $y \cdot (Q^\top x)$, and in order to make the input x appear, we transpose again to obtain $Qy \cdot x = y \cdot (Q^\top x)$.

Remember that we have an isomorphism between $\mathbb{R}^n$ and $\widehat{N}$ given by $y \mapsto \chi_y \circ t$ (with $y \in \mathbb{R}^n$), so a character $\nu \in \widehat{N}$ may be denoted by ν_y. Using this isomorphism, it is easy to determine the orbits and the little groups. By (††), the orbits of the action of $G = \mathbf{SE}(n)$ on $\widehat{N}$ can be viewed as the orbits of the action of $\mathbf{SO}(n)$ on $\mathbb{R}^n$, namely for every $r \in [0, +\infty)$, the sphere $S_r(0)$ of radius r centered at the origin,

$$\mathcal{O}_r = \{x \in \mathbb{R}^n \mid \|x\|_2 = r\}.$$

For $r = 0$, we have $\mathcal{O}_0 = \{0_n\}$. For the countable separation property, we use the G-invariant annuli

$$\{x \in \mathbb{R}^n \mid \alpha < \|x\|_2 < \beta\}$$

with $\alpha < \beta$ rational. For $r > 0$, we pick the special representative re_1 on the sphere $\mathcal{O}_r$ with $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^n$. Then we see immediately that the little group H_{re_1} is isomorphic to $\mathbf{SO}(n-1)$. For $r = 0$, the little group H_0 is $\mathbf{SO}(n)$ and $G_0 = G = \mathbf{SE}(n)$. For $r > 0$, the characters $\nu \in \widehat{N}$ corresponding to points in $\mathcal{O}_r$ are of the form $\chi_y \circ t$, with $y \in \mathbb{R}^n$ and $\|y\|_2 = r$. Earlier we denoted them by ν_y.

Theorem 16.4 yields all irreducible representations of $\mathbf{SE}(n)$.

(1) For $r = 0$, we have $G_0 = G$ and $H_0 = \mathbf{SO}(n)$. We obtain the finite-dimensional irreducible representations $q \circ \sigma$ obtained by lifting the irreducible representations σ of $\mathbf{SO}(n)$ to $\mathbf{SE}(n)$ by composition with the quotient map $q\colon \mathbf{SE}(n) \to \mathbf{SO}(n)$.

(2) For every $r > 0$, for every $y \in \mathbb{R}^n$ with $\|y\|_2 = r$, we have the character $\nu_{r,y}$ given by $\nu_{r,y}(x) = e^{i(y \cdot x)}$. We also have $H_{r,y} = \mathbf{SO}(n-1)$ and $G_{r,y} = \mathbb{R}^n \rtimes \mathbf{SO}(n-1)$. Then for every irreducible representation $\rho\colon \mathbf{SO}(n-1) \to \mathbf{U}(H_\rho)$ of $\mathbf{SO}(n-1)$, we have the irreducible representation $\nu_{r,y}\rho\colon \mathbb{R}^n \rtimes \mathbf{SO}(n-1) \to \mathbf{U}(H_\rho)$ given by

$$(\nu_{r,y}\rho)(xQ) = \nu_{r,y}(x)\rho(Q) = e^{iy\cdot x}\rho(Q), \quad x \in \mathbb{R}^n,\, Q \in \mathbf{SO}(n-1),\, \|y\|_2 = r.$$

The induced representation $\mathrm{Ind}_{\mathbb{R}^n \rtimes \mathbf{SO}(n-1)}^{\mathbf{SE}(n)} \nu_{r,y}\rho$ is an irreducible representation of $\mathbf{SE}(n)$.

In the special case of (2) where $\rho\colon \mathbf{SO}(n-1) \to \mathbf{U}(1)$ is the trivial representation ($\rho(Q) = 1$ for all $Q \in \mathbf{SO}(n-1)$), it can be shown that the induced representation $\mathrm{Ind}_{\mathbb{R}^n \rtimes \mathbf{SO}(n-1)}^{\mathbf{SE}(n)} \nu_{r,y}\rho$ is equivalent to the induced representation $\mathrm{Ind}_{\mathbb{R}^n}^{\mathbf{SE}(n)} \nu_{r,y}$ (see Folland [33], Section 6.3). But we have determined such induced representations in Example 15.3. We found that these are the irreducible representations $\widetilde{\Pi}\colon \mathbf{SE}(n) \to \mathbf{U}(\mathrm{L}^2_\lambda(S^{n-1};\mathbb{C}))$ of class 1 described in Vilenkin [101] (Chapter XI, Section 2) given by

$$(\widetilde{\Pi}_{(a,Q)}(f))(x) = e^{ir(x\cdot a)} f(Q^\top x), \quad (a,Q) \in \mathbf{SE}(n),\, x \in S^{n-1},\, f \in \mathrm{L}^2_\lambda(S^{n-1};\mathbb{C}),\, r > 0.$$

For $n = 2, 3$, we can be more precise.

(1) For $n = 2$, we have $\mathbf{SO}(1) = \{1\}$. Thus for $r > 0$ the irreducible representations of $\mathbf{SE}(2)$ are of the form

$$\mathrm{Ind}_{\mathbb{R}^2}^{\mathbf{SE}(2)} \nu_{r,y}, \quad \text{with } \nu_{r,y}(x) = e^{iyx},\, x, y \in \mathbb{R}^2,\, \|y\|_2 = r.$$

According to the above discussion, they are equivalent to the irreducible representations $\widetilde{\Pi}\colon \mathbf{SE}(2) \to \mathbf{U}(\mathrm{L}^2_\lambda(S^1;\mathbb{C}))$ of class 1 described in Vilenkin [101] (Chapter IV, Section 2) given by

$$(\widetilde{\Pi}_{(a,Q)}(f))(x) = e^{ir(x\cdot a)} f(Q^\top x), \quad (a,Q) \in \mathbf{SE}(2),\, x \in S^1,\, f \in \mathrm{L}^2_\lambda(S^1;\mathbb{C}),\, r > 0.$$

For $r = 0$, $G_0 = \mathbf{SO}(2)$. The group $\mathbf{SO}(2)$ is abelian and $\mathbf{SO}(2) \approx \mathbf{U}(1) \approx \mathbb{T}$, so we know from Proposition 10.9 that the irreducible representations of $\mathbf{SO}(2)$ are the homomorphisms $\rho_k\colon \mathbf{SO}(2) \to \mathbf{U}(1)$ given by

$$\rho_k(e^{i\theta})(z) = e^{ik\theta} z, \quad k \in \mathbb{Z},\, 0 \le \theta < 2\pi,\, z \in \mathbb{C}.$$

We obtain irreducible representations of $\mathbf{SE}(2)$ obtained by lifting the irreducible representations ρ_k of $\mathbf{SO}(2)$ to $\mathbf{SE}(2)$ by composing with the projection map $q\colon \mathbf{SE}(2) \to \mathbf{SO}(2)$.

(2) For $n = 3$, if $r > 0$ then $H_{re_1} = \mathbf{SO}(2)$. As in (1), the irreducible representations of $\mathbf{SO}(2) \approx \mathbf{U}(1)$ are the homomorphisms $\rho_k\colon \mathbf{SO}(2) \to \mathbf{U}(1)$ given by

$$\rho_k(e^{i\theta})(z) = e^{ik\theta} z, \quad k \in \mathbb{Z},\, 0 \le \theta < 2\pi,\, z \in \mathbb{C}.$$

We obtain the irreducible representations $\nu_{r,y}\rho_k \colon \mathbb{R}^3 \rtimes \mathbf{SO}(2) \to \mathbf{U}(1)$ given by

$$(\nu_{r,y}\rho_k)(xe^{i\theta})(z) = e^{i(y\cdot x + k\theta)}z, \;\; x, y \in \mathbb{R}^3, \|y\|_2 = r, \; 0 \leq \theta < 2\pi, \; k \in \mathbb{Z}, \; z \in \mathbb{C},$$

which yield the irreducible representations $\mathrm{Ind}_{\mathbb{R}^3 \rtimes \mathbf{SO}(2)}^{\mathbf{SE}(3)} \nu_{r,y}\rho_k$ of $\mathbf{SE}(3)$. In the special case $k = 0$ these are equivalent to the irreducible representations $\widetilde{\Pi} \colon \mathbf{SE}(2) \to \mathbf{U}(\mathrm{L}^2_\lambda(S^2; \mathbb{C}))$ of class 1 given by

$$(\widetilde{\Pi}_{(a,Q)}(f))(x) = e^{ir(x\cdot a)} f(Q^\top x), \quad (a, Q) \in \mathbf{SE}(3), \; x \in S^2, \; f \in \mathrm{L}^2_\lambda(S^2; \mathbb{C}), \; r > 0.$$

If $r = 0$, We obtain the irreducible representations of $\mathbf{SE}(3)$ obtained by lifting the irreducible representations of $\mathbf{SO}(3)$ to $\mathbf{SE}(3)$ by composing with the projection map $q \colon \mathbf{SE}(3) \to \mathbf{SO}(3)$.

Chapter 17

Harmonic Analysis on Gelfand Pairs

This chapter is the culmination of all of the theories discussed in this book. We are able to present a very general version of the Fourier transform on a homogeneous space G/K, where (G, K) is a Gelfand pair. This chapter presents material discussed in Dieudonné [22] (Chapter XXII, Sections 6-10).

We saw in Section 15.10 that if G is a compact group and if H is a closed subgroup of G, then the algebra $\mathrm{L}^2(H\backslash G/H)$ is commutative if and only if $(\rho : \sigma_0) \leq 1$ for all $\rho \in R(G)$ (where σ_0 is the class of the trivial representation of H). If so, then for every $\rho \in R(G)$ such that $(\rho : \sigma_0) = 1$, the ideal $\mathfrak{a}_{\rho,\sigma_0}$ is one-dimensional and is spanned by the function

$$\omega_\rho(s) = \theta_{\rho,\sigma_0} = \frac{1}{n_\rho} m_{11}^{(\rho)}(s),$$

which is continuous and of positive type. The function ω_ρ is called a (zonal) spherical function.

The goal of this chapter is to generalize the above results for a compact group to a locally compact (metrizable and separable) unimodular group G and to a compact subgroup K of G.

The first difficulty is that if G is not compact, then $\mathrm{L}^2(G)$ is not closed under convolution (in general, $\mathrm{L}^2(G)$ is not contained in $\mathrm{L}^1(G)$). So we have to work with $\mathcal{K}(G)$ instead (recall that $\mathcal{K}(G)$ is the subset of $\mathcal{C}(G)$ consisting of the continuous functions with compact support $f\colon G \to \mathbb{C}$).

There is a bijection between the space $\mathcal{C}(G/K)$ of continuous functions $f\colon G/K \to \mathbb{C}$ and the subspace of continuous functions $g\colon G \to \mathbb{C}$ such that

$$g(st) = g(s), \qquad \text{for all } s \in G \text{ and all } t \in K.$$

We also have a bijection between the space $\mathcal{C}(K\backslash G)$ of continuous functions $f\colon K\backslash G \to \mathbb{C}$ and the subspace of continuous functions $g\colon G \to \mathbb{C}$ such that

$$g(ts) = g(s), \qquad \text{for all } s \in G \text{ and all } t \in K.$$

Let $\mathcal{C}(K\backslash G/K) = \mathcal{C}(G/K) \cap \mathcal{C}(K\backslash G)$, which consists of the continuous functions $f\colon G \to \mathbb{C}$ which are constant on double cosets KsK ($s \in G$), and let $\mathcal{K}(K\backslash G/K)$ be the subspace of $\mathcal{C}(K\backslash G/K)$ consisting of the continuous functions with compact support. The space $\mathcal{K}(K\backslash G/K)$ is an involutive subalgebra of $\mathcal{K}(G)$, and thus of $\mathrm{L}^1(G)$.

The key ingredient is the Banach algebra $\mathrm{L}^1(K\backslash G/K)$, the closure of $\mathcal{K}(K\backslash G/K)$ in $\mathrm{L}^1(G)$. Gelfand's remarkable discovery is that much of the harmonic analysis on abelian locally compact groups and compact groups can be generalized to a pair (G, K) where G is a noncommutative locally compact unimodular (metrizable and separable) group, and K is a compact subgroup of G, *if the algebra* $\mathrm{L}^1(K\backslash G/K)$ *is commutative*. In this case, (G, K) is called a *Gelfand pair*.

Fortunately, there is a sufficient criterion for a pair (G, K) to be a Gelfand pair involving an involutive isomorphism $\sigma\colon G \to G$ such that K is a closed subgroup of the group G^σ of fixed points of σ (see Theorem 17.2). This criterion is reminiscent of Élie Cartan's notion of symmetric space (see Helgason [47] or Gallier and Quaintance [38]), and indeed, many kinds of symmetric spaces are Gelfand pairs. The proof that a pair satisfying this criterion is a Gelfand pair is given in Section 17.1. The conditions of this criterion are flexible enough to apply to three broad classes of pairs (G, K); see Section 17.6.

The purpose of Section 17.2 is to characterize the characters of the algebra $\mathrm{L}^1(K\backslash G/K)$ in terms of certain functions in $\mathcal{C}(K\backslash G/K)$ called *spherical functions*. Every character ζ of the commutative Banach algebra $\mathrm{L}^1(K\backslash G/K)$ is given by a unique function $\omega \in \mathcal{C}(K\backslash G/K)$ which is bounded and continuous on G, with

$$\zeta(f) = (f, \omega) = \int_G f(x)\omega(x)\, d\lambda_G(x), \qquad f \in \mathcal{L}^1(K\backslash G/K);$$

see Proposition 17.4. A function $\omega \in \mathcal{C}(K\backslash G/K)$ as above is called a *spherical function*.

Two criteria for a bounded function $\omega \in \mathcal{C}(K\backslash G/K)$ (different from the zero function) to be a spherical function are given in Theorem 17.6. In particular, the function ω is a spherical function on G relative to K iff

$$\int_K \omega(xty)\, d\lambda_K(t) = \omega(x)\omega(y) \qquad \text{for all } x, y \in G. \tag{s_1}$$

The space of spherical functions on the Gelfand pair (G, K) is denoted $\mathbf{S}(G/K)$. The subspace of characters of the commutative involutive Banach algebra $A = \mathrm{L}^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$ whose restriction to $\mathrm{L}^1(K\backslash G/K)$ is not the zero function is denoted by $\mathsf{X}_0(A)$. This subspace is locally compact in the weak*-topology (metrizable and separable).

The map $\omega \mapsto \zeta_\omega = (f, \omega)$ is a homeomorphism of $\mathbf{S}(G/K)$ equipped with the induced topology of Fréchet space of $\mathcal{C}(G)$ onto $\mathsf{X}_0(A)$ equipped with the topology induced by the weak*-topology of the dual A' of A. Consequently, $\mathbf{S}(G/K)$ is locally compact.

An important class of Lie groups that yield Gelfand pairs are the real forms of a complex semi-simple Lie group, and Sections 17.3–17.5 are devoted to a discussion of these Lie groups.

One needs to understand how to find the *real forms* $\mathfrak{g}_0$ of a complex Lie algebra $\mathfrak{g}$, that is, the real Lie algebras $\mathfrak{g}_0$ such that

$$\mathfrak{g} = \mathfrak{g}_0 \oplus i\mathfrak{g}_0,$$

namely, $\mathfrak{g}$ is the complexification of $\mathfrak{g}_0$. Finding such real algebras $\mathfrak{g}_0$ is equivalent to finding certain semilinear idempotent maps on $\mathfrak{g}$ called *conjugations* (see Proposition 17.11). Now, if we happen to have some semi-simple real form $\mathfrak{g}_u$ such that

$$\mathfrak{g} = \mathfrak{g}_u \oplus i\mathfrak{g}_u,$$

and if c_u is the corresponding conjugation, then it can be shown that *all* the other real forms $\mathfrak{g}_0$ of $\mathfrak{g}$ are given by conjugations c_0 that commute with c_u. Then all this has to be promoted to Lie groups (essentially by using the exponential map).

Three examples of Gelfand pairs are discussed in Section 17.6. In the first example, G is a compact Lie group that has an involutive automorphism σ; this corresponds to symmetric spaces of compact type. In the second example, G_1 arises as a real form of a complex, semi-simple, simply-connected Lie group G, and G_1 has finite center. This corresponds to a symmetric space of noncompact type. The third example is a certain kind of semi-direct product; a typical illustration of this case is the group of rigid motions $\mathbf{SE}(n, \mathbb{R})$.

The Fourier transform and the Fourier cotransform for a Gelfand pair are introduced in Section 17.7. For every function $f \in \mathcal{L}^1(K\backslash G/K)$, the *Fourier cotransform* $\overline{\mathcal{F}}f$ of f is the function $\overline{\mathcal{F}}f \colon \mathbf{S}(G/K) \to \mathbb{C}$ given by

$$(\overline{\mathcal{F}}f)(\omega) = (f, \omega) = \int_G f(x)\omega(x)\, d\lambda_G(x), \qquad \omega \in \mathbf{S}(G/K),$$

and the *Fourier transform* $\mathcal{F}f$ of f is the function $\mathcal{F}f \colon \mathbf{S}(G/K) \to \mathbb{C}$ given by

$$(\mathcal{F}f)(\omega) = (\check{f}, \omega) = \int_G f(x^{-1})\omega(x)\, d\lambda_G(x) = \int_G f(x)\omega(x^{-1})\, d\lambda_G(x), \qquad \omega \in \mathbf{S}(G/K).$$

Using the space $\mathbf{S}(G/K)$ of spherical functions as the domain of $\mathcal{F}$ and $\overline{\mathcal{F}}$ instead of characters yields a simultaneous generalization of the case where G is commutative and the case where G is compact. On $\mathcal{L}^1(K\backslash G/K)$, we have the familar equations

$$\mathcal{F}(f * g) = (\mathcal{F}f)(\mathcal{F}g), \qquad \overline{\mathcal{F}}(f * g) = (\overline{\mathcal{F}}f)(\overline{\mathcal{F}}g). \tag{$*$}$$

In Section 17.8 we generalize Fourier inversion. For this, we use the construction of certain positive σ-Radon measures from measures of positive type (recall Definition 12.18) using the *Plancherel transform*.

When G is not compact, the spherical functions in $\mathbf{S}(G/K)$ are not necessarily of positive type. The subset of $\mathbf{S}(G/K)$ consisting of the *spherical functions of positive type* is denoted by $\mathbf{Z}(G/K)$.

Theorem 17.21 states that given a measure μ of positive type on G, there is a unique (positive) Radon measure $\mu^{\triangle}$ on $\mathbf{Z}(G/K)$ such that for every function $f \in \mathcal{K}(K\backslash G/K)$, the Fourier cotransform $\overline{\mathcal{F}}f$ belong to $\mathcal{L}^2_{\mu^{\triangle}}(\mathbf{Z}(G/K);\mathbb{C})$, and for any two functions $f, g \in \mathcal{K}(K\backslash G/K)$, we have

$$\int_G (g^* * f) d\mu = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}}f)(\omega)\overline{(\overline{\mathcal{F}}g)(\omega)}\, d\mu^{\triangle}(\omega).$$

The measure $\mu^{\triangle}$ is called the *Plancherel transform* of μ.

In particular, if $\mu = \delta_e$, the Dirac measure, we obtain a measure $m_{\mathbf{Z}} = \delta_e^{\triangle}$, called the *canonical measure* on $\mathbf{Z}(G/K)$, and then the linear maps $f \mapsto \mathcal{F}f$ and $f \mapsto \overline{\mathcal{F}}f$, with $f, g \in \mathcal{K}(K\backslash G/K)$, are isometries, and by Theorem 17.21, these maps extend to isomorphisms from the Hilbert space $\mathrm{L}^2(K\backslash G/K)$ onto the Hilbert space $\mathrm{L}^2_{m_{\mathbf{Z}}}(\mathbf{Z}(G/K))$. This is a generalization of the Plancherel theorem (Theorem 10.27).

Another type of Fourier inversion formula is given by Proposition 17.28 The map $p \mapsto (p\,\lambda_G)^{\triangle}$ is a bijection between the space $\mathcal{P}_+(K\backslash G/K)$ of functions in $\mathcal{C}(K\backslash G/K)$ which are of positive type onto the space $\mathcal{M}^1_+(\mathbf{Z}(G/K))$ of bounded positive measures on $\mathbf{Z}(G/K)$.

Section 17.9 discusses an extension of the Plancherel transform to the space $\mathbf{P}(G)$ which is the complex vector space spanned by the union of the compex measures and the Radon measures. In Proposition 17.34 we obtain a Fourier inversion formula which yields the inversion formula of the Pontrjagin duality theorem, Theorem 10.30, as a special case.

Finally, in Section 17.10 we show that functions of positive type induce irreducible representations; see Theorem 17.35. We also state a theorem of Stone characterizing the unitary representations of $\mathbb{R}$ in a separable Hilbert space (Theorem 17.40).

17.1 Gelfand Pairs

In the rest of this chapter we assume that G is locally compact, metrizable, separable and unimodular group, and that K is a compact subgroup of G. Recall that there is a bijection between the space $\mathcal{C}(G/K)$ of continuous functions $f\colon G/K \to \mathbb{C}$ and the subspace of continuous functions $g\colon G \to \mathbb{C}$ such that

$$g(st) = g(s), \qquad \text{for all } s \in G \text{ and all } t \in K,$$

equivalently

$$g * \delta_t = g, \qquad \text{for all } t \in K.$$

This bijection is given by the map $f \mapsto f \circ \pi$, where $\pi\colon G \to G/K$ is the projection map. Observe that $(g * \delta_t)(s) = g(st^{-1})$ (see $(*_{\rho_{s^{-1}}})$ after Definition 8.25; since G is unimodular, the term $\Delta(t^{-1})$ is equal to 1), so the condition $g * \delta_t = g$ is equivalent to $g(st^{-1}) = g(s)$ for all $t \in K$, but this is equivalent to $g(st) = g(s)$ for all $t \in K$ since K is a group. From now on, we identify $\mathcal{C}(G/K)$ with the subspace of $\mathcal{C}(G)$ satisfying the above equivalent properties.

We also have a bijection between the space $\mathcal{C}(K\backslash G)$ of continuous functions $f\colon K\backslash G \to \mathbb{C}$ and the subspace of continuous functions $g\colon G \to \mathbb{C}$ such that

$$g(ts) = g(s), \qquad \text{for all } s \in G \text{ and all } t \in K,$$

equivalently

$$\delta_t * g = g, \qquad \text{for all } t \in K,$$

and we identify $\mathcal{C}(K\backslash G)$ with the subspace of $\mathcal{C}(G)$ satisfying the above equivalent properties. Observe that $(\delta_t * g)(s) = g(t^{-1}s)$, so $\delta_t * g = g$ is equivalent to $g(t^{-1}s) = g(s)$ for all $t \in K$, which is equivalent to $g(ts) = g(s)$ for all $t \in K$.

Let $\mathcal{C}(K\backslash G/K) = \mathcal{C}(G/K) \cap \mathcal{C}(K\backslash G)$, which consists of the continuous functions $f\colon G \to \mathbb{C}$ which are constant on double cosets KsK $(s \in G)$.

Since K is compact, for every compact subset A of G/K, the subset $\pi^{-1}(A)$ is compact in G. The map $f \mapsto f \circ \pi$ is thus a bijection between the subspace $\mathcal{K}(G/K)$ of $\mathcal{C}(G/K)$ onto a subspace of $\mathcal{K}(G)$. This subspace is denoted by $\mathcal{K}(G) \cap \mathcal{C}(G/K)$. Similarly, there is a bijection between the space $\mathcal{K}(K\backslash G)$ and the space $\mathcal{K}(G) \cap \mathcal{C}(K\backslash G)$, and a bijection between the space $\mathcal{K}(K\backslash G/K)$ and the space $\mathcal{K}(G) \cap \mathcal{C}(K\backslash G/K)$.

We saw earlier that $\mathcal{K}(G)$ is an involutive subalgebra under convolution of the algebra $\mathrm{L}^1(G)$ (see Example 9.6(4)). It follows that $\mathcal{K}(G/K)$ is left ideal in $\mathcal{K}(G)$ and that $\mathcal{K}(K\backslash G)$ is right ideal in $\mathcal{K}(G)$, and the involution $f \mapsto \overline{\check{f}}$ maps $\mathcal{K}(G/K)$ onto $\mathcal{K}(K\backslash G)$. As a consequence, $\mathcal{K}(K\backslash G/K)$ is an involutive subalgebra of $\mathcal{K}(G)$ (and so of $\mathrm{L}^1(G)$).

Let λ_K be the Haar measure on K normalized so that $\lambda_K(K) = 1$, and since G is assumed to be unimodular, let λ_G be a left and right-invariant Haar measure on G (since K is compact, it is unimodular so λ_K is also left and right-invariant). In order to study the characters of the Banach algebra $\mathrm{L}^1(K\backslash G/K)$, which is the closure of $\mathcal{K}(K\backslash G/K)$ in $\mathrm{L}^1(G)$, we need to project $\mathcal{C}(G)$ onto $\mathcal{C}(K\backslash G/K)$.

Definition 17.1. We define a projection map from $\mathcal{C}(G)$ onto $\mathcal{C}(K\backslash G/K)$ by

$$f^\sharp(s) = \int_K \int_K f(tst')\, d\lambda_K(t)\, d\lambda_K(t'), \qquad s \in G,$$

for any $f \in \mathcal{C}(G)$.

It is easily checked that if $f \in \mathcal{C}(G)$, then $f^\sharp(t_1 s t_1') = f^\sharp(s)$ for all $t_1, t_1' \in K$, so $f^\sharp \in \mathcal{C}(K\backslash G/K)$. As a consequence, since $\lambda(K) = 1$,

$$f^{\sharp\sharp}(s) = \int_K \int_K f^\sharp(tst')\, d\lambda_K(t)\, d\lambda_K(t') = \int_K \int_K f^\sharp(s)\, d\lambda_K(t)\, d\lambda_K(t') = f^\sharp(s),$$

and the map $f \mapsto f^\sharp$ is indeed a projection. It is also easy to check that for all $f \in \mathcal{C}(K\backslash G/K)$ and for all $g \in \mathcal{C}(G)$, we have

$$(fg)^\sharp = fg^\sharp.$$

Proposition 17.1. *The restriction of the projection $f \mapsto f^\sharp$ to $\mathcal{K}(G)$ maps $\mathcal{K}(G)$ onto $\mathcal{K}(K\backslash G/K)$, and we have*

$$(f * g)^\sharp = f * g^\sharp, \qquad (g * f)^\sharp = g^\sharp * f,$$

for all $f \in \mathcal{K}(K\backslash G/K)$, and all $g \in \mathcal{C}(G)$.

Proof. We leave the first statement as an exercise and prove the first of the two equations. We have

$$(f * g)(x) = \int_G f(s)g(s^{-1}x)\, d\lambda_G(s),$$

and using the left invariance of λ_G and the fact that $f(ts) = f(s)$ and $\lambda_K(K) = 1$, we have

$$\begin{aligned}
(f * g)^\sharp(x) &= \int_K \int_K \int_G f(s)g(s^{-1}txt')\, d\lambda_G(s)\, d\lambda_K(t)\, d\lambda_K(t') \\
&= \int_K \int_K \int_G f(ts)g(s^{-1}xt')\, d\lambda_G(s)\, d\lambda_K(t)\, d\lambda_K(t') \\
&= \int_K \int_K \int_G f(s)g(s^{-1}xt')\, d\lambda_G(s)\, d\lambda_K(t)\, d\lambda_K(t') \\
&= \int_K \int_G f(s)g(s^{-1}xt')\, d\lambda_G(s)\, d\lambda_K(t').
\end{aligned}$$

We also have

$$g^\sharp(x) = \int_K \int_K g(txt')\, d\lambda_K(t)\, d\lambda_K(t'),$$

and using the right invariance of λ_G and the fact that $f(st) = f(s)$ and $\lambda_K(K) = 1$, we have

$$\begin{aligned}
(f * g^\sharp)(x) &= \int_G \int_K \int_K f(s)g(ts^{-1}xt')\, d\lambda_K(t)\, d\lambda_K(t')\, d\lambda_G(s) \\
&= \int_G \int_K \int_K f(st)g(s^{-1}xt')\, d\lambda_K(t)\, d\lambda_K(t')\, d\lambda_G(s) \\
&= \int_G \int_K \int_K f(s)g(s^{-1}xt')\, d\lambda_K(t)\, d\lambda_K(t')\, d\lambda_G(s) \\
&= \int_G \int_K f(s)g(s^{-1}xt')\, d\lambda_K(t')\, d\lambda_G(s),
\end{aligned}$$

and by Fubini, we can interchange the integrations, which shows that $(f * g)^\sharp = f * g^\sharp$. The proof that $(g * f)^\sharp = g^\sharp * f$ is analogous. $\square$

A Gelfand pair (G, K) is defined as follows.

Definition 17.2. Let G be a locally compact (metrizable and separable) unimodular group, and let K be a compact subgroup. We say that the pair (G, K) is a *Gelfand pair* if the algebra $\mathcal{K}(K\backslash G/K)$ is commutative (under convolution).

Obviously, if G is abelian, then (G, K) is a Gelfand pair. The following theorem, which is one of the key theorems of the theory of Gelfand pairs, gives a sufficient criterion for a pair (G, K) to be a Gelfand pair. Using this criterion, we will show later that there are lots of Gelfand pairs.

Theorem 17.2. *(Gelfand) Let G be locally compact (metrizable and separable) unimodular group, $\sigma\colon G \to G$ be an involutive isomorphism of G ($\sigma \circ \sigma = \mathrm{id}_G$), and $G^\sigma = \{s \in G \mid \sigma(s) = s\}$ be the subgroup of elements of G left fixed by σ. Let K be any closed subgroup of G^σ and assume the following properties:*

(1) The subgroup G^σ is compact.

(2) Every $x \in G$ can be written (possibly not uniquely) as $x = yz$, with $y \in K$, $z \in G$, and $\sigma(z) = z^{-1}$.

Then (G, K) is Gelfand pair.

Proof. Since $\sigma^2 = \mathrm{id}_G$, we have $\mathrm{mod}(\sigma)^2 = \mathrm{mod}(\mathrm{id}_G) = 1$, and since $\mathrm{mod}(\sigma) > 0$, we get $\mathrm{mod}(\sigma) = 1$ (see Definition 8.17 and Proposition 8.31). It follows that σ leaves any Haar measure λ_G of G invariant. For every $f \in \mathcal{K}(G)$, let f^σ be the function given by

$$f^\sigma(s) = f(\sigma(s)) = f(\sigma^{-1}(s)), \qquad s \in G.$$

We check immediately that the map $\widehat{\sigma}\colon \mathcal{K}(G) \to \mathcal{K}(G)$ given by $\widehat{\sigma}(f) = f^\sigma$ is an involutive automorphism of the vector space $\mathcal{K}(G)$, and since σ leaves any Haar measure on G invariant, it is an automorphism of the *algebra* $\mathcal{K}(G)$ (under convolution). Since $\sigma(t) = t$ for all $t \in K$, if $f(x) = f(txt')$ for all $t, t \in K$, then $f(\sigma(txt')) = f(t\sigma(x)t') = f(\sigma(x))$, so if $f \in \mathcal{K}(K\backslash G/K)$ then $f^\sigma \in \mathcal{K}(K\backslash G/K)$, which means that the automorphism $\widehat{\sigma}$ leaves the algebra $\mathcal{K}(K\backslash G/K)$ invariant, and its restriction to $\mathcal{K}(K\backslash G/K)$ is an automorphism of this subalgebra. If we can prove that

$$f^\sigma * g^\sigma = g^\sigma * f^\sigma \qquad \text{for all } f, g \in \mathcal{K}(K\backslash G/K),$$

then we will have proven that $\mathcal{K}(K\backslash G/K)$ is commutative.

The trick is that for any function $f \in \mathcal{K}(K\backslash G/K)$, we have $f^\sigma = \check{f}$. Every $x \in G$ can be written as $x = yz$ with $y \in K$ and $\sigma(z) = z^{-1}$, which yields

$$\sigma(x) = \sigma(yz) = \sigma(y)\sigma(z) = yz^{-1} = y(z^{-1}y^{-1})y,$$

and for every $f \in \mathcal{K}(K\backslash G/K)$, we have

$$f(\sigma(x)) = f(y(z^{-1}y^{-1})y) = f(z^{-1}y^{-1}) = f(x^{-1}) = \check{f}(x);$$

that is, $f^\sigma = \check{f}$. Since G is unimodular, the Haar measure is right-invariant, so for any two function $f, g \in \mathcal{K}(G)$, we easily verify that

$$(\check{f} * \check{g})\check{} = g * f.$$

Then for $f, g \in \mathcal{K}(K\backslash G/K)$, we have

$$f^\sigma * g^\sigma = \check{f} * \check{g} = (g * f)\check{} = (g * f)^\sigma = g^\sigma * f^\sigma,$$

as claimed. $\square$

From now on *we assume that* (G, K) *is a Gelfand pair*.

Definition 17.3. The closure of $\mathcal{K}(K\backslash G/K)$ in $\mathrm{L}^1(G)$ is denoted $\mathrm{L}^1(K\backslash G/K)$. Then $\mathrm{L}^1(K\backslash G/K)$ is an involutive, commutative, Banach subalgebra of $\mathrm{L}^1(G)$. Similarly, denote by $\mathrm{L}^2(K\backslash G/K)$ the closure of $\mathcal{K}(K\backslash G/K)$ in $\mathrm{L}^2(G)$.

The following results are proven in Dieudonné [22] (Chapter XXII, Section 6).

Proposition 17.3. *The projection $f \mapsto f^\sharp$ of $\mathcal{K}(G)$ onto $\mathcal{K}(K\backslash G/K)$ extends to a continuous projection of $\mathrm{L}^1(G)$ onto $\mathrm{L}^1(K\backslash G/K)$. For any $f \in \mathcal{L}^1(G)$, we have $\left\|f^\sharp\right\|_1 \leq \|f\|_1$, and the class $[f^\sharp]$ is the class of the function $f^\sharp$ equal almost everywhere to the function*

$$s \mapsto \int_K \int_K f(tst')\, d\lambda_K(t)\, d\lambda_K(t').$$

Similarly, the projection $f \mapsto f^\sharp$ of $\mathcal{K}(G)$ onto $\mathcal{K}(K\backslash G/K)$ extends to a continuous projection of $\mathrm{L}^2(G)$ onto $\mathrm{L}^2(K\backslash G/K)$. For any $f \in \mathcal{L}^2(G)$, we have $\left\|f^\sharp\right\|_2 \leq \|f\|_2$, and the class $[f^\sharp]$ is the class of the function $f^\sharp$ equal almost everywhere to the function

$$s \mapsto \int_K \int_K f(tst')\, d\lambda_K(t)\, d\lambda_K(t').$$

Definition 17.4. We denote by $\mathcal{L}^1(G/K)$ and $\mathcal{L}^2(G/K)$ the subspaces of $\mathcal{L}^1(G)$ and $\mathcal{L}^2(G)$ consisting of the functions f such that for almost all $s \in G$, we have

$$f(st) = f(s) \qquad \text{for all } t \in K.$$

Similarly, we denote by $\mathcal{L}^1(K\backslash G)$ and $\mathcal{L}^2(K\backslash G)$ the subspaces of $\mathcal{L}^1(G)$ and $\mathcal{L}^2(G)$ consisting of the functions f such that for almost all $s \in G$, we have

$$f(ts) = f(s) \qquad \text{for all } t \in K.$$

Let $\mathcal{L}^1(K\backslash G/K) = \mathcal{L}^1(G/K) \cap \mathcal{L}^1(K\backslash G)$ and $\mathcal{L}^2(K\backslash G/K) = \mathcal{L}^2(G/K) \cap \mathcal{L}^2(K\backslash G)$.

If $f \in \mathcal{L}^1(G)$ (resp. $f \in \mathcal{L}^2(G)$), then $f^\sharp \in \mathcal{L}^1(K\backslash G/K)$ (resp. $f^\sharp \in \mathcal{L}^2(K\backslash G/K)$), and $\mathrm{L}^1(K\backslash G/K)$ (resp. $\mathrm{L}^2(K\backslash G/K)$) is the canonical image in $\mathrm{L}^1(G)$ (resp. $\mathrm{L}^2(G)$) of $\mathcal{L}^1(K\backslash G/K)$ (resp. $\mathcal{L}^2(K\backslash G/K)$).

We also obtain an alternative description of $\mathrm{L}^1(K\backslash G/K)$ (resp. $\mathrm{L}^2(K\backslash G/K)$) in terms of $\mathcal{L}^1(G/K)$ and $\mathcal{L}^1(K\backslash G)$ (resp. $\mathcal{L}^2(G/K)$ and $\mathcal{L}^2(K\backslash G)$).

If we denote by $\mathrm{L}^1(G/K)$, $\mathrm{L}^1(K\backslash G)$ (resp. $\mathrm{L}^2(G/K)$, $\mathrm{L}^2(K\backslash G)$) the canonical images in $\mathrm{L}^1(G)$ (resp. $\mathrm{L}^2(G)$) of $\mathcal{L}^1(G/K), \mathcal{L}^1(K\backslash G)$ (resp. $\mathcal{L}^2(G/K), \mathcal{L}^2(K\backslash G)$), then we have $\mathrm{L}^1(K\backslash G/K) = \mathrm{L}^1(G/K) \cap \mathrm{L}^1(K\backslash G)$ (resp. $\mathrm{L}^2(K\backslash G/K) = \mathrm{L}^2(G/K) \cap \mathrm{L}^2(K\backslash G)$).

17.2 Spherical Functions

Our next goal is to characterize the characters of the algebra $\mathrm{L}^1(K\backslash G/K)$ in terms of certain functions in $\mathcal{C}(K\backslash G/K)$ called spherical functions. In this chapter we will use the notation (f, g) as an abbreviation for

$$\int_G f(x)g(x)\, d\lambda_G(x),$$

whenever such an integral makes sense for some functions $f, g\colon G \to \mathbb{C}$.

Proposition 17.4. *Every character ζ of the commutative Banach algebra $\mathrm{L}^1(K\backslash G/K)$ is given by a unique function $\omega \in \mathcal{C}(K\backslash G/K)$ which is bounded and continuous on G, with*

$$\zeta(f) = (f, \omega) = \int_G f(x)\omega(x)\, d\lambda_G(x), \qquad f \in \mathcal{L}^1(K\backslash G/K).$$

Furthermore, ω is uniformly continuous for every left-invariant metric on G, and $|\omega(s)| \leq \omega(e) = 1$ for all $s \in G$.

Proof. Every character ζ of the commutative subalgebra $\mathrm{L}^1(K\backslash G/K)$ of $\mathcal{M}^1(G)$ can be extended to a character ζ' of the unital commutative Banach algebra $\mathrm{L}^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$ by setting $\zeta'(f + \lambda\delta_e) = \zeta(f) + \lambda$. By Theorem 9.19, we have $|\zeta(f)| \leq \|f\|_1$ for all $f \in \mathrm{L}^1(K\backslash G/K)$. As a consequence, the map $\Phi\colon \mathrm{L}^1(G) \to \mathbb{C}$ given by

$$\Phi(f) = \zeta(f^\sharp), \qquad f \in \mathrm{L}^1(G)$$

is a linear form of norm ≤ 1 because by Proposition 17.3, $\left\|f^\sharp\right\|_1 \leq \|f\|_1$, and by Theorem 5.51, there is a unique function $\omega_0 \in \mathcal{L}^\infty(G)$ with $\|\omega_0\|_\infty \leq 1$, such that

$$\Phi(f) = \zeta(f^\sharp) = \int_G f(x)\omega_0(x)\, d\lambda_G(x).$$

The problem is that $\omega_0 \in \mathcal{L}^\infty(G)$ is in the wrong space because we need it to be in $\mathcal{C}(K\backslash G/K)$ (and to be bounded by 1). To remedy this problem, we define another function ω in terms of ω_0, which we regularize by integrating against some function $f_0 \in \mathcal{K}(K\backslash G/K)$. In the end, we will see that ω_0 and ω are equal almost everywhere, but this will take some work.

First, observe that for all $t, t' \in K$ and all $f \in \mathcal{K}(G)$, we have

$$\int f(txt')\omega_0(x)\, d\lambda_G(x) = \int f(x)\omega_0(x)\, d\lambda_G(x). \qquad (*)$$

To prove this, if we let $h_{t,t'}$ be the function given by $h_{t,t'}(x) = f(txt')$, then will prove that $h^\sharp_{t,t'} = f^\sharp$, and so $\zeta(f^\sharp) = \zeta(h^\sharp_{t,t'})$. Indeed, using the left an right invariance of the Haar

measure λ_K, we have

$$\begin{aligned}
h^\sharp_{t,t'}(x) &= \int_K\int_K h_{t,t'}(t_1xt_2)\, d\lambda_K(t_1)\, d\lambda_K(t_2) \\
&= \int_K\int_K f(tt_1xt_2t')\, d\lambda_K(t_1)\, d\lambda_K(t_2) \\
&= \int_K\int_K f(t_1xt_2)\, d\lambda_K(t_1)\, d\lambda_K(t_2) = f^\sharp(x).
\end{aligned}$$

By definition, the character ζ is not identically zero, so there is some function $f_0 \in \mathcal{K}(K\backslash G/K)$ such that $\zeta(f_0) \neq 0$. Since ζ is a character of $\mathcal{L}^1(K\backslash G/K)$, for every function $g \in \mathcal{K}(K\backslash G/K)$, since $f_0 \in \mathcal{K}(K\backslash G/K)$ we also have $g*f_0 \in \mathcal{K}(K\backslash G/K)$, so $(g*f_0)^\sharp = g*f_0$, and using Fubini's theorem and the fact that ζ is a character of $\mathcal{K}(K\backslash G/K)$, we have

$$\begin{aligned}
\zeta(g) &= \zeta(f_0)^{-1}\zeta(f_0)\zeta(g) = \zeta(f_0)^{-1}\zeta(g)\zeta(f_0) \\
&= \zeta(f_0)^{-1}\zeta(g * f_0) = \zeta(f_0)^{-1}\zeta((g * f_0)^\sharp) \\
&= \zeta(f_0)^{-1}\iint_{G\times G} f_0(s^{-1}x)g(s)\omega_0(x)\, d\lambda_G(s)\, d\lambda_G(x) \\
&= \int_G g(s)\omega(s)\, d\lambda_G(s),
\end{aligned}$$

with

$$\omega(s) = \zeta(f_0)^{-1}\int_G f_0(s^{-1}x)\omega_0(x)\, d\lambda_G(x) = \zeta(f_0)^{-1}\int_G f_0(x)\omega_0(sx)\, d\lambda_G(x).$$

It follows by Theorem 14.10.6(ii) of Dieudonné [24] (Chapter XIV, Section 10) that ω is bounded in G and uniformly continuous for every left-invariant distance on G. Observe that for the integral on the right-hand side of

$$\zeta(g) = \int_G g(s)\omega(s)\, d\lambda_G(s)$$

to make sense, we need $g \in \mathcal{K}(K\backslash G/K)$, since we only know that $\omega_0 \in \mathcal{L}^\infty(G)$. We will extend the above equation to functions in $\mathrm{L}^1(K\backslash G/K)$ by density.

Next we need to prove that $\omega \in \mathcal{C}(K\backslash G/K)$. For all $t, t' \in K$, since $f_0 \in \mathcal{K}(K\backslash G/K)$, we have

$$\begin{aligned}
\omega(tst') &= \zeta(f_0)^{-1}\int_G f_0(t'^{-1}s^{-1}t^{-1}x)\omega_0(x)\, d\lambda_G(x) \\
&= \zeta(f_0)^{-1}\int_G f_0(s^{-1}t^{-1}x)\omega_0(x)\, d\lambda_G(x) \\
&= \zeta(f_0)^{-1}\int_G f_0(s^{-1}x)\omega_0(x)\, d\lambda_G(x) = \omega(x),
\end{aligned}$$

where we used $(*)$ with $f(x) = f_0(s^{-1}x)$ in the last step. Since $\mathcal{K}(K\backslash G/K)$ is dense in $\mathrm{L}^1(K\backslash G/K)$, we proved that

$$\zeta(f) = (f, \omega) = \int_G f(x)\omega(x)\, d\lambda_G(x), \qquad f \in \mathcal{L}^1(K\backslash G/K). \tag{$\dagger_1$}$$

We still need to prove that $|\omega(s)| \leq 1$ for all $s \in G$. Since we know that this is true of ω_0, we prove that $[\omega] = [\omega_0]$. By Theorem 7.10 and Theorem 5.51, it suffices to prove that

$$\int_G f(x)\omega(x)\, d\lambda(x) = \int_G f(x)\omega_0(x)\, d\lambda(x) = \zeta(f^\sharp) \qquad \text{for all } f \in \mathcal{K}(G).$$

Observe that since $\omega^\sharp = \omega$ (because $\omega \in \mathcal{C}(K\backslash G/K)$), and

$$\zeta(g) = \int_G g(s)\omega(s)\, d\lambda_G(s), \qquad g \in \mathrm{L}^1(K\backslash G/K),$$

if we can prove that

$$(f^\sharp, \psi) = \int_G f^\sharp(x)\psi(x)\, d\lambda_G(x) = \int_G f(x)\psi^\sharp(x)\, d\lambda_G(x) = (f, \psi^\sharp), \tag{$**$}$$

for all $f \in \mathcal{K}(G)$ and all $\psi \in \mathcal{C}(G)$, we will be done, because

$$\begin{aligned}
\int_G f(x)\omega_0(x)\, d\lambda(x) = \zeta(f^\sharp) &= \int_G f^\sharp(s)\omega(s)\, d\lambda_G(s) \\
&= \int_G f(s)\omega^\sharp(s)\, d\lambda_G(s) \\
&= \int_G f(s)\omega(s)\, d\lambda_G(s),
\end{aligned}$$

as claimed. To prove $(**)$, using Fubini, and the fact that G is unimomdular, we have

$$\begin{aligned}
\int_G f^\sharp(x)\psi(x)\, d\lambda_G(x) &= \int_G \int_K \int_K f(tst')\psi(s)\, d\lambda_K(t)\, d\lambda_K(t')\, d\lambda_G(s) \\
&= \int_G \int_K \int_K f(s)\psi(t^{-1}st'^{-1})\, d\lambda_K(t)\, d\lambda_K(t')\, d\lambda_G(s) \\
&= \int_G \int_K \int_K f(s)\psi(tst')\, d\lambda_K(t)\, d\lambda_K(t')\, d\lambda_G(s) \\
&= \int_G f(x)\psi^\sharp(x)\, d\lambda_G(x).
\end{aligned}$$

Now, since $\omega = \omega_0$ almost everywhere and since ω is continuous, we have proven that there is unique function $\omega \in \mathcal{C}(K\backslash G/K)$ satisfying the condition of the proposition and that $|\omega(s)| \leq 1$, since $\|\omega_0\|_\infty \leq 1$. If we let $s = e$ in

$$\omega(s) = \zeta(f_0)^{-1} \int_G f_0(s^{-1}x)\omega_0(x)\, d\lambda_G(x),$$

since

$$\zeta(f_0^\sharp) = \int_G f_0(x)\omega_0(x)\, d\lambda_G(x)$$

and $f_0^\sharp = f_0$ because $f_0 \in \mathcal{K}(K\backslash G/K)$, we get

$$\begin{aligned}\omega(e) &= \zeta(f_0)^{-1} \int_G f_0(x)\omega_0(x)\, d\lambda_G(x) \\ &= \zeta(f_0)^{-1}\zeta(f_0^\sharp) = \zeta(f_0)^{-1}\zeta(f_0) = 1,\end{aligned}$$

as claimed. $\square$

Remark: The above proof shows that

$$\omega(s) = (f_0, \omega)^{-1} \int_G f_0(s^{-1}x)\omega(x)\, d\lambda_G(x),$$

with $(f_0, \omega) = \int_G f_0(s)\omega(s)\, d\lambda_G(s)$, for any function $f_0 \in \mathcal{K}(K\backslash G/K)$ such that $(f_0, \omega) \neq 0$.

Definition 17.5. A bounded function $\omega \in \mathcal{C}(K\backslash G/K)$ is a *spherical (or zonal spherical) function* on G relative to K, if the function

$$f \mapsto (f, \omega) = \int_G f(s)\omega(s)\, d\lambda_G(s), \qquad f \in \mathrm{L}^1(K\backslash G/K)$$

is a character of $\mathrm{L}^1(K\backslash G/K)$, which means that the map $f \mapsto (f, \omega)$ is linear in $f \in \mathrm{L}^1(K\backslash G/K)$ and that

$$(f * g, \omega) = (f, \omega)(g, \omega), \qquad \text{for all } f, g \in \mathrm{L}^1(K\backslash G/K). \tag{$\dagger_2$}$$

Proposition 17.4 shows that if ω is a spherical function, then $\omega(e) = 1$ and $|\omega(s)| \leq 1$ for all $s \in G$.

Proposition 17.5. *If ω is a spherical function, then $\overline{\omega}$ and $\check{\omega}$ are also spherical functions.*

Proof. For every $f \in \mathcal{K}(K\backslash G/K)$, by Proposition 7.24, we have

$$\int f(s)\overline{\omega}(s)\, d\lambda_G(s) = \overline{\int \overline{f}(s)\omega(s)\, d\lambda_G(s)},$$

and for all $f, g \in \mathcal{K}(K\backslash G/K)$ we have

$$\overline{f * g} = \overline{f} * \overline{g},$$

which proves that $\overline{\omega}$ induces a character, because

$$\begin{aligned}\int (f*g)(s)\overline{\omega}(s)\,d\lambda_G(s) &= \overline{\int \overline{(f*g)}(s)\omega(s)\,d\lambda_G(s)} = \overline{\int \overline{f}*\overline{g}(s)\omega(s)\,d\lambda_G(s)}\\ &= \overline{\int \overline{f}(s)\omega(s)\,d\lambda_G(s)}\;\overline{\int \overline{g}(s)\omega(s)\,d\lambda_G(s)}\\ &= \int f(s)\overline{\omega}(s)\,d\lambda_G(s)\int g(s)\overline{\omega}(s)\,d\lambda_G(s).\end{aligned}$$

Since G is unimodular, we also have

$$\begin{aligned}\int_G f(s)\check{\omega}(s)\,d\lambda_G(s) &= \int_G f(s)\omega(s^{-1})\,d\lambda_G(s)\\ &= \int_G f(s^{-1})\omega(s)\,d\lambda_G(s)\\ &= \int_G \check{f}(s)\omega(s)\,d\lambda_G(s),\end{aligned}$$

and if $f, g \in \mathcal{K}(K\backslash G/K)$, we have

$$(f*g)\check{} = \check{g}*\check{f} = \check{f}*\check{g},$$

so $\check{\omega}$ induces a character. $\square$

Observe that Proposition 17.5 shows that a spherical function is uniformly continuous for every left-invariant as as well as every right-invariant distance on G. In general, a spherical function *does not have compact support*.

The next theorem gives criteria for a bounded function in $\mathcal{C}(K\backslash G/K)$ (different from the zero function) to be a spherical function.

Theorem 17.6. *Let ω be a bounded function in $\mathcal{C}(K\backslash G/K)$ not equal to the zero function. The following properties are equivalent:*

(1) The function ω is a spherical function on G relative to K.

(2) We have

$$\int_K \omega(xty)\,d\lambda_K(t) = \omega(x)\omega(y) \qquad \text{for all } x, y \in G. \tag{s_1}$$

(3) We have $\omega(e) = 1$, and for every function $f \in \mathcal{K}(K\backslash G/K)$, there is some $\lambda_f \in \mathbb{C}$ such that

$$f*\omega = \lambda_f\omega. \tag{s_2}$$

In fact

$$\lambda_f = (\check{f}, \omega) = \int_G \check{f}(x)\omega(x)\,d\lambda_G(x).$$

Similarly, $\omega * f = \lambda_f\omega$, for the same λ_f.

Proof. First we prove that (3) implies (1). Assume that $f * \omega = \lambda_f \omega$ for every $f \in \mathcal{K}(K\backslash G/K)$. Since $\omega(e) = 1$, we have

$$\lambda_{\check{f}} = (\check{f} * \omega)(e) = \int_G f(s)\omega(s)\, d\lambda_G(s) = (f, \omega).$$

Therefore,

$$\lambda_f = (\check{f}, \omega).$$

For all $f, g \in \mathcal{K}(K\backslash G/K)$, we have $(f * g)\check{} * \omega = (\check{g} * \check{f}) * \omega = \check{g} * (\check{f} * \omega)$, which implies that

$$\lambda_{\check{g}*\check{f}} = \lambda_{\check{g}}\lambda_{\check{f}},$$

and so

$$(f * g, \omega) = (f, \omega)(g, \omega),$$

which shows that the map $f \mapsto (f, \omega)$ is a character. The proof is similar if $\omega * f = \lambda_f \omega$.

We now prove that (1) implies (3). We claim that for all $f \in \mathcal{K}(K\backslash G/K)$, we have

$$f * \omega = (\check{f}, \omega)\omega = \int_G f(s^{-1})\omega(s)\, d\lambda_G(s).$$

Since $f * \omega \in \mathcal{C}(K\backslash G/K)$, in view of $(**)$, namely

$$(h^\sharp, \psi) = (h, \psi^\sharp), \qquad \text{for all } h \in \mathcal{K}(G) \text{ and all } \psi \in \mathcal{K}(G),$$

and since $\omega^\sharp = \omega$, it suffices to prove that

$$(g, f * \omega) = (\check{f}, \omega)(g, \omega), \qquad \text{for all } g \in \mathcal{K}(K\backslash G/K),$$

because by Proposition 17.1, $(f * \omega)^\sharp = f * \omega^\sharp = f * \omega$, so

$$\begin{aligned}
(h, f * \omega) &= (h, (f * \omega)^\sharp) \\
&= (h^\sharp, f * \omega) \\
&= (\check{f}, \omega)(h^\sharp, \omega) \\
&= (\check{f}, \omega)(h, \omega^\sharp) \\
&= (\check{f}, \omega)(h, \omega) = (h, (\check{f}, \omega)\omega)
\end{aligned}$$

for all $h \in \mathcal{K}(G)$, and thus $f * \omega = (\check{f}, \omega)\omega$. In that last step, we used the fact that the map $f_1, f_2 \mapsto (f_1, f_2)$ is bilinear, so for every $\lambda \in \mathbb{C}$ we have $\lambda(f_1, f_2) = (\lambda f_1, f_2) = (f_1, \lambda f_2)$.

Using Fubini and the fact that G is unimodular, we have

$$(g, f * \omega) = (\check{f} * g, \omega),$$

since

$$\int_G \int_G g(s)f(t)\omega(t^{-1}s) d\lambda_G(t)\, d\lambda_G(s) = \int_G \int_G g(ts)f(t)\omega(s) d\lambda_G(t)\, d\lambda_G(s)$$
$$= \int_G \int_G f(t^{-1})g(t^{-1}s)\omega(s) d\lambda_G(t)\, d\lambda_G(s).$$

Since ω is spherical function,

$$(g, f * \omega) = (\check{f} * g, \omega) = (\check{f}, \omega)(g, \omega).$$

Since $\check{\overline{\omega}}$ is also spherical (by Proposition 17.5), and since G is unimodular, we have

$$\omega * f = \overline{(\check{\overline{f}} * \check{\overline{\omega}})}^{\vee} = \overline{(\overline{f}, \check{\overline{\omega}})}\omega = (\check{f}, \omega)\omega.$$

Let us now prove that (2) and (3) are equivalent. For any $\omega \in \mathcal{C}(K\backslash G/K)$, define the function h by

$$h(x, y) = \int_K \omega(xty)\, d\lambda_K(t), \qquad x, y \in G.$$

A simple adaptation of the proof of Proposition 8.20 shows the map $x \mapsto h(x, y)$ is continuous. For all $t' \in K$, since $\omega(t'xty) = \omega(xty)$, we have $h(t'x, y) = h(x, y)$, and due to the invariance of λ_K, we have $h(xt', y) = h(x, y)$. It follows that the function $x \mapsto h(x, y)$ is in $\mathcal{C}(K\backslash G/K)$. Let us show that for every function $f \in \mathcal{K}(K\backslash G/K)$, we have

$$\int_G \check{f}(x)h(x, y)\, d\lambda_G(x) = (f * \omega)(y). \tag{$\dagger_3$}$$

Since $f * \omega \in \mathcal{C}(K\backslash G/K)$ and G is unimodular, by Fubini, we have

$$\int_G \check{f}(x)h(x, y)\, d\lambda_G(x) = \int_G \int_K \check{f}(x)\omega(xty)\, d\lambda_K(t)\, d\lambda_G(x)$$
$$= \int_G \int_K f(x)\omega(x^{-1}ty)\, d\lambda_G(x)\, d\lambda_K(t)$$
$$= \int_K (f * \omega)(ty)\, d\lambda_K(t) = \int_K (f * \omega)(y)\, d\lambda_K(t) = (f * \omega)(y).$$

If (2) holds, Equation (s_1), namely

$$\int_K \omega(xty)\, d\lambda_K(t) = \omega(x)\omega(y),$$

implies that $h(x, y) = \omega(x)\omega(y)$, so by $(\dagger_3)$ we have

$$(f * \omega)(y) = \int_G \check{f}(x)h(x, y)\, d\lambda_G(x) = \int_G \check{f}(x)\omega(x)\omega(y) d\lambda_G(x)$$
$$= \omega(y) \int_G \check{f}(x)\omega(x) d\lambda_G(x) = (\check{f}, \omega)\omega(y),$$

which proves (3).

Conversely, if (3) holds, we saw in proving that (3) implies (1) that $f * \omega = (\check{f}, \omega)\omega$. We prove that this implies that $h(x, y) = \omega(x)\omega(y)$, which is (s_1). For any function $g \in \mathcal{K}(G)$, by $(**)$, using the fact that $h, \omega \in \mathcal{C}(K\backslash G/K)$ and $(\dagger_3)$, we have

$$\begin{aligned}(g, h(-, y)) &= (g, h(-, y)^\sharp) = (g^\sharp, h(-, y)) = (\check{g}^\sharp * \omega)(y) \\ &= (g^\sharp, \omega)\omega(y) = (g, \omega^\sharp)\omega(y) = (g, \omega)\omega(y) = (g, \omega(y)\omega),\end{aligned}$$

and so $h(x, y) = \omega(x)\omega(y)$, as claimed. $\square$

Remark: If a bounded continuous function ω on G not equal to the zero function satisfies the equation

$$\int_K \omega(xty)\, d\lambda_K(t) = \omega(x)\omega(y)$$

for all $x, y \in G$, then it must belong to $\mathcal{C}(K\backslash G/K)$, and thus is a spherical function. Indeed, since λ_K is left and right invariant, for any $t' \in K$, we have

$$\omega(xt')\omega(y) = \omega(x)\omega(y) = \omega(x)\omega(t'y)$$

for all $x, y \in G$ an all $t' \in K$, which shows that $\omega(xt') = \omega(x)$ and $\omega(t'y) = \omega(y)$.

Definition 17.6. Let $\mathbf{S}(G/K)$, or simply $\mathbf{S}$, denote the space of spherical functions on G relative to K. This is a subspace of $\mathcal{C}(K\backslash G/K) \cap \mathrm{L}^\infty(G)$.

Let $A = \mathrm{L}^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$, a commutative, involutive, unital Banach algebra. In the degenerate case where $\delta_e \in \mathrm{L}^1(K\backslash G/K)$, the group G is a discrete group and δ_e invariant by translation by the elements of K. This implies that $K = \{e\}$, and that G is commutative and discrete. Otherwise, there is exactly one character of the algebra $A = \mathrm{L}^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$ whose restriction to $\mathrm{L}^1(K\backslash G/K)$ is the zero function.

Definition 17.7. The subspace of characters of the commutative involutive Banach algebra $A = \mathrm{L}^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$ whose restriction to $\mathrm{L}^1(K\backslash G/K)$ is not the zero function is denoted by $\mathsf{X}_0(A)$. This subspace is locally compact (metrizable and separable).

For every spherical function $\omega \in \mathbf{S}(G/K)$, let $\zeta_\omega \in \mathsf{X}_0(A)$ be the character given by

$$\zeta_\omega(f) = (f, \omega) = \int_G f(x)\omega(x)\, d\lambda_G(x), \qquad f \in \mathrm{L}^1(K\backslash G/K).$$

The map $\omega \mapsto \zeta_\omega$ is a bijection between $\mathbf{S}(G/K)$ and $\mathsf{X}_0(A)$. In fact, we have the following stronger result.

Theorem 17.7. *The following properties hold:*

(1) The map $\omega \mapsto \zeta_\omega$ is a homeomorphism of $\mathbf{S}(G/K)$ equipped with the induced topology of Fréchet space of $\mathcal{C}(G)$ (see Definition 2.18 and Proposition 2.23) onto $\mathsf{X}_0(A)$ equipped with the topology induced by the weak*-topology of the dual A' of A (see Definition 9.13). Consequently, $\mathbf{S}(G/K)$ is locally compact.

(2) Every compact subset L of $\mathbf{S}(G/K)$ is equicontinuous.

(3) The map $(x, \omega) \mapsto \omega(x)$ from $G \times \mathbf{S}(G/K)$ to $\mathbb{C}$ is continuous.

Proof sketch. The proof of (1) is given in Dieudonné [22] (Chapter XXII, Section 6, no. 22.6.9). This proof is very technical and makes use of the following results proven in Dieudonné [22] (Chapter XXII, Section 1, no. 22.1.11.2 and 22.1.11.2).

Proposition 17.8. *For every subset B of $\mathcal{C}(G)$ consisting of uniformly bounded functions, the topology induced by the weak*-topology on $\mathrm{L}^\infty(G)$ (see Definition 12.16) is coarser than the topology induced by the topology of $\mathcal{C}(G)$ as a Fréchet space.*

Using Proposition 17.8, the following result can be shown.

Proposition 17.9. *Let B be a subset of $\mathcal{C}(G)$ consisting of uniformly bounded functions and having the following property: for every $p_0 \in B$, for every compact subset K of G, for every $\epsilon > 0$, there is some neighborhood U of p_0 in B for the weak*-topology on $\mathrm{L}^\infty(G)$ and some compact neighborhood W of e in G, such that for every function $p \in U$, we have*

$$|(a^{-1}\chi_W * p)(s) - p(s)| \leq \epsilon \qquad \text{for all } s \in K,$$

where $a = \lambda_G(W)$. Then the topology on B induced by the weak*-topology on $\mathrm{L}^\infty(G)$ is identical to the topology induced by the topology of $\mathcal{C}(G)$ as a Fréchet space.

Proposition 17.9 follows from the following result also proven in Dieudonné [22] (Chapter XXII, Section 1, no. 22.1.11.5).

Proposition 17.10. *For every function $f \in \mathcal{L}^1(G)$ and for every bounded subset B of the Banach space $\mathrm{L}^\infty(G)$, the map $g \mapsto f * g$ is a continuous map from B equipped with the weak*-topology on $\mathrm{L}^\infty(G)$ to the Fréchet space $\mathcal{C}(G)$.*

The proof of Proposition 17.10 uses the trick that

$$\begin{aligned}(f * g)(s) &= \int_G f(t)g(t^{-1}s)\, d\lambda_G(t) = \int_G f(st)g(t^{-1})\, d\lambda_G(t) \\ &= \int_G (\lambda_{s^{-1}} f)(t)\check{g}(t)\, d\lambda_G(t) = (\check{g}, \lambda_{s^{-1}} f).\end{aligned}$$

(2) Pick any $x_0 \in G$. For every compact neighborhood V_0 of x_0, by definition of the Fréchet topology on $\mathcal{C}(G)$, the restriction map $f \mapsto f|V_0$ from $\mathcal{C}(G)$ to $\mathcal{C}(V; \mathbb{C})$ is continuous, thus the image L_0 of L under this map is compact. By Ascoli III (Theorem 2.14, Dieudonné's version), L_0 is equicontinuous. Consequently, for every $\epsilon > 0$ there is a neighborhood $V \subseteq V_0$ of x_0 such that $|\omega(x) - \omega(x_0)| \leq \epsilon$ for all $x \in V$ and all $\omega \in L$.

(3) Let (x_0, ω_0) be an element of $G \times \mathbf{S}(G/K)$. By (2), for every $\epsilon > 0$, there is a compact neigbhorhood V of x_0 in G and a compact neighborhood W of ω_0 in $\mathbf{S}(G/K)$, such that $|\omega(x) - \omega(x_0)| \leq \epsilon$ for all $x \in V$ and all $\omega \in W$. By definition of the Fréchet topology on $\mathcal{C}(G)$, there is a neighborhood $U \subseteq W$ of ω_0 in $\mathbf{S}(G/K)$ such that $|\omega(x) - \omega_0(x)| \leq \epsilon$ for all $x \in V$ and all $\omega \in U$. We deduce that

$$|\omega(x) - \omega_0(x_0)| \leq |\omega(x) - \omega(x_0)| + |\omega(x_0) - \omega_0(x_0)| \leq 2\epsilon$$

for all $x \in V$ and all $\omega \in U$. $\square$

A theory of spherical functions for Lie groups, in particular symmetric spaces, not based on Gelfand pairs but instead on certain invariant differential operators is discussed in Helgason [46] (Chapter 4).

In order to present some of the examples of Gelfand pairs involving Lie groups, we need to discuss some material about semi-simple Lie groups.

17.3 Real Forms of a Complex Semi-Simple Lie Algebra

This section assumes some background of Lie algebras and Lie groups. Such material is discussed extrensively in Carter, Segal and Macdonald [17], Dieudonné [21], Duistermaat and Kolk [29], Fulton and Harris [36], Gallier and Quaintance [38, 39], Hall [42], Helgason [47], Humphreys [51], Knapp [57, 56], Samelson [82], Serre [92, 91], and Varadarajan [98]. The most elementary presentations occur in Carter, Segal and Macdonald [17], Hall [42], and Gallier and Quaintance [38]. We need to review the process of "complexifying" a real vector space V. But first we recall how to view a complex vector space as a real vector space.

Definition 17.8. If V is a complex vector space, then we denote by $V|_{\mathbb{R}}$ the vector space whose underlying abelian group is V, but with the scalar multiplication restricted to $\mathbb{R}$.

We can define the complexification $V_{\mathbb{C}}$ of the real vector space V as the complex vector space whose carrier is the tensor product $\mathbb{C} \otimes_{\mathbb{R}} V$, but more directly as $V \times V$, with the addition operation

$$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1, +v_2), \qquad u_1, u_2, v_1, v_2 \in V,$$

and the scalar product given by

$$(a + ib)(u, v) = (au - bv, av + bu), \qquad a, b \in \mathbb{R}, u, v \in V.$$

Observe that

$$(0, v) = i(v, 0),$$

so we can write

$$(u, v) = (u, 0) + i(v, 0),$$

and if $j\colon V \to V_{\mathbb{C}}$ is the injection given by $j(u) = (u, 0)$, for all $u \in V$, then we have an isomorphism

$$V_{\mathbb{C}} \cong V \oplus iV,$$

as a direct sum of real subspaces. More precisely, the injection j induces an isomorphism

$$(V_{\mathbb{C}})|_{\mathbb{R}} \cong V \oplus iV,$$

but by abuse of notation, we usualy write $V_{\mathbb{C}} = V \oplus iV$. Using the above isomorphism, the scalar multiplication of a vector $u + iv \in V_{\mathbb{C}}$ by a complex number $a + ib$ is given by

$$(a + ib)(u + iv) = au - bv + i(av + bu).$$

The map $c_V\colon V_{\mathbb{C}} \to V_{\mathbb{C}}$ given by

$$c_V(u + iv) = u - iv, \qquad u, v \in V,$$

is semi-linear and an involution; that is, $c_V^2 = \mathrm{id}_{V_{\mathbb{C}}}$. The map c_V is called the *conjugation* of $V_{\mathbb{C}}$ associated with V. Observe that

$$V = \{w \in V_{\mathbb{C}} \mid c_V(w) = w\}, \quad iV = \{w \in V_{\mathbb{C}} \mid c_V(w) = -w\}.$$

To simplify notation we usually write c instead of c_V.

If $\mathfrak{g}$ is a real Lie algebra, then its complexification $\mathfrak{g}_{\mathbb{C}}$ is the complex Lie algebra whose carrier is the complex vector space such that $(\mathfrak{g}_{\mathbb{C}})|_{\mathbb{R}} = \mathfrak{g} \oplus i\mathfrak{g}$ as a direct sum of real subspaces, with the Lie bracket given by

$$[u + iv, x + iy]_{\mathbb{C}} = [u, x] - [v, y] + i([u, y] + [v, x]).$$

If c is the conjugation of $\mathfrak{g}_{\mathbb{C}}$ associated with $\mathfrak{g}$, then

$$\begin{aligned} c([u + iv, x + iy]_{\mathbb{C}}) &= c([u, x] - [v, y] + i([u, y] + [v, x])) \\ &= c([u, x]) - c([v, y]) - i(c([u, y]) + c([v, x])) \\ &= [u, x] - [v, y] - i([u, y] + [v, x]) \\ &= [u - iv, x - iy]_{\mathbb{C}} = [c(u + iv), c(x + iy)]_{\mathbb{C}}. \end{aligned}$$

This shows that c is an automorphism of the real Lie algebra $\mathfrak{g}_{\mathbb{C}}|_{\mathbb{R}}$.

Definition 17.9. Given a complex Lie algebra $\mathfrak{g}$, a real Lie algebra $\mathfrak{g}_0$ such that

$$\mathfrak{g}|_{\mathbb{R}} \cong \mathfrak{g}_0 \oplus i\mathfrak{g}_0$$

as a direct sum of real subspaces is called a *real form* of $\mathfrak{g}$. The complex Lie algebra $\mathfrak{g}$ is the *complexification* of $\mathfrak{g}_0$.

The following proposition shows that finding a real form of a complex Lie algebra $\mathfrak{g}$ is equivalent to finding an automorphism c of the real Lie algebra $\mathfrak{g}|_{\mathbb{R}}$ satisfying the properties stated in the following proposition.

Proposition 17.11. *Let $\mathfrak{g}$ be a complex Lie algebra, $\mathfrak{g}_0$ be a real Lie algebra, and assume that $\mathfrak{g}|_{\mathbb{R}} = \mathfrak{g}_0 \oplus i\mathfrak{g}_0$; that is, $\mathfrak{g}$ is the complexification of $\mathfrak{g}_0$. Then the conjugation $c_0 \colon \mathfrak{g} \to \mathfrak{g}$ given by $c_0(u + iv) = u - iv$ ($u, v \in \mathfrak{g}_0$) has the following properties:*

(a) The map c_0 is semi-linear, which means that $c_0(x + y) = c_0(x) + c_0(y)$, and $c_0(ix) = -ic_0(x)$, for all $x, y \in \mathfrak{g}$.

(b) The map c_0 is idempotent; that is, $c_0^2 = \mathrm{id}_{\mathfrak{g}}$.

(c) We have

$$c_0([x, y]) = [c_0(x), c_0(y)], \qquad \text{for all } x, y \in \mathfrak{g}.$$

Conversely, if a map $c \colon \mathfrak{g} \to \mathfrak{g}$ has Properties (a), (b), (c), then if we consider the linear automorphism $c \colon \mathfrak{g}|_{\mathbb{R}} \to \mathfrak{g}|_{\mathbb{R}}$ of $\mathfrak{g}$ viewed as a real vector space, and if we let

$$\mathfrak{g}_1 = \{x \in \mathfrak{g} \mid c(x) = x\},$$

the subspace $\mathfrak{g}_1$ is a real Lie subalgebra of $\mathfrak{g}|_{\mathbb{R}}$, and we have

$$\mathfrak{g}|_{\mathbb{R}} = \mathfrak{g}_1 \oplus i\mathfrak{g}_1;$$

that is, $\mathfrak{g}$ is the complexification of $\mathfrak{g}_1$.

Proof. We already proved the first part of the proposition. Conversely, since c is an involutive automorphism of $\mathfrak{g}|_{\mathbb{R}}$, we know by linear algebra that

$$\mathfrak{g}|_{\mathbb{R}} = \mathfrak{g}_1 \oplus \mathfrak{g}_{-1}$$

where $\mathfrak{g}_1$ and $\mathfrak{g}_{-1}$ are the real eigenspaces of c given by

$$\mathfrak{g}_1 = \{x \in \mathfrak{g} \mid c(x) = x\}, \qquad \mathfrak{g}_{-1} = \{x \in \mathfrak{g} \mid c(x) = -x\}.$$

Since c is semi-linear, for every $x \in \mathfrak{g}_1$, we have

$$c(ix) = -ic(x) = -ix,$$

so $ix \in \mathfrak{g}_{-1}$, which shows that

$$i\mathfrak{g}_1 \subseteq \mathfrak{g}_{-1}. \tag{$*$}$$

For every $x \in \mathfrak{g}_{-1}$, we have

$$c(ix) = -ic(x) = ix$$

so $ix \in \mathfrak{g}_1$, which shows that

$$i\mathfrak{g}_{-1} \subseteq \mathfrak{g}_1.$$

But the above inclusion implies that

$$\mathfrak{g}_{-1} \subseteq i\mathfrak{g}_1. \tag{$**$}$$

By ($*$) and ($**$), we get

$$\mathfrak{g}_{-1} = i\mathfrak{g}_1,$$

and so

$$\mathfrak{g}|_{\mathbb{R}} = \mathfrak{g}_1 \oplus i\mathfrak{g}_1,$$

as claimed. Since c is the identity on $\mathfrak{g}_1$ and satisfies (c), we conclude that $\mathfrak{g}_1$ is a (real) subalgebra of $\mathfrak{g}$. $\square$

Definition 17.10. Given a complex Lie algebra $\mathfrak{g}$, a map $c_0\colon \mathfrak{g} \to \mathfrak{g}$ satisfying Conditions (a), (b), (c) of Proposition 17.11 is called a *conjugation*.

Recall that a semi-simple Lie algebra has no commutative ideals other than (0). A Lie group is semi-simple if its Lie algebra is semi-simple; see Gallier and Quaintance [38], Section 21.5, Definition 21.8. If the reader is familiar with the notion of Killing form, Cartan's criterion says that a Lie algebra is semi-simple iff its Killing form is nondegenerate; see Gallier and Quaintance [38], Section 21.6, Theorem 21.26, Knapp [57], and Serre [92].

Let G_u be a real compact semi-simple simply-connected Lie group, let $\mathfrak{g}_u$ be its real semi-simple Lie algebra, and let $\mathfrak{g} = \mathfrak{g}_u \oplus i\mathfrak{g}_u$ be the complex Lie algebra which is its complexification. Then it is possible to determine all the real forms $\mathfrak{g}_0$ of $\mathfrak{g}$ (up to isomorphism).

Recall that the conjugation $c_u\colon \mathfrak{g} \to \mathfrak{g}$ associated with $\mathfrak{g}_u$ is given by

$$c_u(x + iy) = x - iy, \qquad x, y \in \mathfrak{g}_u.$$

It can be proven that in order to find all conjugations $c_0\colon \mathfrak{g} \to \mathfrak{g}$ of $\mathfrak{g}$, it suffices to consider conjugations that commute with c_u; see Dieudonné [21] (Chapter XXI, no. 21.18.3). The key to the proof is that the Killing form associated with the Lie algebra of a compact semi-simple connected Lie group is negative definite; see Gallier and Quaintance [38], Section 21.6, Theorem 21.27. Technically, we have the following result.

Proposition 17.12. *Let G_u be a real compact semi-simple simply-connected Lie group, let $\mathfrak{g}_u$ be its real semi-simple Lie algebra, and let $\mathfrak{g} = \mathfrak{g}_u \oplus i\mathfrak{g}_u$ be its complexification. For any conjugation c of $\mathfrak{g}$, there is an automorphism $\varphi\colon \mathfrak{g} \to \mathfrak{g}$ such that c_u and $\varphi \circ c \circ \varphi^{-1}$ commute.*

Proof. The proof assumes some familiarity with the properties of the KiIling form and may be safely omitted. The Killing form is the bilinear form on $\mathfrak{g}$ defined by

$$B_{\mathfrak{g}}(u, v) = \mathrm{tr}(\mathrm{ad}_u \circ \mathrm{ad}_v), \quad u, v \in \mathfrak{g},$$

where $\mathrm{ad}_u(w) = [u, w]$. The first fact is that the mapping

$$(u, v) \mapsto \langle u, v\rangle = -B_{\mathfrak{g}}(u, c_u(v))$$

is a Hermitian inner product on $\mathfrak{g}$, where $B_{\mathfrak{g}}$ is the Killing form of $\mathfrak{g}$; see Dieudonné [21] (Chapter XXI, no. 21.17.2.1). Consider the map $h = c \circ c_u$. For simplicity of notation we suppress the composition symbol. Since c and c_u are semi-linear maps that preserve the Lie bracket, $h = cc_u$ is actually linear, and since $h^{-1} = (cc_u)^{-1} = c_u c$, we see that cc_u is an automorphism of the Lie algebra $\mathfrak{g}$. We will prove that h is self-adjoint with respect to the inner product $\langle -, - \rangle$. As a consequence, h is diagonalizable and has real eigenvalues, so $S = h^2$ has strictly positive eigenvalues. Then we will see that $\varphi = S^{-1/4}$ does the job.

First observe that

$$h^{-1} c_u = c_u c c_u = c_u h.$$

Since h and h^{-1} preserve the Lie bracket, they also preserves the Killing form so we have

$$\begin{aligned}
\langle h(u), v \rangle &= -B_{\mathfrak{g}}(h(u), c_u(v)) \\
&= -B_{\mathfrak{g}}(u, h^{-1}(c_u(v)) \\
&= -B_{\mathfrak{g}}(u, c_u(h(v))) \\
&= \langle u, h(v) \rangle,
\end{aligned}$$

which shows that h is self-adjoint. Thus h is diagonalizable with respect to a an orthonormal basis of eigenvectors and its eigenvalues are real. But then $S = h^2$ is a self-adjoint linear map with strictly positive eigenvalues, so with respect to an orthonormal basis $(e_1, \ldots, e_n)$ of eigenvectors, S is represented by a diagonal matrix

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

with $\lambda_i > 0$ for $i = 1, \ldots, n$. For any real number $t > 0$, define the matrix Λ^t as

$$\Lambda^t = \mathrm{diag}(\lambda_1^t, \ldots, \lambda_n^t).$$

Obviously the linear isomorphisms S^t represented by the matrices Λ^t commute with h. We claim that they are also Lie algebra isomorphisms of $\mathfrak{g}$. The Lie bracket on $\mathfrak{g}$ is determined by its values on the basis $(e_1, \ldots, e_n)$, that is, by equations

$$[e_j, e_k] = \sum_{l=1}^{n} a_{jkl} e_l, \quad 1 \leq j, k \leq n,$$

for some $a_{jkl} \in \mathbb{C}$. To express that S is an automorphism of $\mathfrak{g}$ is equivalent to stating the equations

$$[S(e_j), S(e_k)] = S([e_j, e_k]), \quad 1 \leq j, k \leq n,$$

and since

$$[S(e_j), S(e_k)] = \lambda_j \lambda_k [e_j, e_k], \qquad S(e_l) = \lambda_l e_l,$$

to stating the equations

$$\lambda_j \lambda_k \sum_{l'=1}^{n} a_{jkl'} e_{l'} = \sum_{l=1}^{n} \lambda_l a_{jkl} e_l,$$

that is, to the equations

$$\lambda_j \lambda_k a_{jkl} = a_{jkl} \lambda_l, \quad 1 \leq j, k, l \leq n. \tag{$*_1$}$$

These equations are nontrivial only when $a_{jkl} \neq 0$, in which case

$$\lambda_j \lambda_k = \lambda_l.$$

These equations imply that

$$\lambda_j^t \lambda_k^t = \lambda_l^t$$

for all $t \geq 0$, so

$$\lambda_j^t \lambda_k^t a_{jkl} = a_{jkl} \lambda_l^t, \quad 1 \leq j, k, l \leq n,$$

which shows that the S^t are Lie algebra isomorphisms of $\mathfrak{g}$. If we let $S^{-t} = (S^t)^{-1}$ for $t > 0$, consider the conjugation of $\mathfrak{g}$ given by

$$c^{(t)} = S^t c_u S^{-t}.$$

Beware that in general $c^{(1)} \neq c$, this is why we use the notation $c^{(t)}$ instead of c^t. Observe that since $h = cc_u$,

$$c_u h c_u^{-1} = c_u c c_u c_u^{-1} = c_u c = h^{-1}$$

so

$$c_u h c_u^{-1} c_u h c_u^{-1} = h^{-1} h^{-1},$$

namely

$$c_u h^2 c_u^{-1} = h^{-2},$$

and since $S = h^2$, we get $c_u S c_u^{-1} = S^{-1}$. Since $c_u^2 = \mathrm{id}$, the equation $c_u S c_u^{-1} = S^{-1}$ is equivalent to

$$S c_u = c_u S^{-1}. \tag{$*_2$}$$

We prove that the above equation implies that

$$S^t c_u = c_u S^{-t} \quad \text{for all } t > 0. \tag{$*_3$}$$

The map c_u is linear over $\mathbb{R}$, so we can express $(*_2)$ in terms of matrices over the basis $(e_1, \ldots, e_n)$. If (c_{ij}) is the matrix representing c_u, we know that S is represented by the diagonal matrix Λ, so $(*_2)$ is equivalent to the equations

$$\lambda_i c_{ij} = c_{ij} \lambda_j^{-1}, \qquad 1 \leq i, j \leq n. \tag{$*_4$}$$

These equations are trivially satisfied if $c_{ij} = 0$ and otherwise they imply

$$\lambda_i = \lambda_j^{-1},$$

which in turn imply

$$\lambda_i^t = \lambda_j^{-t}, \qquad \text{for all } t > 0,$$

and thus

$$\lambda_i^t c_{ij} = c_{ij}\lambda_j^{-t}, \qquad 1 \leq i, j \leq n, \tag{$*_5$}$$

which means that

$$S^t c_u = c_u S^{-t},$$

as claimed. Using the equation $S^t c_u = c_u S^{-t}$, we have

$$\begin{aligned} cc^{(t)} &= cS^t c_u S^{-t} = cc_u S^{-2t} = hS^{-2t} \\ c^{(t)} c &= (cc^{(t)})^{-1} = S^{2t} h^{-1} = h^{-1} S^{2t}. \end{aligned}$$

If we let $t = 1/4$, then

$$cc^{(t)} = c^{(t)} c = h^{-1} S^{1/2},$$

since $S = h^2$ and so

$$hS^{-1/2} = hS^{-1} S^{1/2} = hh^{-2} S^{1/2} = h^{-1} S^{1/2}.$$

Therefore, with $\varphi = S^{-1/4}$, we see that $c^{(1/4)} = S^{1/4} c_u S^{-1/4} = \varphi^{-1} \circ c_u \circ \varphi$ commutes with c, namely

$$c \circ \varphi^{-1} \circ c_u \circ \varphi = \varphi^{-1} \circ c_u \circ \varphi \circ c,$$

which implies

$$\varphi \circ c \circ \varphi^{-1} \circ c_u = c_u \circ \varphi^{-1} \circ c \circ \varphi^{-1},$$

that is, c_u and $\varphi \circ c \circ \varphi^{-1}$ commute.

Beware that even though $S = h^2$, in general, $S^{1/2} \neq h$ because h may have some negative eigenvalues, but S is positive definite and so are all of its powers S^t. $\square$

Let $c_0 \colon \mathfrak{g} \to \mathfrak{g}$ be a conjugation of $\mathfrak{g}$ that commutes with c_u. In this case, for any $x \in \mathfrak{g}_u$, since

$$c_0(x) = c_0(c_u(x)) = c_u(c_0(x)),$$

we see that $c_0(x) \in \mathfrak{g}_u$, so $\mathfrak{g}_u$ is invariant under c_0, and similarly, for any $x \in \mathfrak{g}_u$, since

$$c_0(ix) = -c_0(c_u(ix)) = -c_u(c_0(ix)),$$

so $i\mathfrak{g}_u$ is also invariant under c_0.

Since the restriction of c_0 to $\mathfrak{g}_u$ is an involutive automorphism of $\mathfrak{g}_u$, we know by linear algebra that

$$\mathfrak{g}_u = E_1 \oplus E_{-1},$$

where E_1 and E_{-1} are the real eigenspaces of c_0 given by

$$E_1 = \{x \in \mathfrak{g}_u \mid c_0(x) = x\}, \qquad E_{-1} = \{x \in \mathfrak{g}_u \mid c_0(x) = -x\}.$$

It is customary to denote E_1 by $\mathfrak{k}_0$ and E_{-1} by $i\mathfrak{p}_0$, where both are real vector spaces, so that

$$\mathfrak{g}_u = \mathfrak{k}_0 \oplus i\mathfrak{p}_0,$$

and $i\mathfrak{g}_u = i\mathfrak{k}_0 \oplus \mathfrak{p}_0$. Since c_0 is semi-linear, for any $ix \in i\mathfrak{p}_0$, we have

$$-ix = c_0(ix) = -ic_0(x),$$

so $c_0(x) = x$ if $x \in \mathfrak{p}_0$, and for $ix \in i\mathfrak{k}_0$, we have

$$c_0(ix) = -ic_0(x) = -ix,$$

since $c_0(x) = x$ for $x \in \mathfrak{k}_0$. Consequently the (real) Lie algebra $\mathfrak{g}_0$ of fixed points of c_0 is

$$\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0,$$

and

$$\mathfrak{g} = \mathfrak{g}_0 + i\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0 \oplus i\mathfrak{k}_0 \oplus i\mathfrak{p}_0 = \mathfrak{k}_0 \oplus i\mathfrak{p}_0 \oplus i\mathfrak{k}_0 \oplus \mathfrak{p}_0 = \mathfrak{g}_u \oplus i\mathfrak{g}_u,$$

so $\mathfrak{g}_0$ is a semi-simple Lie algebra.

Definition 17.11. The decomposition

$$\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0$$

is called a *Cartan decomposition* of the real Lie algebra $\mathfrak{g}_0$ (with respect to the conjugation c_0).

Next we give several examples of Cartan decompositions. The group

$$G = \mathbf{SL}(n, \mathbb{C}) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid \det(X) = 1\}$$

is one of the simplest and most important example of a complex semi-simple Lie group and the Lie group

$$G_u = \mathbf{SU}(n) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid XX^* = X^*X = I_n,\ \det(X) = 1\}$$

is one of the the simplest and most important example of a real semi-simple Lie group, so we use these groups in our examples. The Lie group $\mathbf{SL}(n, \mathbb{C})$ has Lie algebra $\mathfrak{sl}(n, \mathbb{C})$ and the Lie group $\mathbf{SU}(n)$ has Lie algebra $\mathfrak{su}(n)$, both defined in the next section.

17.4 Examples of Cartan Decompositions

Example 17.1. Consider the real Lie algebra $\mathfrak{g}_u = \mathfrak{su}(n)$ of $n \times n$ complex skew-hermitian matrices with zero trace,

$$\mathfrak{su}(n) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* = -X,\ \mathrm{tr}(X) = 0\}.$$

We claim that the complexification $\mathfrak{g} = \mathfrak{su}(n)_{\mathbb{C}}$ of $\mathfrak{su}(n)$ is the complex Lie algebra $\mathfrak{sl}(n, \mathbb{C})$ of $n \times n$ complex matrices with zero trace,

$$\mathfrak{sl}(n, \mathbb{C}) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid \mathrm{tr}(X) = 0\}.$$

First observe that $i\mathfrak{su}(n)$ is the (real) vector space (not a Lie algebra) of hermitian matrices with zero trace,

$$i\mathfrak{su}(n) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* = X,\ \mathrm{tr}(X) = 0\}.$$

Indeed, if $X^* = -X$, then $(iX)^* = -iX^* = iX$, and if $\mathrm{tr}(X) = 0$, then $\mathrm{tr}(iX) = i\mathrm{tr}(X) = 0$, so

$$i\mathfrak{su}(n) \subseteq \{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* = X,\ \mathrm{tr}(X) = 0\}.$$

Conversely, if $X^* = X$ then $(iX)^* = -iX^* = -iX$, and if $\mathrm{tr}(X) = 0$, then $\mathrm{tr}(iX) = i\mathrm{tr}(X) = 0$, so

$$i\{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* = X,\ \mathrm{tr}(X) = 0\} \subseteq \mathfrak{su}(n).$$

But the above equation implies that

$$\{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* = X,\ \mathrm{tr}(X) = 0\} \subseteq i\mathfrak{su}(n),$$

so

$$i\mathfrak{su}(n) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* = X,\ \mathrm{tr}(X) = 0\}.$$

Every complex matrix $X \in \mathrm{M}_n(\mathbb{C})$ can be written as

$$X = \frac{1}{2}(X + X^*) + \frac{1}{2}(X - X^*),$$

and we have

$$\frac{1}{2}(X + X^*)^* = \frac{1}{2}(X^* + X^{**}) = \frac{1}{2}(X^* + X) = \frac{1}{2}(X + X^*),$$

and

$$\frac{1}{2}(X - X^*)^* = \frac{1}{2}(X^* - X^{**}) = \frac{1}{2}(X^* - X) = -\frac{1}{2}(X - X^*).$$

Also, if $\mathrm{tr}(X) = 0$, then

$$\mathrm{tr}\left(\frac{1}{2}(X + X^*)\right) = \frac{1}{2}\left(\mathrm{tr}(X) + \mathrm{tr}(X^*)\right) = \frac{1}{2}\left(\mathrm{tr}(X) + \mathrm{tr}(X)\right) = 0,$$

and

$$\mathrm{tr}\left(\frac{1}{2}(X - X^*)\right) = \frac{1}{2}\left(\mathrm{tr}(X) - \mathrm{tr}(X^*)\right) = \frac{1}{2}\left(\mathrm{tr}(X) - \mathrm{tr}(X)\right) = 0.$$

Thus if $X \in \mathfrak{sl}(n, \mathbb{C})$, then $\frac{1}{2}(X + X^*) \in \mathfrak{su}(n)$ and $\frac{1}{2}(X - X^*) \in i\mathfrak{su}(n)$, which proves that

$$\mathfrak{sl}(n, \mathbb{C}) = \mathfrak{su}(n) \oplus i\mathfrak{su}(n). \tag{$*$}$$

The sum is a direct sum because the only matrix such that $X^* = -X$ and $X^* = X$ is the matrix $X = 0$.

Since $\mathfrak{su}(n) = \{X \in \mathfrak{sl}(n, \mathbb{C}) \mid X^* = -X\}$, the congugation c_u of $\mathfrak{sl}(n, \mathbb{C})$ associated with $\mathfrak{su}(n)$ is the map given by $c_u(X) = -X^*$.

We now consider three types of conjugations on $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$ which lead to interesting real forms of $\mathfrak{sl}(n, \mathbb{C})$.

Example 17.2. Consider the conjugation c_0 of $\mathfrak{sl}(n, \mathbb{C})$ given by $c_0(X) = \overline{X}$. Obviously, c_0 commutes with c_u. The restriction of c_0 to $\mathfrak{g}_u = \mathfrak{su}(n)$ is also c_0, and we obtain

$$\begin{aligned} \mathfrak{k}_0 &= \mathfrak{so}(n) = \{X \in \mathfrak{su}(n) \mid \overline{X} = X\} \\ i\mathfrak{p}_0 &= \{X \in \mathfrak{su}(n) \mid \overline{X} = -X\}, \end{aligned}$$

where $\mathfrak{so}(n)$ is the Lie algebra of $n \times n$ real skew-symmetric matrices.

But for any $X = (x_{jk}) \in \mathfrak{su}(n)$, if $x_{jk} = a_{jk} + ib_{jk}$, with $a_{jk}, b_{jk} \in \mathbb{R}$, since $X^* = -X$, we have $x_{kj} = -a_{jk} + ib_{jk}$, so $a_{kk} = 0$. If we also have $\overline{X} = -X$, then $a_{jk} + ib_{jk} = -a_{jk} + ib_{jk}$, so $a_{jk} = 0$ for all j, k, which means that $X = (ib_{jk})$, with (b_{jk}) a real symmetric matrix. Thus

$$\mathfrak{p}_0 = \mathfrak{s}(n) = \{X \in \mathrm{M}_n(\mathbb{R}) \mid X^\top = X, \, \mathrm{tr}(X) = 0\},$$

the vector space of $n \times n$ real symmetric matrices with zero trace (not a Lie algebra). We obtain

$$\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0 = \mathfrak{so}(n) \oplus \mathfrak{s}(n) = \mathfrak{sl}(n, \mathbb{R}),$$

with

$$\mathfrak{sl}(n, \mathbb{R}) = \{X \in \mathrm{M}_n(\mathbb{R}) \mid \mathrm{tr}(X) = 0\}.$$

It turns out that a Gelfand pair arises from two Lie groups G_0 and K_0 whose Lie algebras are $\mathfrak{sl}(n, \mathbb{R})$ and $\mathfrak{sl}(n, \mathbb{R}) \cap \mathfrak{su}(n) = \mathfrak{so}(n)$. To describe these groups, first we need to consider the complex simply-connected Lie group $G = \mathbf{SL}(n, \mathbb{C})$ whose complex Lie algebra is $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$.

Definition 17.12. A complex Lie group G of (complex) dimension n can also be viewed as a real Lie group of (real) dimension $2n$ denoted $G|_{\mathbb{R}}$, by viewing a holomorphic chart $\varphi\colon U \to \mathbb{C}^n$ of G as a real smooth function $\varphi\colon U \to \mathbb{R}^{2n}$. More generally, a complex manifold M of (complex) dimension n can be viewed as a real manifold of dimension $2n$ denoted $M|_{\mathbb{R}}$.

Using the correspondence between simply-connected real Lie groups and real Lie algebras (see Gallier and Quaintance [38], Section 19.4), there is a unique automorphism $\sigma_0\colon \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}} \to \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}}$ such that $d(\sigma_0)_I = c_0$, where $c_0\colon \mathfrak{sl}(n, \mathbb{C})_{\mathbb{R}} \to \mathfrak{sl}(n, \mathbb{C})|_{\mathbb{R}}$ is the map $c_0(X) = \overline{X}$, and σ_0 is also given by

$$\sigma_0(X) = \overline{X}.$$

The real Lie group G_0 is the set of fixed points of $\mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}}$ under the automorphism σ_0, given by

$$G_0 = \{X \in \mathbf{SL}(n, \mathbb{C})_{\mathbb{R}} \mid \overline{X} = X\} = \mathbf{SL}(n, \mathbb{R}).$$

Note that the Lie algebra of $G_0 = \mathbf{SL}(n, \mathbb{R})$ is $\mathfrak{g}_0 = \mathfrak{sl}(n, \mathbb{R})$. The simply-connected real Lie group G_u whose Lie algebra is $\mathfrak{su}(n)$ is $G_u = \mathbf{SU}(n)$. We define K_0 by

$$K_0 = G_0 \cap G_u = \mathbf{SL}(n, \mathbb{R}) \cap \mathbf{SU}(n) = \mathbf{SO}(n).$$

Note that

$$\begin{aligned}\mathbf{SL}(n, \mathbb{R}) &= \{X \in \mathrm{M}_n(\mathbb{R}) \mid \det(X) = 1\} \\ \mathbf{SO}(n) &= \{X \in \mathrm{M}_n(\mathbb{R}) \mid XX^\top = X^\top A = I_n,\ \det(X) = 1\},\end{aligned}$$

and that the Lie algebra of the compact Lie group $K_0 = \mathbf{SO}(n)$ is

$$\mathfrak{k}_0 = \mathfrak{so}(n) = \{X \in \mathrm{M}_n(\mathbb{R}) \mid X^\top = -X\}.$$

The following paragraph is meant for readers well acquainted with Lie groups and Lie algebras and can be safely omitted. The real Lie group $G_0 = \mathbf{SL}(n, \mathbb{R})$ is semi-simple and the real Lie group $K_0 = \mathbf{SO}(n)$ is semi-simple for $n \geq 3$. These groups are connected but not simply-connected. For $n = 2$, the universal cover of $\mathbf{SL}(2, \mathbb{R})$ is $\mathbb{R}^3$ and the universal cover of $\mathbf{SO}(2)$ is $\mathbb{R}$. For $n \geq 3$, the universal cover $\widetilde{G}_0 = \widetilde{\mathbf{SL}}(n, \mathbb{R})$ of $G_0 = \mathbf{SL}(n, \mathbb{R})$ is not a matrix group and the universal cover of $K_0 = \mathbf{SO}(n)$ is $\widetilde{K}_0 = \mathbf{Spin}(n)$. The real semi-simple (connected) Lie group $G_0 = \mathbf{SL}(n, \mathbb{R})$ is called a *real form* of the complex semi-simple (simply-connected) Lie group $G = \mathbf{SL}(n, \mathbb{C})$. We will show later that the pair $(G_0, K_0) = (\mathbf{SL}(n, \mathbb{R}), \mathbf{SO}(n))$ is a Gelfand pair.

Example 17.3. Again, consider the real Lie algebra $\mathfrak{g}_u = \mathfrak{su}(n)$ of $n \times n$ complex skew-hermitian matrices with zero trace. In this example we also need the Lie algebra $\mathfrak{u}(n)$ of the real Lie group

$$\mathbf{U}(n) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid XX^* = X^*X = I_n\},$$

given by

$$\mathfrak{u}(n) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* = -X\}.$$

In other words, $\mathfrak{u}(n)$ consists of all skew-Hermitian complex matrices. Observe that

$$\mathfrak{su}(n) = \{X \in \mathfrak{u}(n) \mid \mathrm{tr}(X) = 0\} = \mathfrak{u}(n) \cap \mathfrak{sl}(n, \mathbb{C}).$$

We showed in Example 17.2 that the complexification $\mathfrak{su}(n)_\mathbb{C}$ of $\mathfrak{su}(n)$ is the complex Lie algebra $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$ of $n \times n$ complex matrices with zero trace. This time, let $c_0 \colon \mathfrak{sl}(n, \mathbb{C}) \to \mathfrak{sl}(n, \mathbb{C})$ be the conjugation given by

$$c_0(X) = -I_{p,n-p} X^* I_{p,n-p},$$

where

$$I_{p,n-p} = \begin{pmatrix} I_p & 0_{p,n-p} \\ 0_{n-p,p} & -I_{p,n-p} \end{pmatrix},$$

with $1 \leq p \leq n-1$. Obviously $I_{p,n-p}^* = I_{p,n-p}$ and $I_{p,n-p}^2 = I_n$, and c_0 commutes with c_u (given by $c_u(X) = -X^*$). Since matrices in $\mathfrak{su}(n)$ satisfy the property $X^* = -X$, the restriction of c_0 to $\mathfrak{su}(n)$ is given by $c_0(X) = I_{p,n-p} X I_{p,n-p}$. If we wite

$$X = \begin{pmatrix} U & B \\ A & V \end{pmatrix},$$

where $U \in \mathrm{M}_p(\mathbb{C}), V \in \mathrm{M}_{n-p}(\mathbb{C}), A \in \mathrm{M}_{n-p,p}(\mathbb{C}), B \in \mathrm{M}_{p,n-p}(\mathbb{C})$, then

$$I_{p,n-p} X I_{p,n-p} = \begin{pmatrix} I_p & 0_{p,n-p} \\ 0_{n-p,p} & -I_{n-p} \end{pmatrix} \begin{pmatrix} U & B \\ A & V \end{pmatrix} \begin{pmatrix} I_p & 0_{p,n-p} \\ 0_{n-p,p} & -I_{n-p} \end{pmatrix} = \begin{pmatrix} U & -B \\ -A & V \end{pmatrix}.$$

Therefore

$$\begin{aligned} \mathfrak{k}_0 &= \{X \in \mathfrak{su}(n) \mid I_{p,n-p} X I_{p,n-p} = X\} \\ &= \left\{ \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \;\middle|\; U^* = -U,\ V^* = -V,\ \mathrm{tr}(U) + \mathrm{tr}(V) = 0 \right\} \\ &= \left\{ \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \;\middle|\; U \in \mathfrak{u}(p),\ V \in \mathfrak{u}(n-p),\ \mathrm{tr}(U) + \mathrm{tr}(V) = 0 \right\}, \end{aligned}$$

and

$$\begin{aligned} i\mathfrak{p}_0 &= \{X \in \mathfrak{su}(n) \mid -I_{p,n-p} X I_{p,n-p} = X\} \\ &= \left\{ \begin{pmatrix} 0 & -A^* \\ A & 0 \end{pmatrix} \;\middle|\; A \in \mathrm{M}_{n-p,p}(\mathbb{C}) \right\}, \end{aligned}$$

so

$$\begin{aligned} \mathfrak{p}_0 &= \left\{ \begin{pmatrix} 0 & -iA^* \\ iA & 0 \end{pmatrix} \;\middle|\; A \in \mathrm{M}_{n-p,p}(\mathbb{C}) \right\} \\ &= \left\{ \begin{pmatrix} 0 & (iA)^* \\ iA & 0 \end{pmatrix} \;\middle|\; A \in \mathrm{M}_{n-p,p}(\mathbb{C}) \right\} \\ &= \left\{ \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix} \;\middle|\; A \in \mathrm{M}_{n-p,p}(\mathbb{C}) \right\}. \end{aligned}$$

Consequently the real Lie algebra $\mathfrak{g}_0$ corresponding to the conjugation c_0 is given by

$$\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0 = \left\{ \begin{pmatrix} U & A^* \\ A & V \end{pmatrix} \;\middle|\; U \in \mathfrak{u}(p),\ V \in \mathfrak{u}(n-p),\ A \in \mathrm{M}_{n-p,p}(\mathbb{C}),\ \mathrm{tr}(U) + \mathrm{tr}(V) = 0 \right\}.$$

Thus $\mathfrak{g}_0 = \mathfrak{su}(p, n-p)$, the Lie algebra of the real Lie group $\mathbf{SU}(p, n-p)$ defined by

$$\mathbf{SU}(p, n-p) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* I_{p,n-p} X = I_{p,n-p}, \quad \det(X) = 1\}.$$

Given any $p \times p$ matrix U, if $\alpha = \mathrm{tr}(U)$, we let

$$U_1 = U - \frac{\alpha}{p} I_p$$

and then we have

$$\mathrm{tr}(U_1) = \mathrm{tr}(U) - p\frac{\alpha}{p} = \alpha - \alpha = 0,$$

so we can write

$$U = U_1 + \frac{\alpha}{p} I_p$$

with $\mathrm{tr}(U_1) = 0$, and since $\mathrm{tr}(U) + \mathrm{tr}(V) = 0$, we also let

$$V_1 = V + \frac{\alpha}{n-p} I_{n-p},$$

so that $\mathrm{tr}(V_1) = 0$ and

$$V = V_1 - \frac{\alpha}{n-p} I_{n-p},$$

which shows that we can also write every matrix $X \in \mathfrak{k}_0$ as

$$X = \begin{pmatrix} U_1 & 0 \\ 0 & 0 \end{pmatrix} + i \begin{pmatrix} \frac{\alpha}{p} I_p & 0 \\ 0 & -\frac{\alpha}{n-p} I_{n-p} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & V_1 \end{pmatrix}, \qquad U_1 \in \mathfrak{su}(p),\ V_1 \in \mathfrak{su}(n-p),\ \alpha \in \mathbb{R},$$

and

$$\mathfrak{k}_0 \cong \mathfrak{su}(p) \oplus i\mathbb{R} \oplus \mathfrak{su}(n-p),$$

where all the summands are ideals.

The derivative $d(\sigma_0)_I$ of the automorphism $\sigma_0 \colon \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}} \to \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}}$ given by

$$\sigma_0(X) = I_{p,n-p}(X^*)^{-1} I_{p,n-p}$$

is the map $c_0 \colon \mathfrak{sl}(n, \mathbb{C})_{\mathbb{R}} \to \mathfrak{sl}(n, \mathbb{C})_{\mathbb{R}}$ also given by $c_0(X) = -I_{p,n-p} X^* I_{p,n-p}$. The real Lie group G_0 is the set of fixed points of $\mathbf{SL}(n, \mathbb{C})_{\mathbb{R}}$ under the automorphism σ_0, given by

$$\begin{aligned} G_0 &= \{X \in \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}} \mid X = I_{p,n-p}(X^*)^{-1} I_{p,n-p}\} \\ &= \{X \in \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}} \mid X^* I_{p,n-p} X = I_{p,n-p}\} = \mathbf{SU}(p, n-p). \end{aligned}$$

If we write

$$X = \begin{pmatrix} U & B \\ A & V \end{pmatrix}$$

for any $X \in \mathbf{SU}(p, n-p)$, it is easy to check that if $X \in \mathbf{SU}(p, n-p) \cap \mathbf{SU}(n)$, then

$$\begin{aligned} U^*U - A^*A &= I_p \\ V^*V - B^*B &= I_{n-p} \\ U^*U + A^*A &= I_p \\ V^*V + B^*B &= I_{n-p} \\ U^*B - A^*V &= 0 \\ U^*B + A^*V &= 0, \end{aligned}$$

which implies that $A = 0$ and $B = 0$. Therefore,

$$\begin{aligned} K_0 &= G_0 \cap G_u = \mathbf{SU}(p, n-p) \cap \mathbf{SU}(n) \\ &= \left\{ \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \;\middle|\; U \in \mathbf{U}(p),\; V \in \mathbf{U}(n-p),\; \det(U)\det(V) = 1 \right\}. \end{aligned}$$

This group is usually denoted $S(\mathbf{U}(p) \times \mathbf{U}(n-p))$. For any $X \in S(\mathbf{U}(p) \times \mathbf{U}(n-p))$ we can write

$$X = \begin{pmatrix} U_1 & 0 \\ 0 & I_{n-p} \end{pmatrix} \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & I_{p-1} & 0 & 0 \\ 0 & 0 & \lambda^{-1} & 0 \\ 0 & 0 & 0 & I_{n-p-1} \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & V_1 \end{pmatrix},$$

with $U_1 \in \mathbf{SU}(p), V_1 \in \mathbf{SU}(n-p)$ and $\lambda \in \mathbb{C}$ with $|\lambda| = 1$. Therefore

$$K_0 = S(\mathbf{U}(p) \times \mathbf{U}(n-p)) \cong \mathbf{SU}(p) \times \mathbf{U}(1) \times \mathbf{SU}(n-p).$$

Again the following paragraph is meant for readers well acquainted with Lie groups and Lie algebras and can be safely omitted. The Lie algebra of the real compact Lie group $K_0 = \mathbf{SU}(p) \times \mathbf{U}(1) \times \mathbf{SU}(n-p)$ is $\mathfrak{k}_0 = \mathfrak{su}(p) \oplus i\mathbb{R} \oplus \mathfrak{su}(n-p)$. The real Lie groups $G_0 = \mathbf{SU}(p, n-p)$ and $K_0 = \mathbf{SU}(p) \times \mathbf{U}(1) \times \mathbf{SU}(n-p)$ are semi-simple and connected but not simply-connected. The universal cover of K_0 is $\widetilde{K}_0 = \mathbf{SU}(p) \times \mathbb{R} \times \mathbf{SU}(n-p)$, and the universal cover of $G_0 = \mathbf{SU}(p, n-p)$ is $G_0 = \mathbf{Spin}(p, n-p)$. The real semi-simple (connected) Lie group $G_0 = \mathbf{SU}(p, n-q)$ is called a real form of the complex semi-simple (simply-connected) Lie group $G = \mathbf{SL}(n, \mathbb{C})$. We will show later that the pair $(G_0, K_0) = (\mathbf{SU}(p, n-p), \mathbf{SU}(p) \times \mathbf{U}(1) \times \mathbf{SU}(n-p))$ is a Gelfand pair.

Example 17.4. Again, consider the real Lie algebra $\mathfrak{g}_u = \mathfrak{su}(n)$ of $n \times n$ complex skew-hermitian matrices with zero trace and its complexification $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$. This time, assume $n = 2m$, and let $c_0 \colon \mathfrak{sl}(n, \mathbb{C}) \to \mathfrak{sl}(n, \mathbb{C})$ be the conjugation given by

$$c_0(X) = -J_m \overline{X} J_m,$$

with

$$J_m = \begin{pmatrix} 0_m & I_m \\ -I_m & 0_m \end{pmatrix}.$$

Since $J_m^2 = -I_{2m}$, $\overline{J_m} = J_m$, and $J_m^\top = -J_m$, the conjugation c_0 commutes with c_u. The automorphism $\sigma_0 \colon \mathbf{SL}(2m, \mathbb{C})|_{\mathbb{R}} \to \mathbf{SL}(2m, \mathbb{C})_{\mathbb{R}}$ such that $(d\sigma_0)_I = c_0$ is also $\sigma_0(X) = -J_m \overline{X} J_m$. In this example, since we also consider matrices whose entries are quaternions, we denote the group $\mathbf{SU}(n)$ as $\mathbf{SU}(n, \mathbb{C})$ and the Lie algebra $\mathfrak{su}(n)$ as $\mathfrak{su}(n, \mathbb{C})$ as to avoid confusion. We will determine the Lie algebras $\mathfrak{k}_0 = \{X \in \mathfrak{su}(2m, \mathbb{C}) \mid c_0(X) = X\}$, the group $G_0 = \{X \in \mathbf{SL}(2m, \mathbb{C})|_{\mathbb{R}} \mid \sigma_0(X) = X\}$, and the group $K_0 = G_0 \cap \mathbf{SU}(2m, \mathbb{C})$. The group G_0 is a Lie group known as $\mathbf{SU}^*(2m)$. We give another description of the group $\mathbf{SU}^*(2m)$ as a group of matrices with quaternion entries ($\mathbf{SL}(m, \mathbb{H})$). We also give two descriptions of

the group K_0; one as a group of matrices with quaternion entries ($\mathbf{SU}(m, \mathbb{H})$), and the other as a symplectic group ($\mathbf{Sp}(m)$).

If we write

$$X = \begin{pmatrix} U & V \\ A & B \end{pmatrix} \in \mathbf{SL}(2m, \mathbb{C}),$$

where $U, V, A, B \in \mathrm{M}_m(\mathbb{C})$, then we have

$$\begin{aligned} -J_m \overline{X} J_m &= \begin{pmatrix} 0 & -I_m \\ I_m & 0 \end{pmatrix} \begin{pmatrix} \overline{U} & \overline{V} \\ \overline{A} & \overline{B} \end{pmatrix} \begin{pmatrix} 0 & I_m \\ -I_m & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -I_m \\ I_m & 0 \end{pmatrix} \begin{pmatrix} -\overline{V} & \overline{U} \\ -\overline{B} & \overline{A} \end{pmatrix} = \begin{pmatrix} \overline{B} & -\overline{A} \\ -\overline{V} & \overline{U} \end{pmatrix}, \end{aligned}$$

so $X = -J_m \overline{X} J_m$ iff

$$U = \overline{B}, \quad V = -\overline{A}, \quad A = -\overline{V}, \quad B = \overline{U},$$

which simplifies to

$$B = \overline{U}, \qquad A = -\overline{V}.$$

Therefore, X is of the form

$$X = \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix},$$

and the real Lie group $G_0 = \{X \in \mathbf{SL}(2m, \mathbb{C})|_{\mathbb{R}} \mid X = -J_m \overline{X} J_m\}$ is given by

$$G_0 = \left\{ \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \in \mathbf{SL}(2m, \mathbb{C}) \;\middle|\; U, V \in \mathrm{M}_m(\mathbb{C}) \right\}.$$

Definition 17.13. The real Lie group $\mathbf{SU}^*(2m)$ is defined by

$$\mathbf{SU}^*(2m) = \left\{ \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \in \mathbf{SL}(2m, \mathbb{C}) \;\middle|\; U, V \in \mathrm{M}_m(\mathbb{C}) \right\}.$$

The notation $\mathbf{SU}^*(2m)$ for this real Lie group is introduced in Helgason [47], Chapter X, Section 2. We will show later that the (real) Lie group $\mathbf{SU}^*(2m)$ is isomorphic to the quaternionic (real) Lie group $\mathbf{SL}(m, \mathbb{H})$.

Since $X \in \mathbf{SU}^*(2m)$ is invertible, for any vector $z = \binom{x}{y} \in \mathbb{C}^{2m}$, if $Xz = 0$ then $z = 0$, so in particular, for $y = 0$, since

$$Xz = \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} Ux + Vy \\ -\overline{V}x + \overline{U}y \end{pmatrix},$$

we have $Xz = 0$ iff $Ux + Vy = 0$ and $\overline{V}x + \overline{U}y = 0$. So with $y = 0$, if $Ux = 0$ then $x = 0$, and with $x = 0$, if $Vy = 0$ then $y = 0$ which implies that U and V are invertible. Therefore, the group G_0, the set of fixed points of σ_0, is also given by

$$G_0 = \left\{ X = \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \;\middle|\; U, V \in \mathbf{GL}(m, \mathbb{C})|_{\mathbb{R}},\ \det(X) = 1 \right\}.$$

Since the conjugation c_0 on $\mathfrak{sl}(2m,\mathbb{C})$ has the same expression as the conjugation σ_0 on $\mathbf{SL}(2m,\mathbb{C})|_{\mathbb{R}}$, the same computation as above shows that

$$\mathfrak{su}^*(2m) = \left\{ \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \;\middle|\; U, V \in \mathrm{M}_m(\mathbb{C}),\ \mathrm{tr}(U) + \mathrm{tr}(\overline{U}) = 0 \right\}.$$

We also have

$$\begin{aligned}
\mathfrak{k}_0 &= \{X \in \mathfrak{su}(2m,\mathbb{C}) \mid X = -J_m \overline{X} J_m\} \\
&= \left\{ \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \;\middle|\; U \in \mathfrak{u}(m,\mathbb{C}),\ V \in \mathrm{M}_m(\mathbb{C}),\ V^\top = V \right\} \\
&= \left\{ \begin{pmatrix} U & V \\ -\overline{V} & -U^\top \end{pmatrix} \;\middle|\; U \in \mathfrak{u}(m,\mathbb{C}),\ V \in \mathrm{M}_m(\mathbb{C}),\ V^\top = V \right\},
\end{aligned}$$

and

$$\begin{aligned}
i\mathfrak{p}_0 &= \{X \in \mathfrak{su}(2m,\mathbb{C}) \mid X = J_m \overline{X} J_m\} \\
&= \left\{ \begin{pmatrix} U & V \\ \overline{V} & -\overline{U} \end{pmatrix} \;\middle|\; U \in \mathfrak{u}(m,\mathbb{C}),\ V \in \mathrm{M}_m(\mathbb{C}),\ V^\top = -V,\ \mathrm{tr}(U) = 0 \right\}.
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
\mathfrak{p}_0 &= \left\{ \begin{pmatrix} iU & iV \\ i\overline{V} & -i\overline{U} \end{pmatrix} \;\middle|\; U \in \mathfrak{u}(m,\mathbb{C}),\ V \in \mathrm{M}_m(\mathbb{C}),\ V^\top = -V,\ \mathrm{tr}(U) = 0 \right\} \\
&= \left\{ \begin{pmatrix} A & B \\ -\overline{B} & \overline{A} \end{pmatrix} \;\middle|\; A \in i\mathfrak{u}(m,\mathbb{C}),\ B \in \mathrm{M}_m(\mathbb{C}),\ B^\top = -B,\ \mathrm{tr}(A) = 0 \right\} \\
&= \left\{ \begin{pmatrix} A & B \\ -\overline{B} & \overline{A} \end{pmatrix} \;\middle|\; A,\ B \in \mathrm{M}_m(\mathbb{C}),\ A^* = A,\ B^\top = -B,\ \mathrm{tr}(A) = 0 \right\} \\
&= \left\{ \begin{pmatrix} A & B \\ -\overline{B} & A^\top \end{pmatrix} \;\middle|\; A,\ B \in \mathrm{M}_m(\mathbb{C}),\ A^* = A,\ B^\top = -B,\ \mathrm{tr}(A) = 0 \right\}.
\end{aligned}$$

Observe that if $X \in \mathfrak{k}_0$, then automatically $\mathrm{tr}(X) = 0$. We immediately check that

$$\mathfrak{su}^*(2m) = \mathfrak{k}_0 \oplus \mathfrak{p}_0.$$

The (real) Lie group $\mathbf{SU}^*(2m)$ can also be viewed as the (real) Lie group $\mathbf{SL}(m,\mathbb{H})$, where $\mathbb{H}$ is the skew-field of quaternions. To see this, it is convenient to view the real algebra $\mathbb{H}$ as

$$\mathbb{H} = \mathbb{C} \oplus \mathbb{C}j.$$

using the fact that every quaternion can be written uniquely as

$$q = a + xi + yj + zk = a + xi + (y + zi)j, \quad a, x, y, z \in \mathbb{R},$$

since $ij = k$. Since the conjugate $\overline{q}$ of q is $\overline{q} = a - xi - yj - zk = a - xi - (y + zi)j$, if we write $q = \alpha + \beta j$ with $\alpha, \beta \in \mathbb{C}$, then

$$\overline{q} = \overline{\alpha} - \beta j.$$

Also, since $ij = -ji$, for any $\alpha = a + bi \in \mathbb{C}$, we have

$$j\alpha = j(a + bi) = ja + bji = aj - bij = (a - bi)j = \overline{\alpha}j.$$

In summary, we have

$$\overline{\alpha + j\beta} = \overline{\alpha} - \beta j, \qquad j\alpha = \overline{\alpha}j. \tag{conj}$$

Every $m \times m$ matrix $X = (\alpha_{k\ell} + \beta_{k\ell}j) \in \mathrm{M}_m(\mathbb{H})$ can be written uniquely as

$$X = U + Vj,$$

with $U = (\alpha_{k\ell}) \in \mathrm{M}_m(\mathbb{C})$ and $V = (\beta_{k\ell}) \in \mathrm{M}_m(\mathbb{C})$. Then in view of the equations (conj), if $X = (\alpha_{k\ell} + \beta_{k\ell}j) \in \mathrm{M}_m(\mathbb{H})$ and if we define

$$\overline{X} = \left(\overline{\alpha_{\ell k} + \beta_{\ell k}j}\right), \quad \text{and} \quad X^* = \overline{X}^\top,$$

then we have

$$X^* = (U + Vj)^* = U^* - V^\top j.$$

We also have

$$jV = \overline{V}j.$$

In summary, we have

$$(U + Vj)^* = U^* - V^\top j, \qquad jV = \overline{V}j. \tag{conj'}$$

The (real) Lie group $\mathbf{GL}(m, \mathbb{H})$ is the group of matrices $X \in \mathrm{M}_m(\mathbb{H})$ such that there is some $Y \in \mathrm{M}_m(\mathbb{H})$ with

$$XY = YX = I_m.$$

Observe that since $j^2 = -1$, we have

$$\begin{aligned}
(U_1 + V_1j)(U_2 + V_2j) &= U_1V_1 + U_1V_2j + V_1jU_2 + V_1jV_2j \\
&= U_1V_1 + U_1V_2j + V_1\overline{U_2}j + V_1\overline{V_2}j^2 \\
&= U_1V_1 - V_1\overline{V_2} + (U_1V_2 + V_1\overline{U_2})j.
\end{aligned}$$

Since

$$\begin{pmatrix} U_1 & V_1 \\ -\overline{V_1} & \overline{U_1} \end{pmatrix} \begin{pmatrix} U_2 & V_2 \\ -\overline{V_2} & \overline{U_2} \end{pmatrix} = \begin{pmatrix} U_1U_2 - V_1\overline{V_2} & U_1V_2 + V_1\overline{U_2} \\ -\overline{U_1}\,\overline{V_2} - \overline{V_1}U_2 & \overline{U_1}\,\overline{U_2} - \overline{V_1}V_2 \end{pmatrix},$$

the map $\varphi\colon \mathrm{M}_m(\mathbb{H}) \to \mathrm{M}_{2m}(\mathbb{C})$ given by

$$\varphi(U + Vj) = \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix}$$

is an injective $\mathbb{R}$-algebra homomorphism. Observe that this is the matrix in the definition of G_0.

What we did previously shows that this homomorphism restricts to an injective homomorphism $\varphi\colon \mathbf{GL}(m,\mathbb{H}) \to \mathbf{GL}(2m,\mathbb{C})$, which allows us to view the group $\mathbf{GL}(m,\mathbb{H})$ as

$$\varphi(\mathbf{GL}(m,\mathbb{H})) = \left\{ X = \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \;\middle|\; X \in \mathbf{GL}(2m,\mathbb{C}),\ U, V \in \mathbf{GL}(m,\mathbb{C}) \right\}.$$

Definition 17.14. We define the (real) Lie the group $\mathbf{SL}(m,\mathbb{H})$ as

$$\mathbf{SL}(m,\mathbb{H}) = \{X \in \varphi(\mathbf{GL}(m,\mathbb{H})) \mid \det(X) = 1\} = \varphi(\mathbf{GL}(m,\mathbb{H})) \cap \mathbf{SL}(2m,\mathbb{C}).$$

Therefore, we conclude that

$$G_0 = \mathbf{SU}^*(2m) = \mathbf{SL}(m,\mathbb{H}).$$

This is a real, semi-simple, simply-connected Lie group.

Technically, $\mathbf{SL}(m,\mathbb{H})$ should be defined as the subgroup of $\mathbf{GL}(m,\mathbb{H})$ given by

$$\{X \in \mathbf{GL}(m,\mathbb{H}) \mid \det(\varphi(X)) = 1\},$$

but for our purpose it is more convenient to view $\mathbf{GL}(m,\mathbb{H})$ and its various subgroups as subgroups of $\mathbf{GL}(2m,\mathbb{C})$.

With this identification in mind the Lie algebras $\mathfrak{gl}(m,\mathbb{H})$ of $\mathbf{GL}(m,\mathbb{H})$ and $\mathfrak{sl}(m,\mathbb{H})$ of $\mathbf{SL}(m,\mathbb{H})$ are given by

$$\mathfrak{gl}(m,\mathbb{H}) = \left\{ \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \;\middle|\; U, V \in \mathbf{M}_m(\mathbb{C}) \right\},$$

and

$$\mathfrak{sl}(m,\mathbb{H}) = \left\{ \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \;\middle|\; U, V \in \mathbf{M}_m(\mathbb{C}),\ \mathrm{tr}(U) + \mathrm{tr}(\overline{U}) = 0 \right\}.$$

Observe that

$$\mathfrak{sl}(m,\mathbb{H}) = \mathfrak{su}^*(2n),$$

as it should be.

We can also identify

$$K_0 = G_0 \cap \mathbf{SU}(2m,\mathbb{C}) = \mathbf{SU}^*(2m) \cap \mathbf{SU}(2m,\mathbb{C}) = \mathbf{SL}(m,\mathbb{H}) \cap \mathbf{SU}(2m,\mathbb{C}).$$

There are two descriptions for the Lie group K_0; one as a quaternionic group; another in terms of the complex symplectic group $\mathbf{Sp}(m,\mathbb{C})$.

For the quaternionic description, define $\mathbf{U}(m, \mathbb{H})$ as the (real) Lie group

$$\mathbf{U}(m, \mathbb{H}) = \{X = U + Vj \in \mathbf{GL}(m, \mathbb{H}) \mid X^*X = XX^* = I_m\}.$$

It is easy to see that only the first equation is needed. Using (conj'), this is equivalent to

$$\begin{aligned}(U + Vj)^*(U + Vj) &= (U^* - V^\top j)(U + Vj) \\ &= U^*U + U^*Vj - V^\top jU - V^\top jVj \\ &= U^*U + U^*Vj - V^\top \overline{U} j - V^\top \overline{V} j^2 \\ &= U^*U + V^\top \overline{V} + (U^*V - V^\top \overline{U})j = I_m,\end{aligned}$$

so we get

$$U^*U + V^\top \overline{V} = I_m, \qquad U^*V - V^\top \overline{U} = 0.$$

On the other hand, for

$$X = \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix},$$

we have $X^*X = I_{2m}$ iff

$$\begin{pmatrix} U^* & -V^\top \\ V^* & U^\top \end{pmatrix} \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} = \begin{pmatrix} U^*U + V^\top \overline{V} & U^*V - V^\top \overline{U} \\ V^*U - U^\top \overline{V} & V^*V + U^\top \overline{U} \end{pmatrix} = \begin{pmatrix} I_m & 0 \\ 0 & I_m \end{pmatrix},$$

and these are equivalent to the same conditions as above,

$$U^*U + V^\top \overline{V} = I_m, \qquad U^*V - V^\top \overline{U} = 0.$$

We conclude that

$$\varphi(\mathbf{U}(m, \mathbb{H})) = \varphi(\mathbf{GL}(m, \mathbb{H})) \cap \mathbf{U}(2m, \mathbb{C}).$$

Definition 17.15. The (real) Lie group $\mathbf{SU}(m, \mathbb{H})$ is defined as

$$\mathbf{SU}(m, \mathbb{H}) = \varphi(\mathbf{GL}(m, \mathbb{H})) \cap \mathbf{SU}(2m, \mathbb{C}).$$

Since

$$\mathbf{SL}(m, \mathbb{H}) = \varphi(\mathbf{GL}(m, \mathbb{H})) \cap \mathbf{SL}(2m, \mathbb{C})$$

we have

$$\mathbf{SU}(m, \mathbb{H}) = \mathbf{SL}(m, \mathbb{H}) \cap \mathbf{SU}(2m, \mathbb{C}).$$

Again, for our purpose, it is more convenient to view $\mathbf{SL}(m, \mathbb{H})$ and $\mathbf{SU}(m, \mathbb{H})$ as subgroups of $\mathbf{GL}(2m, \mathbb{C})$.

In summary, we see that

$$\begin{aligned}K_0 &= \mathbf{SU}^*(2m) \cap \mathbf{SU}(2m, \mathbb{C}) = \mathbf{SL}(m, \mathbb{H}) \cap \mathbf{SU}(2m, \mathbb{C}) \\ &= \mathbf{SU}(m, \mathbb{H}).\end{aligned}$$

This is a real, semi-simple, simply-connected Lie group.

The Lie algebra $\mathfrak{u}(m, \mathbb{H})$ of $\mathbf{U}(m, \mathbb{H})$ consists of the space of matrices

$$\mathfrak{u}(m, \mathbb{H}) = \{X \in \mathrm{M}_m(\mathbb{H}) \mid X^* = -X\}.$$

If we write $X = U + Vj$ (with $U, V \in \mathrm{M}_n(\mathbb{C})$), then $X^* = -X$ is equivalent to

$$(U + Vj)^* = U^* - V^\top j = -U - Vj,$$

that is,

$$U^* = -U, \qquad V^\top = V.$$

Thus we can also write

$$\begin{aligned} \mathfrak{u}(m, \mathbb{H}) &= \left\{ \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \,\middle|\, U, V \in \mathbf{M}_m(\mathbb{C}),\ U^* = -U,\ V^\top = V \right\} \\ &= \left\{ \begin{pmatrix} U & V \\ -\overline{V} & -U^\top \end{pmatrix} \,\middle|\, U \in \mathfrak{u}(m, \mathbb{C}),\ V \in \mathbf{M}_m(\mathbb{C}),\ V^\top = V \right\}, \end{aligned}$$

and so

$$\mathfrak{u}(m, \mathbb{H}) = \mathfrak{k}_0.$$

Since

$$\mathfrak{su}(m, \mathbb{H}) = \{X \in \mathfrak{u}(m, \mathbb{H}) \mid \mathrm{tr}(X) = 0\},$$

we find that

$$\mathfrak{su}(m, \mathbb{H}) = \mathfrak{u}(m, \mathbb{H}) = \mathfrak{k}_0,$$

as it should be.

The real Lie group $\mathbf{SU}(m, \mathbb{H})$ has another description in terms of the complex symplectic group $\mathbf{Sp}(m, \mathbb{C})$. Since

$$\mathbf{SU}(m, \mathbb{H}) = \{X \in \mathrm{M}_{2m}(\mathbb{C}) \mid X^* X = I_{2m},\ X = -J_m \overline{X} J_m,\ \det(X) = 1\},$$

first from $X^* X = I_{2m}$ we get

$$I_{2m} = X^* X = -X^* J_m \overline{X} J_m;$$

since $J_m^2 = -I_{2m}$ we get

$$X^* J_m \overline{X} = J_m,$$

and since $\overline{J_m} = J_m$, by conjugating both sides we get

$$X^\top J_m X = J_m.$$

Definition 17.16. The *complex symplectic group* $\mathbf{Sp}(m, \mathbb{C})$ is defined as

$$\mathbf{Sp}(m, \mathbb{C}) = \{X \in \mathrm{M}_{2m}(\mathbb{C}) \mid X^\top J_m X = J_m\}.$$

It can be shown that the Lie algebra $\mathfrak{sp}(m, \mathbb{C})$ of the Lie group $\mathbf{Sp}(m, \mathbb{C})$ is given by

$$\mathfrak{sp}(m, \mathbb{C}) = \left\{ \begin{pmatrix} U & V_1 \\ V_2 & -U^\top \end{pmatrix} \;\middle|\; U, V_1, V_2 \in \mathbf{M}_m(\mathbb{C}),\ V_1^\top = V_1,\ V_2^\top = V_2 \right\};$$

see Helgason [47] (Chapter X, Section 2, page 446).

The preceding argument showed that

$$\mathbf{SU}(m, \mathbb{H}) = \mathbf{SL}(m, \mathbb{H}) \cap \mathbf{SU}(2m, \mathbb{C}) \subseteq \mathbf{Sp}(m, \mathbb{C}) \cap \mathbf{U}(2m, \mathbb{C}).$$

It can also be shown that the group $\mathbf{Sp}(m, \mathbb{C})$ is connected; see Helgason [47] (Chapter X, Section 2, Lemma 2.4) or Knapp [57] (Chapter I, Proposition 1.145). Since J_m is invertible ($J_m^{-1} = -J_m$), the equation $X^\top J_m X = J_m$ shows that $\det(X) = \pm 1$. Since $\det(I) = 1$ and $\mathbf{Sp}(m, \mathbb{C})$ is connected, we deduce that for every $X \in \mathbf{Sp}(m, \mathbb{C})$, we have $\det(X) = 1$. (There are also purely algebraic proofs of this property using the fact that the symplectic transvections generate $\mathbf{Sp}(m, \mathbb{C})$ and that they have determinant 1.) Thus if $X \in \mathbf{Sp}(m, \mathbb{C}) \cap \mathbf{U}(2m, \mathbb{C})$, then in fact $X \in \mathbf{Sp}(m, \mathbb{C}) \cap \mathbf{SU}(2m, \mathbb{C})$, and from $X^\top J_m X = J_m$ and $X^* X = I_{2m}$, by reversing the above argument, we deduce that $X = -J_m \overline{X} J_m$, and so

$$\mathbf{Sp}(m, \mathbb{C}) \cap \mathbf{U}(2m, \mathbb{C}) \subseteq \mathbf{SL}(m, \mathbb{H}) \cap \mathbf{SU}(2m, \mathbb{C}) = \mathbf{SU}(m, \mathbb{H}).$$

We just showed that

$$\mathbf{SU}(m, \mathbb{H}) = \mathbf{SL}(m, \mathbb{H}) \cap \mathbf{SU}(2m, \mathbb{C}) = \mathbf{Sp}(m, \mathbb{C}) \cap \mathbf{U}(2m, \mathbb{C}).$$

Definition 17.17. We define the real Lie group $\mathbf{Sp}(m)$ as

$$\mathbf{Sp}(m) = \mathbf{Sp}(m, \mathbb{C}) \cap \mathbf{U}(2m, \mathbb{C});$$

see Helgason [47] (Chapter X, Section 2).

Since we showed that $\det(X) = 1$ for all $X \in \mathbf{Sp}(m, \mathbb{C})$, we also have

$$\mathbf{Sp}(m) = \mathbf{Sp}(m, \mathbb{C}) \cap \mathbf{SU}(2m, \mathbb{C}).$$

In view of Definition 17.17, we just proved that

$$\mathbf{Sp}(m) = \mathbf{SU}(m, \mathbb{H}),$$

and the Lie algebra $\mathfrak{sp}(m)$ of $\mathbf{Sp}(m)$ is equal to $\mathfrak{sp}(m, \mathbb{C}) \cap \mathfrak{u}(2m, \mathbb{C})$, so it is given by

$$\begin{aligned}
\mathfrak{sp}(m) &= \left\{ \begin{pmatrix} U & V \\ -V^* & -U^\top \end{pmatrix} \;\middle|\; U \in \mathfrak{u}(m, \mathbb{C}),\ V \in \mathrm{M}_m(\mathbb{C}),\ V^\top = V \right\} \\
&= \left\{ \begin{pmatrix} U & V \\ -\overline{V} & -U^\top \end{pmatrix} \;\middle|\; U \in \mathfrak{u}(m, \mathbb{C}),\ V \in \mathrm{M}_m(\mathbb{C}),\ V^\top = V \right\}.
\end{aligned}$$

Again,

$$\mathfrak{sp}(m) = \mathfrak{k}_0,$$

as it should be.

The real semi-simple simply-connected Lie group $G_0 = \mathbf{SL}(m, \mathbb{H}) = \mathbf{SU}^*(2m)$ is another real form of the complex semi-simple (simply-connected) Lie group $G = \mathbf{SL}(2m, \mathbb{C})$. We will show later that the pair $(G_0, K_0) = (\mathbf{SL}(m, \mathbb{H}), \mathbf{SU}(m, \mathbb{H})) = (\mathbf{SU}^*(2m), \mathbf{Sp}(m))$ is a Gelfand pair.

It can be shown that up to isomorphism, $\mathbf{SL}(n, \mathbb{R})$, $\mathbf{SU}(p, n-p)$, and $\mathbf{SL}(m, \mathbb{H}) = \mathbf{SU}^*(2m)$ (when $n = 2m$), are the only real forms of $\mathbf{SL}(n, \mathbb{C})$; see Helgason [47] and Dieudonné [21] (Chapter XXI, Section 21.18.11).

17.5 Real Forms of Complex Semi-Simple Simply-Connected Lie Groups

A general method to find real forms of a complex, semi-simple, simply-connected Lie group G with complex semi-simple Lie algebra $\mathfrak{g}$ goes as follows. Suppose we have real, compact, semi-simple, simply-connected Lie group G_u with Lie algebra $\mathfrak{g}_u$ such that $\mathfrak{g} = \mathfrak{g}_u \oplus i\mathfrak{g}_u$ is the complexification of the real Lie algebra $\mathfrak{g}_u$. In our previous examples, $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$, $G = \mathbf{SL}(n, \mathbb{C})$, $\mathfrak{g}_u = \mathfrak{su}(n, \mathbb{C})$, and $G_u = \mathbf{SU}(n, \mathbb{C})$. By a famous theorem of Hermann Weyl, such a real semi-simple Lie algebra $\mathfrak{g}_u$ and such a compact semi-simple Lie group G_u always exist; see Dieudonné [21] (Chapter XXI, no. 21.20.7).

Also assume that we have a conjugation $c_0 \colon \mathfrak{g} \to \mathfrak{g}$ that commutes with the conjugation c_u associated with $\mathfrak{g}_u$. Using the correspondence between (real) simply-connected Lie groups and (real) Lie algebras, there is a unique involutive automorphism $\sigma_0 \colon G|_{\mathbb{R}} \to G|_{\mathbb{R}}$ such that $d(\sigma_0)_e = c_0$, where $c_0 \colon \mathfrak{g}|_{\mathbb{R}} \to \mathfrak{g}|_{\mathbb{R}}$. If P is the closed submanifold of $G|_{\mathbb{R}}$ which is the image of the real vector space $i\mathfrak{g}_u$ by the exponential map $\exp_G \colon \mathfrak{g} \to G$ ($ix \mapsto \exp(ix), x \in \mathfrak{g}_u$), then the following result can be shown.

Proposition 17.13. *The exponential map* $\exp_G$ *is a diffeomorphism of* $i\mathfrak{g}_u$ *onto* P*, and the map* $(x, y) \mapsto xy$ *is a diffeomorphism of* $G_u \times P$ *onto* $G|_{\mathbb{R}}$ *(as real manifolds).*

The automorphism σ_0 of $G|_{\mathbb{R}}$ leaves G_u and $P|_{\mathbb{R}}$ invariant since c_0 leaves $\mathfrak{g}_u$ and $i\mathfrak{g}_u$ invariant.

If $G = \mathbf{SL}(n, \mathbb{C})$, $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$, $G_u = \mathbf{SU}(n, \mathbb{C})$, and $\mathfrak{g}_u = \mathfrak{su}(n, \mathbb{C})$, then

$$i\mathfrak{g}_u = i\mathfrak{su}(n, \mathbb{C}) = \{X \in \mathrm{M}_n(\mathbb{C}) \mid X^* = X\},$$

and $P = \exp_G(i\mathfrak{g}_u)$ is the manifold of hermitian positive definite matrices with determinant $+1$. The factorization $\mathbf{SL}(n, \mathbb{C}) = \mathbf{SU}(n, \mathbb{C})P$ is the polar form for matrices in $\mathbf{SL}(n, \mathbb{C})$.

Let G_0 be the real Lie subgroup of $G|_{\mathbb{R}}$ given by

$$G_0 = \{s \in G|_{\mathbb{R}} \mid \sigma_0(s) = s\}.$$

The group G_0 is a real semi-simple Lie group because its Lie algebra is $\mathfrak{g}_0$. Since σ_0 leaves G_u invariant, the group

$$K_0 = G_0 \cap G_u = \{s \in G_u \mid \sigma_0(s) = s\}$$

is a real compact subgroup of G_u consisting of the fixed points of G_u under σ_0. The group G_0 also contains the image P_0 of $\mathfrak{p}_0 \subseteq i\mathfrak{g}_u$ under the exponential map $v \mapsto \exp_{G_0}(v)$, and since $\exp_{G_0}(c_0(v)) = \sigma_0(\exp_{G_0}(v))$ (see Proposition 19.7 in Gallier and Quaintance [38]), the manifold P_0 is the set of fixed points of $P|_{\mathbb{R}}$ under σ_0.

Proposition 17.14. *The map $v \mapsto \exp_{G_0}(v)$ is a diffeomorphism of $\mathfrak{p}_0$ onto the closed manifold P_0, and the map $(y, z) \mapsto yz$ from $K_0 \times P_0$ to G_0 is a diffeomorphism.*

If $G = \mathbf{SL}(n, \mathbb{C})$, $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$, $G_u = \mathbf{SU}(n, \mathbb{C})$, and $\mathfrak{g}_u = \mathfrak{su}(n, \mathbb{C})$, and $c_0(X) = \overline{X}$, as Example 17.2, we have

$$G_0 = \mathbf{SL}(n, \mathbb{R}), \quad K_0 = \mathbf{SO}(n), \quad \mathfrak{k}_0 = \mathfrak{so}(n), \quad \mathfrak{p}_0 = \mathfrak{s}(n) = \{X \in \mathrm{M}_n(\mathbb{R}) \mid X^\top = X\},$$

and $P_0 = \exp_{G_0}(\mathfrak{p}_0)$ is the manifold of symmetric positive definite matrices with determinant $+1$. The factorization $\mathbf{SL}(n, \mathbb{R}) = \mathbf{SO}(n)P_0$ is the polar form for matrices in $\mathbf{SL}(n, \mathbb{R})$.

Thus G_0 is a real noncompact semi-simple connected Lie group, diffeomorphic to the product of a compact semi-simple Lie group K_0 and some $\mathbb{R}^N$. It can be shown that the group K_0 is a maximal compact subgroup of G_0, is connected, but in general not semi-simple nor simply-connected.

Definition 17.18. The real, semi-simple, simply-connected Lie group G_u is called the *compact form* of the complex semi-simple simply-connected Lie group G, and the real, semi-simple, connected Lie group G_0 is a *real form* of G.

The pair (G_0, K_0) is a Gelfand pair, but not with respect to the involution σ_0, because σ_0 is the identity on P_0. However, there is also a unique involutive automorphism $\sigma_u \colon G|_{\mathbb{R}} \to G|_{\mathbb{R}}$ such that $d(\sigma_u)_e = c_u$, and this involution makes (G_0, K_0) a Gelfand pair.

Since $\mathfrak{k}_0 \subseteq \mathfrak{g}_u$ and since $\mathfrak{g}_u$ is the set of fixed points of c_u, the Lie group K_0 is a set of fixed points of σ_u. Similarly, since $\mathfrak{p}_0 \subseteq i\mathfrak{g}_u$, and since $c_u(ix) = -ix$ for all $x \in \mathfrak{g}_u$, we see that $\sigma_u(\exp_{G_0}(ix)) = \exp_{G_0}(c_u(ix)) = \exp_{G_0}(-ix) = (\exp_{G_0}(ix))^{-1}$ for all $ix \in \mathfrak{p}_0$, so $\sigma_u(s) = s^{-1}$ for all $s \in P_0$ (since $P_0 = \exp_{G_0}(\mathfrak{p}_0)$).

Proposition 17.15. *We have $G_0 = K_0P_0$, $\sigma_u(s) = s$ for all $s \in K_0$, and $\sigma_u(s) = s^{-1}$ for all $s \in P_0$. The group K_0 is the set of fixed points of σ_u,*

$$K_0 = \{s \in G_0 \mid \sigma_u(s) = s\}.$$

Proof. Since every $s \in G_0 = K_0 P_0$ can be written as $s = xy$ with $x \in K_0$ and $y \in P_0$, we have $\sigma_u(s) = s$ iff $\sigma_u(xy) = xy$, but then

$$xy = \sigma_u(xy) = \sigma_u(x)\sigma_u(y) = xy^{-1},$$

so we deduce that

$$y^2 = e.$$

However, $y \in P_0 = \exp_{G_0}(\mathfrak{p}_0)$, so for $y = \exp_{G_0}(w)$ we have $y^2 = \exp_{G_0}(2w) = e$, and since $\exp_{G_0}$ is a diffeomorphism on $\mathfrak{p}_0$, we must have $w = 0$, and thus $y = e$. $\square$

In summary, since K_0 is compact, we proved that (G_0, K_0) *is a Gelfand pair with involution* σ_u.

The real semi-simple connected Lie group G_0 is called a real form of the complex semi-simple simply-connnected Lie group G because its real Lie algebra $\mathfrak{g}_0$ is a real form of the complex Lie algebra $\mathfrak{g}$ of G.

There are other real semi-simple connected Lie groups having $\mathfrak{g}_0$ as Lie algebra, and they can all be found (up to isomorphism) as follows; see Dieudonné [21] (Chapter XXI, no. 21.18.8-21.18.12).

Proposition 17.16. *Let $\widetilde{G}_0$ be the universal cover of G_0, let $\pi\colon \widetilde{G}_0 \to G_0$ be the covering map, and let $\widetilde{K}_0 = \pi^{-1}(K_0)$. Then $\widetilde{K}_0$ is isomorphic to the universal cover of the compact Lie group K_0, the exponential map $\exp_{\widetilde{G}_0}$ is a diffeomorphism of $\mathfrak{p}_0$ onto a closed submanifold $\widetilde{P}_0$ of $\widetilde{G}_0$ such that $\widetilde{K}_0 \cap \widetilde{P}_0 = \{e\}$, and the map $(x, y) \mapsto xy$ is a diffeomorphism of $\widetilde{K}_0 \times \widetilde{P}_0$ onto $\widetilde{G}_0$. The center Z of $\widetilde{G}_0$ is a discrete subgroup contained in the center of $\widetilde{K}_0$.*

If K_0 is not semi-simple, then Z is not equal to the center of $\widetilde{K}_0$ and $\widetilde{K}_0$ is not compact.

Theorem 17.17. *Every real semi-simple connected Lie group G_1 having $\mathfrak{g}_0$ as real Lie algebra is of the form $G_1 = \widetilde{G}_0/D$, where D is a (discrete) subgroup of the center Z of $\widetilde{G}_0$. The center C_1 of G_1 is given by $C_1 = Z/D$. The Lie group $K_1 = \widetilde{K}_0/D$ is a connected subgroup of G_1 which contains C_1 (in general, C_1 is not equal to the center of G_1), and whose Lie algebra is $\mathfrak{k}_0$. The Lie group K_1 is compact iff C_1 is finite. The exponential map $\exp_{G_1}$ is a diffeomorphism of $\mathfrak{p}_0$ onto a closed submanifold P_1 of G_1 such that $K_1 \cap P_1 = \{e\}$, and the map $(x, y) \mapsto xy$ is a diffeomorphism of $K_1 \times P_1$ onto G_1.*

Definition 17.19. The factorization $G_1 = K_1 P_1$ is called a *Cartan decomposition* of G_1.

The Cartan decomposition is a generalization of the polar form for invertible matrices.

It can also be shown that K_1 is isomorphic to the product of a compact Lie group with some $\mathbb{R}^m$. Thus G_1 is diffeomorphic to the product of a compact Lie group with some $\mathbb{R}^M$ (in fact, this compact group is maximal in G_1).

The reasoning in the proof of Proposition 17.15 involving the conjugation σ_u can be used to show that the conjugation σ_u on $\widetilde{G}_0$ such that $(d\sigma_u)_e = c_u$ induces a conjugation on $G_1 = \widetilde{G}_0/D$ by passing to the quotient, and because Z is contained in $\widetilde{K}_0$, that

$$K_1 = \{s \in G_1 \mid \sigma_u(s) = s\},$$

and that $\sigma_u(s) = s^{-1}$ for all $s \in P_1$. Since $G_1 = K_1 P_1$, if the center C_1 of G_1 is finite, then K_1 is compact, and then (G_1, K_1) is a Gelfand pair with involution σ_u.

17.6 Examples of Gelfand Pairs

There are three important cases for which Gelfand's theorem (Theorem 17.2) applies.

(1) Let G be a compact, connected real Lie group, and let σ be an involutive automorphism of G ($\sigma \neq \mathrm{id}_G$); see Dieudonné [21] (Chapter XXI, no. 21.18.13). Let G^σ be the closed (thus compact) Lie subgroup of G consisting of the fixed points of σ,

$$G^\sigma = \{s \in G \mid \sigma(s) = s\}.$$

The derivative $\theta = d\sigma_e$ of σ is an involution of the Lie algebra $\mathfrak{g}$ of G. Then as in Section 17.3, we know by linear algebra that

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_{-1}$$

where $\mathfrak{g}_1$ and $\mathfrak{g}_{-1}$ are the real eigenspaces of θ given by

$$\mathfrak{g}_1 = \{x \in \mathfrak{g} \mid \theta(x) = x\}, \qquad \mathfrak{g}_{-1} = \{x \in \mathfrak{g} \mid \theta(x) = -x\}.$$

Let G_0^σ be the connected component of e in G^σ. For any closed subgroup K (thus compact) such that $G_0^\sigma \subseteq K \subseteq G^\sigma$, using some differential geometry, it can be shown that $\mathfrak{g}_1$ is the Lie algebra of K; see O'Neill [76] or Gallier and Quaintance [38] (Proposition 23.33). Let $P = \exp(\mathfrak{g}_{-1})$. Then $\sigma(s) = s$ for all $s \in K$ and $\sigma(s) = s^{-1}$ for all $s \in P$ (since $\theta(x) = -x$ for all $x \in \mathfrak{g}_{-1}$). It remains to prove that $G = KP$. For this, again we use some differential geometry.

Since K is compact, G/K has some G-invariant metric. In fact, G/K is a naturally reductive homogeneous space. Since G is compact, G/K is compact, and by Hopf–Rinov, it is geodesically complete. But since G/K is naturally reductive, the tangent space $T_e(G/K) \cong \mathfrak{g}_{-1}$, and every geodesic γ_x through e with initial velocity $x \in \mathfrak{g}_{-1}$ is given by

$$\gamma_x(t) = \pi(\exp_G(tx));$$

see Proposition 23.27 in Gallier and Quaintance [38]. Consequently, $\pi(P) = G/K$, or equivalently $G = PK$. But since $P = \exp(\mathfrak{g}_{-1})$, we see that P is closed under the

map $s \mapsto s^{-1}$, and since $G = PK$, for every $s \in G$, we have $s = xy$ with $x \in P$ and $y \in K$, so $s^{-1} = y^{-1}x^{-1} \in KP$, and since this holds for any $s \in G$, we have $G = KP$. Therefore, (G, K) is a Gelfand pair for the involution σ.

If the compact Lie group G is also semi-simple, then its Killing form is negative definite, so G/K is a symmetric space of compact type.

(2) Let G_u be a real, compact, semi-simple, simply-connected Lie group, $\mathfrak{g}_u$ be its Lie algebra, $\mathfrak{g} = \mathfrak{g}_u \oplus i\mathfrak{g}_u$ be the complexification of $\mathfrak{g}_u$, and let G be the complex, semi-simple, simply-connected Lie group with Lie algebra $\mathfrak{g}$. For any conjugation $c_0 \colon \mathfrak{g} \to \mathfrak{g}$ that commutes with the conjugation c_u associated with $\mathfrak{g}_u$, let $\mathfrak{g}_0$ be the real form of $\mathfrak{g}$ induced by c_0. For any real form G_1 of G with Lie algebra $\mathfrak{g}_0$, let σ_u be the involution of G_1 such that $(d\sigma_u)_e = c_u$, as explained in Section 17.5. If the center C_1 of G_1 is finite, then $K_1 = \{s \in G_1 \mid \sigma_u(s) = s\}$ is a compact subgroup of G_1 such that (G_1, K_1) is Gelfand pair. The space G_1/K_1 is a symmetric space of non-compact type.

A typical example is given by $G = \mathbf{SL}(n, \mathbb{C})$, $G_u = \mathbf{SU}(n)$, $G_1 = \mathbf{SL}(n, \mathbb{R})$, and $K_1 = \mathbf{SO}(n, \mathbb{R})$; the maps c_0, c_u, and σ_0, are given by $c_0(X) = \overline{X}$, $c_u(X) = -X^*$, $\sigma_0(X) = \overline{X}$, $\sigma_u(X) = -X^*$.

(3) The group G is unimodular and contains

(a) A closed *commutative, normal* subgroup A such that $s^2 = e$ implies $s = e$ for all $s \in A$, and

(b) A compact subgroup K such that the mapping $(t, s) \mapsto ts$ from $K \times A$ to G is a homeomorphism. This implies that G is a semi-direct product of K and A, with A the normal factor. But beware that due to the order of the factors, since every element $g \in G = KA$ is written uniquely as $g = ka$ with $k \in K$ and $a \in A$, the multiplication in $G = KA$ is given by

$$(k_1a_1)(k_2a_2) = (k_1k_2)([k_2^{-1}a_1k_2]a_2),$$

where $k_1, k_2 \in K$ and $a_1, a_2 \in A$. So K acts on A by conjugation *on the right*. See Gallier and Quaintance [38], Section 19.5, Definition 19.20 and the remarks that follows. A typical example is $G = \mathbf{SE}(n, \mathbb{R})$.

Let σ be given by $\sigma(ts) = ts^{-1}$, for all $t \in K$ and all $s \in A$. Obviously σ is continuous, and we have

$$\sigma^2(ts) = \sigma(ts^{-1}) = ts,$$

so $\sigma^2 = \mathrm{id}_G$. For $t, t \in K$ and $s, s' \in A$, since A is a normal subgroup of G and $s \in A$, we have $t'^{-1}s^{-1}t' \in A$, and since $s' \in A$, so we also have $t'^{-1}st's' \in A$, and we have

$$\sigma(tst's') = \sigma((tt')(t'^{-1}st's')) = tt'(s'^{-1}t'^{-1}s^{-1}t'),$$

and
$$\sigma(ts)\sigma(t's') = ts^{-1}t's'^{-1} = tt'(t'^{-1}s^{-1}t's'^{-1}).$$

Since from above $t'^{-1}s^{-1}t' \in A$, and since A is abelian, $t'^{-1}s^{-1}t's'^{-1} = s'^{-1}t'^{-1}s^{-1}t'$, and so $\sigma(tst's') = \sigma(ts)\sigma(t's')$. Thus, σ is an involutive automorphism of G. By definition, $K = \{t \in G \mid \sigma(t) = t\}$, and $\sigma(s) = s^{-1}$ for $s \in A$. Therefore, (G, K) is a Gelfand pair.

Example 17.5. Assume the group G is compact. If (G, K) is a Gelfand pair, then the closure $\mathrm{L}^2(K\backslash G/K)$ of $\mathcal{K}(K\backslash G/K)$ in $\mathrm{L}^2(G)$ is commutative, which corresponds to the situation considered in Proposition 15.21. Furthermore, the restriction of every character of $\mathrm{L}^1(K\backslash G/K)$ to $\mathrm{L}^2(K\backslash G/K)$ is a (continuous) character of the algebra $\mathrm{L}^2(K\backslash G/K)$. We know that the direct sum
$$B = \bigoplus_{(\rho:\sigma_0)=1} \mathbb{C}\omega_\rho$$
is a dense algebra in $\mathrm{L}^2(K\backslash G/K)$, and it is easy to see that the only homomorphisms from B to $\mathbb{C}$ different from the zero function are the maps of the form $f \mapsto (f, \overline{\omega}_\rho)$, which shows that the spherical functions of G relative to K are the partial traces ω_ρ for all $\rho \in R(G)$ such that $(\rho : \sigma_0) = 1$ (see Definition 15.16 and Proposition 15.21). Since $\omega_\rho \in \mathrm{L}^1(G)$, the set of functions $f \in \mathrm{L}^\infty(G)$ such that $|(f - \omega_\rho, \overline{\omega}_\rho)| \leq \frac{1}{2}$ is a neighborhood of ω_ρ in the weak$\,^*$-topology of $\mathrm{L}^\infty(G)$. Since $(\omega_\rho, \overline{\omega}_\rho) = 1$ and $(\omega_{\rho'}, \overline{\omega}_\rho) = 0$ when $\rho \neq \rho'$, we conclude that the space $\mathbf{S}(G/K)$ is *discrete*.

The Grassmannians constitue a very good example. Let $G = \mathbf{SO}(n)$ (with $n \geq 2$), let
$$I_{k,n-k} = \begin{pmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{pmatrix},$$
where I_k is the $k \times k$-identity matrix, and let σ be given by
$$\sigma(P) = I_{k,n-k} P I_{k,n-k}, \quad P \in \mathbf{SO}(n).$$

It is clear that σ is an involutive automorphism of G. Let us find the set G^σ of fixed points of σ. If we write
$$P = \begin{pmatrix} Q & U \\ V & R \end{pmatrix}, \qquad Q \in M_{k,k}(\mathbb{R}),\ \ U \in M_{k,n-k}(\mathbb{R}),\ \ V \in M_{n-k,k}(\mathbb{R}),\ \ R \in M_{n-k,n-k}(\mathbb{R}),$$
then $P = I_{k,n-k} P I_{k,n-k}$ iff
$$\begin{pmatrix} Q & U \\ V & R \end{pmatrix} = \begin{pmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{pmatrix} \begin{pmatrix} Q & U \\ V & R \end{pmatrix} \begin{pmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{pmatrix}$$
iff
$$\begin{pmatrix} Q & U \\ V & R \end{pmatrix} = \begin{pmatrix} Q & -U \\ -V & R \end{pmatrix},$$

so $U = 0$, $V = 0$, $Q \in \mathbf{O}(k)$ and $R \in \mathbf{O}(n-k)$. Since $P \in \mathbf{SO}(n)$, we conclude that $\det(Q)\det(R) = 1$, so

$$G^\sigma = \left\{ \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix} \;\middle|\; Q \in \mathbf{O}(k),\ R \in \mathbf{O}(n-k),\ \det(Q)\det(R) = 1 \right\};$$

that is,

$$G^\sigma = S(\mathbf{O}(k) \times \mathbf{O}(n-k)).$$

We also have

$$G^\sigma_0 = \mathbf{SO}(k) \times \mathbf{SO}(n-k).$$

For $K = G^\sigma$, the homogeneous space

$$G/K = \mathbf{SO}(n)/(S(\mathbf{O}(k) \times \mathbf{O}(n-k)))$$

is the Grassmannian $G(k,n)$ of k-subspaces in $\mathbb{R}^n$. For $K = \mathbf{SO}(k) \times \mathbf{SO}(n-k)$, the homogeneous space

$$G/K = \mathbf{SO}(n)/(\mathbf{SO}(k) \times \mathbf{SO}(n-k))$$

is the Grassmannian $G^0(k,n)$ of oriented k-subspaces in $\mathbb{R}^n$. In particular, for $k = 1$, $G^0(1, n-1) = S^{n-1}$ and $G(1, n-1) = \mathbb{RP}^{n-1}$.

Example 17.6. Let G be a real semi-simple connected noncompact Lie group with finite center and let K be a maximal compact subgroup of K, so that (G,K) is a Gelfand pair, as in (2) above.

We showed that $G = KP$ where P is closed manifold in G, and P is closed under the map $s \mapsto s^{-1}$, but in general is not a group. However, it is known that there is closed solvable subgroup S of G such that $G = KS$, and that the map $(x,y) \mapsto xy$ from $K \times S$ to G is a diffeomorphism; for the definition of a solvable lie algebra and a solvable Lie group, see Gallier and Quaintance [38], Section 21.5, Definition 21.12. This is a corollary of the Iwasawa decomposition, which is a generalization of the QR-decomposition for invertible matrices; see Dieudonné [21] (Chapter XXI, no. 21.21.10). Since $(yx)^{-1} = x^{-1}y^{-1}$, the map $(y,x) \mapsto yx$ from $S \times K$ to G is also a diffeomorphism since it is the composition of the diffeomorphisms $(y,x) \mapsto (x^{-1}, y^{-1})$ from $S \times K$ to $K \times S$, the map $(x,y) \mapsto xy$ from $K \times S$ to G, and the map $s \mapsto s^{-1}$ from G to G. A way to construct spherical functions goes as follows.

Suppose we have a continuous homomorphism $\alpha\colon S \to \mathbb{C}^*$ (called an *exponential* of S). Then we can extend α to G as follows:

$$\alpha(st) = \alpha(s) \qquad \text{for all } s \in S \text{ and all } t \in K \tag{$*$}$$

We claim that the following properties hold

$$\begin{aligned} \alpha(xt) &= \alpha(x) && \text{for all } x \in G \text{ and all } t \in K \\ \alpha(sx) &= \alpha(s)\alpha(x) && \text{for all } s \in S \text{ and all } x \in G. \end{aligned}$$

Since for $x \in G$ we can write $x = st'$ with $s \in S$ and $t' \in K$, by $(*)$ we have

$$\alpha(xt) = \alpha(st't) = \alpha(s) = \alpha(st') = \alpha(x).$$

Similarly, we can write $x = s't$ with $s' \in S$ and $t \in K$, so by $(*)$, we have

$$\alpha(sx) = \alpha(ss't) = \alpha(ss') = \alpha(s)\alpha(s') = \alpha(s)\alpha(s't) = \alpha(s)\alpha(x).$$

Define the function $\omega\colon G \to \mathbb{C}$ by

$$\omega(x) = \int_K \alpha(tx)\, d\lambda_K(t). \tag{$\dagger_4$}$$

The function ω is continuous, and we we claim that if it is bounded, then it is a spherical function. By the remark just after Proposition 17.6, is suffices to prove that the equation

$$\int_K \omega(xt'y)\, d\lambda_K(t') = \omega(x)\omega(y) \qquad \text{for all } x, y \in G \tag{s_1}$$

holds. The left-hand side of this equation is

$$\int_K \omega(xt'y)\, d\lambda_K(t') = \int_K \int_K \alpha(txt'y)\, d\lambda_K(t)\, d\lambda_K(t').$$

We can also write for every fixed $x \in G$ and all $t \in K$, $tx = s(t)u(t)$, with $s(t) \in S$ and $u(t) \in K$, where s and u are continuous in $t \in K$, and using Fubini, the invariance of λ_K, and the equations just after $(*)$,

$$\begin{aligned}
\int_K \int_K \alpha(txt'y)\, d\lambda_K(t)\, d\lambda_K(t') &= \int_K \int_K \alpha(s(t)u(t)t'y)\, d\lambda_K(t)\, d\lambda_K(t') \\
&= \int_K \int_K \alpha(s(t)t'y)\, d\lambda_K(t)\, d\lambda_K(t') \\
&= \int_K \int_K \alpha(s(t))\alpha(t'y)\, d\lambda_K(t)\, d\lambda_K(t') \\
&= \int_K \int_K \alpha(s(t)u(t))\alpha(t'y)\, d\lambda_K(t)\, d\lambda_K(t') \\
&= \int_K \int_K \alpha(tx)\alpha(t'y)\, d\lambda_K(t)\, d\lambda_K(t') = \omega(x)\omega(y).
\end{aligned}$$

Harish–Chandra has shown that *all* continuous solutions of the functional equation (s_1) are given by $(\dagger_4)$. Such functions may be called generalized spherical functions. He also determined explicitly the exponentials $\alpha\colon S \to \mathbb{C}^*$ of S, by a very deep study of the Lie algebra of G. One also knows exactly when the generalized spherical functions are bounded, and thus the spherical functions are completely known.

Consider the groups $G = \mathbf{SL}(2,\mathbb{R})$, $K = \mathbf{SO}(2)$, and

$$S = S_0 = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \;\middle|\; a, b \in \mathbb{R},\ a > 0 \right\},$$

from Example 15.5. We know that $G = \mathbf{SL}(2,\mathbb{R})$ acts transitively on the upper half plane $P = \{z = x + iy \in \mathbb{C} \mid y > 0\}$ by the action given by

$$X \cdot z = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d}, \qquad z = x + iy \in P,$$

and that the stabilizer of i is $\mathbf{SO}(2)$. Given any $z = x + iy \in P$, there is a unique coset $X\mathbf{SO}(2) \subseteq \mathbf{SL}(2,\mathbb{R})$ (where $X \in \mathbf{SL}(2,\mathbb{R})$) that maps i to z, and in view of the unique factorization of matrix X in $\mathbf{SL}(2,\mathbb{R})$ as $X = st$ with $s \in S_0$ and $t \in K$, we can pick as a representative of this coset $X\mathbf{SO}(2)$ the matrix $s_z \in S_0$ such that

$$s_z \cdot i = z = x + iy,$$

namely

$$s_z = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix}.$$

We will show that the functions f in $\mathcal{C}(K\backslash G/K)$ are those which may be written as $f(1/2(a^2 + b^2 + c^2 + d^2))$, where f is a continuous function defined on the interval $[1, +\infty)$. The proof is fairly tedious and involves a geometric argument which identifies a double coset KXK (with $X \in \mathbf{SL}(2,\mathbb{R})$) as a circle, the orbit of a point ir on the imaginary axis $(0 < r \leq 1)$ under the action of K. We will also determine the exponentials α in terms of the continuous homomorphisms from $\mathbb{R}_+^*$ to $\mathbb{C}^*$.

Since every coset XK (with $X \in \mathbf{SL}(2,\mathbb{R})$) corresponds to a unique point $z \in P$ in the upper half plane, and since there is a unique matrix $s_z \in S_0$ such that $s_z \cdot i = z = x + iy$, the coset XK is uniquely represented by the matrix $s_z \in S_0$. It follows that the double coset KXK is uniquely determined by the set of matrices Ks_z, and geometrically this set of matrices corresponds to the orbit in P of the (left) action of $K = \mathbf{SO}(2)$ on the point z. Although this is not obvious, such an orbit is a circle centered on the y-axis. To show this, we will prove that it suffices to prove that the orbit of a point ir on the imaginary axis $(0 < r \leq 1)$ under the action of K is a circle of center iv and radius R, with

$$v = \frac{1}{2}\left(r + \frac{1}{r}\right), \qquad R = \frac{1}{2}\left|\frac{1}{r} - r\right|.$$

If so, the equation of this circle is $x^2 + (y - v)^2 = R^2$, that is,

$$x^2 + y^2 - 2yv = R^2 - v^2.$$

But

$$v^2 - 1 = \frac{1}{4}r^2 + \frac{1}{2} + \frac{1}{4r^2} - 1 = \frac{1}{4}r^2 - \frac{1}{2} + \frac{1}{4r^2} = \left(\frac{1}{2}\left(\frac{1}{r} - r\right)\right)^2 = R^2,$$

so $R^2 - v^2 = -1$, and the equation of our circle is

$$x^2 + y^2 + 1 = 2yv. \tag{$*_1$}$$

We also showed that $R = \sqrt{v^2 - 1}$.

We can now prove that the orbit of any point $z = x + iy$ in the upper half plane under the action of K is a circle centered on the y-axis, which is also the orbit of a point ir with $0 < r \leq 1$ under the action of K. Observe that since $R = \sqrt{v^2 - 1}$, $R > 0$ and $v > 0$, we have $v \geq 1$. We have to find $r > 0$ such that

$$r + \frac{1}{r} = 2v,$$

that is

$$r^2 - 2vr + 1 = 0,$$

and the zeros of this equation are

$$r = v \pm \sqrt{v^2 - 1}.$$

If $r = v - \sqrt{v^2 - 1}$, then $0 < r \leq 1$. Therefore, we found that $z \in P$ is on the circle of center iv and radius R, the orbit of ir by the action of K, with

$$v = \frac{x^2 + y^2 + 1}{2y}, \qquad R = \sqrt{v^2 - 1}, \qquad r = v - \sqrt{v^2 - 1} = v - R.$$

It remains to prove that the orbit of the point ir with $0 < r \leq 1$ under the action of K is the circle of center iv and radius R, with

$$v = \frac{1}{2}\left(r + \frac{1}{r}\right), \qquad R = \sqrt{v^2 - 1}.$$

Since a rotation in $K = \mathbf{SO}(2)$ is of the form

$$t_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

the orbit of ir consists of the points $x_\theta + iy_\theta = t_\theta \cdot ir$, with

$$x_\theta + iy_\theta = \frac{ir\cos\theta - \sin\theta}{ir\sin\theta + \cos\theta}. \tag{$*_2$}$$

From ($*_2$), we have

$$\begin{aligned}
x_\theta + iy_\theta &= \frac{ir\cos\theta - \sin\theta}{ir\sin\theta + \cos\theta} = \frac{ir\cos\theta - \sin\theta}{ir\sin\theta + \cos\theta}\,\frac{-ir\sin\theta + \cos\theta}{-ir\sin\theta + \cos\theta} \\
&= \frac{(r^2 - 1)\sin\theta\cos\theta + ir}{r^2\sin^2\theta + \cos^2\theta} = \frac{\frac{(r^2-1)}{2}2\sin\theta\cos\theta + ir}{(1 - r^2)\cos^2\theta + r^2} \\
&= \frac{\frac{(r^2-1)}{2}\sin 2\theta + ir}{\frac{(1-r^2)}{2}(2\cos^2\theta - 1) + r^2 + \frac{1-r^2}{2}} \\
&= \frac{-(1 - r^2)\sin 2\theta + i2r}{(1 - r^2)\cos 2\theta + r^2 + 1}.
\end{aligned}$$

Let us compute $x_\theta + iy_\theta - iv$. We have

$$\begin{aligned}
x_\theta + iy_\theta - iv &= \frac{-(1-r^2)\sin 2\theta + i2r}{(1-r^2)\cos 2\theta + r^2 + 1} - \frac{i(r^2+1)}{2r} \\
&= \frac{-2r(1-r^2)\sin 2\theta + i4r^2 - i(r^2+1)((1-r^2)\cos 2\theta + r^2 + 1)}{2r((1-r^2)\cos 2\theta + r^2 + 1)} \\
&= \frac{-2r(1-r^2)\sin 2\theta + i4r^2 - i(r^2+1)^2 - i(r^2+1)(1-r^2)\cos 2\theta}{2r((1-r^2)\cos 2\theta + r^2 + 1)} \\
&= \frac{(1-r^2)(-2r\sin 2\theta - i(1-r^2+(r^2+1)\cos 2\theta))}{2r((1-r^2)\cos 2\theta + r^2 + 1)}.
\end{aligned}$$

The numerator of $|x_\theta + i(y_\theta - v)|^2$ is

$$\begin{aligned}
N &= (1-r^2)^2(4r^2\sin^2 2\theta + (1-r^2)^2 + 2(1-r^2)(r^2+1)\cos 2\theta + (r^2+1)^2\cos^2 2\theta) \\
&= (1-r^2)^2(4r^2(1-\cos^2 2\theta) + (1-r^2)^2 + 2(1-r^2)(r^2+1)\cos 2\theta + (r^2+1)^2\cos^2 2\theta) \\
&= (1-r^2)^2((1-r^2)^2\cos^2 2\theta + 2(1-r^2)(r^2+1)\cos 2\theta + (1+r^2)^2) \\
&= (1-r^2)^2((1-r^2)\cos 2\theta + (1+r^2))^2.
\end{aligned}$$

The denominator of $|x_\theta + i(y_\theta - v)|^2$ is

$$4r^2((1-r^2)\cos 2\theta + (1+r^2))^2.$$

Therefore

$$|x_\theta + i(y_\theta - v)|^2 = \frac{(1-r^2)^2}{4r^2} = \left(\frac{1}{2}\left(\frac{1}{r} - r\right)\right)^2 = R^2 = v^2 - 1,$$

or equivalenty

$$x_\theta^2 + (y_\theta - v)^2 = v^2 - 1,$$

which confirms our claim that the point $x_\theta + iy_\theta$ is on the circle of center v and radius $R = \sqrt{v^2-1}$.

Therefore the matrices of the double class KXK are exactly the matrices $s_\theta t$ where $t \in K = \mathbf{SO}(2)$ and $s_\theta \in S_0$ is the matrix

$$s_\theta = \begin{pmatrix} \sqrt{y_\theta} & x_\theta/\sqrt{y_\theta} \\ 0 & 1/\sqrt{y_\theta} \end{pmatrix}$$

corresponding to the point $x_\theta + iy_\theta = t_\theta \cdot ir$ in $(*_2)$. The matrices s_θ satisfy the property

$$\mathrm{tr}(s_\theta^\top s_\theta) = \frac{x_\theta^2 + y_\theta^2 + 1}{y_\theta} = 2v,$$

which characterizes them. Now one has $(s_\theta t)^\top(s_\theta t) = t^\top(s_\theta^\top s_\theta)t$, and as $t^\top = t^{-1}$, we get

$$\mathrm{tr}((s_\theta t)^\top(s_\theta t)) = \mathrm{tr}(s_\theta^\top s_\theta).$$

We deduce that the matrices

$$X = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G = \mathbf{SL}(2, \mathbb{R}), \quad \text{with} \quad X = s_\theta t,$$

which form the double class KXK, are exactly those for which

$$\mathrm{tr}(X^\top X) = a^2 + b^2 + c^2 + d^2 = 2v,$$

with $v \geq 1$. Hence the functions $f \in \mathcal{C}(K \backslash G/K)$ are those which may be written as $f(1/2(a^2 + b^2 + c^2 + d^2))$, where f is a continuous function defined on the interval $[1, +\infty)$.

We can also determine the exponentials $\alpha\colon S \to \mathbb{C}^*$ of S. Given any group G, recall that for any two elements $a, b \in G$, the element $a^{-1}b^{-1}ab$ is the *commutator* of a and b, and that the subgroup of G generated by the commutators is called the *commutator subgroup* of G and is denoted by DG. If $\alpha\colon S \to \mathbb{C}^*$ is a (continuous) homomorphism, then obviously α has the value 1 on the commutator subgroup DS of S (since $\alpha(a^{-1}b^{-1}ab) = \alpha(a^{-1})\alpha(b^{-1})\alpha(a)\alpha(b) = \alpha(a)^{-1}\alpha(b)^{-1}\alpha(a)\alpha(b) = 1$). But since

$$\begin{pmatrix} a_1 & b_1 \\ 0 & a_1^{-1} \end{pmatrix} \begin{pmatrix} a_2 & b_2 \\ 0 & a_2^{-1} \end{pmatrix} = \begin{pmatrix} a_1 a_2 & a_1 b_2 + b_1 a_2^{-1} \\ 0 & a_1^{-1} a_2^{-1} \end{pmatrix}$$

for any two matrices in S, we see that

$$DS = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \;\middle|\; b \in \mathbb{R} \right\}.$$

Furthermore, every matrix $X \in S$ can be factored as

$$X = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & a^{-1}b \\ 0 & 1 \end{pmatrix},$$

with $a > 0$, so the homomorphism α is determined by its restriction to the subgroup

$$\left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \;\middle|\; a > 0 \right\} \cong \mathbb{R}_+^*,$$

and thus it corresponds to a continuous homomorphism from $\mathbb{R}_+^*$ to $\mathbb{C}^*$, which is well known to be of the form $t \mapsto t^\lambda = e^{\lambda \log t}$, for some $\lambda \in \mathbb{C}$. This is because the map $x \mapsto e^x$ is a continuous homomorphism from $(\mathbb{R}, +)$ to $\mathbb{R}_+^*$, and every continuous homomorphism from $(\mathbb{R}, +)$ to $\mathbb{C}^*$ is of the form $x \mapsto e^{\lambda x}$ for some $\lambda \in \mathbb{C}$, as the proof of Proposition 10.9(4) shows. We showed earlier that

$$\begin{aligned} x + iy &= \frac{ir\cos\theta - \sin\theta}{ir\sin\theta + \cos\theta} \\ &= \frac{(r^2 - 1)\sin\theta\cos\theta + ir}{r^2\sin^2\theta + \cos^2\theta}, \end{aligned}$$

so we get

$$\frac{1}{y} = r\sin^2\theta + \frac{1}{r}\cos^2\theta.$$

Using the fact that

$$v = \frac{1}{2}\left(r + \frac{1}{r}\right),$$

we get

$$\begin{aligned}
\frac{1}{y} &= r\sin^2\theta + \frac{1}{r}\cos^2\theta \\
&= v + r\sin^2\theta - \frac{1}{2}r + \frac{1}{r}\cos^2\theta - \frac{1}{2r} = v + \frac{r}{2}(2\sin^2\theta - 1) + \frac{1}{2r}(2\cos^2\theta - 1) \\
&= v + \frac{r}{2}(1 - 2\cos^2\theta) + \frac{1}{2r}(2\cos^2\theta - 1) = v + \frac{1}{2}\left(\frac{1}{r} - r\right)\cos 2\theta \\
&= v + \sqrt{v^2 - 1}\cos 2\theta.
\end{aligned}$$

Since

$$X = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix},$$

the above reasoning shows that

$$\alpha(XU) = \alpha(X) = e^{\lambda\log(\sqrt{y})} = e^{\frac{1}{2}\lambda\log y},$$

for some complex number $\lambda \in \mathbb{C}$, where $U \in \mathbf{SO}(2)$. However it is more convenient to express $\alpha(X)$ in terms of $1/y$, so we write

$$\alpha(X) = e^{-\rho\log y} = y^{-\rho},$$

with $\rho = -\frac{1}{2}\lambda$, so finally, the generalized spherical function given by ($\dagger_4$) is

$$P_\rho(v) = \frac{1}{2\pi}\int_0^{2\pi} (v + \sqrt{v^2 - 1}\cos\varphi)^\rho \, d\varphi, \qquad v \geq 1,\ \rho \in \mathbb{C}.$$

The function P_ρ a *Legendre function of (complex) index* ρ. When ρ is a positive integer n, it can be shown that up to a constant, P_n is a Legendre polynomial.

In order for the function P_ρ to be bounded when $v \geq 1$, some conditions must be imposed on ρ. If ρ is purely imaginary, then P_ρ is bounded, but if ρ is a positive real, then it is not bounded.

One can check that the functional equation (s_1) becomes

$$\frac{1}{2\pi}\int_0^{2\pi} P_\rho(\cosh t\cosh u + \sinh t\sinh u\cos\varphi)\, d\varphi = P_\rho(\cosh t)P_\rho(\cosh u),$$

for all $t, u \in \mathbb{R}$.

Example 17.7. Let now now consider Case (3) above, where G is a unimodular group containing an abelian normal subgroup A and a compact subgroup K such that the map $(t,s) \mapsto ts$ is a homeomorphism from $K \times A$ to G. Let $\alpha\colon A \to \mathbb{C}^*$ be a continuous homomorphism (an exponential of A). By analogy with Example 17.6, define the function $\omega\colon G \to \mathbb{C}$ by

$$\omega(x) = \int_K \alpha(usu^{-1})\, d\lambda_K(u), \qquad x = ts,\ t \in K,\ s \in A. \tag{$\dagger_5$}$$

The function ω is continuous, and we claim that if ω is bounded, then it is a spherical function for (G,K). For this, we verify that the functional equation (s_1) holds.

Let $x = t_1s_1, y = t_2s_2$, with $t_1,t_2 \in K,\ s_1,s_2 \in A$. For $v \in K$, we may write

$$xvy = t_1s_1vt_2s_2 = (t_1vt_2)(((vt_2)^{-1}s_1(vt_2))s_2),$$

with $t_1vt_2 \in K$ and $((vt_2)^{-1}s_1(vt_2))s_2 \in A$, because A is normal, so $(vt_2)^{-1}s_1(vt_2) \in A$, and $((vt_2)^{-1}s_1(vt_2))s_2 \in A$. Consequently, since α is a homomorphism of A, the subgroup A is a normal subgroup, and by Fubini, we have

$$\begin{aligned}
\int_K \omega(xvy)\, d\lambda_K(v) &= \int_K\int_K \alpha(u((vt_2)^{-1}s_1(vt_2))s_2u^{-1})\, d\lambda_K(u)\, d\lambda_K(v)\\
&= \int_K\int_K \alpha((((vt_2u^{-1})^{-1}s_1(vt_2u^{-1}))(us_2u^{-1}))\, d\lambda_K(u)\, d\lambda_K(v)\\
&= \int_K\int_K \alpha((vt_2u^{-1})^{-1}s_1(vt_2u^{-1}))\alpha(us_2u^{-1})\, d\lambda_K(u)\, d\lambda_K(v)\\
&= \int_K \alpha(us_2u^{-1}) \int_K \alpha((vt_2u^{-1})^{-1}s_1(vt_2u^{-1}))\, d\lambda_K(v)\, d\lambda_K(u).
\end{aligned}$$

But since K is unimodular, we have

$$\int_K \alpha((vt_2u^{-1})^{-1}s_1(vt_2u^{-1}))\, d\lambda_K(v) = \int_K \alpha(v^{-1}s_1v)\, d\lambda_K(v) = \int_K \alpha(vs_1v^{-1})\, d\lambda_K(v) = \omega(x),$$

and thus

$$\begin{aligned}
\int_K \omega(xvy)\, d\lambda_K(v) &= \int_K \alpha(us_2u^{-1}) \int_K \alpha((vt_2u^{-1})^{-1}s_1(vt_2u^{-1}))\, d\lambda_K(v)\, d\lambda_K(u)\\
&= \int_K \alpha(us_2u^{-1})\omega(x)\, d\lambda_K(u) = \omega(x)\omega(y),
\end{aligned}$$

as claimed.

Conversely, it can be shown that all spherical functions are given by $(\dagger_5)$. A proof is sketched in Dieudonné [20] (Chapter 16).

Consider the example $G = \mathbf{SE}(2,\mathbb{R})$ of the group of rigid motions of $\mathbb{R}^2$. Since we view this group as the semi-direct product of $\mathbf{SO}(2)$ and $\mathbb{R}^2$ (instead of $\mathbb{R}^2$ and $\mathbf{SO}(2)$), we want

a matrix representation in which every rigid motion is written as the product of a rotation and a translation so we view $\mathbf{SE}(2,\mathbb{R})$ as

$$\mathbf{SE}(2,\mathbb{R}) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ a & b & 1 \end{pmatrix} \;\middle|\; a, b \in \mathbb{R},\ 0 \leq \theta \leq 2\pi \right\},$$

instead of

$$\begin{pmatrix} \cos\theta & -\sin\theta & a \\ \sin\theta & \cos\theta & b \\ 0 & 0 & 1 \end{pmatrix}.$$

Clearly,

$$\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ a & b & 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & 1 \end{pmatrix},$$

so $\mathbf{SE}(2,\mathbb{R}) = KA$, where K (the rotations) is isomorphic to $\mathbf{SO}(2)$ and A (the translations) is isomorphic to $\mathbb{R}^2$. Note that in this representation of $\mathbf{SE}(2,\mathbb{R}^2)$, we use matrices

$$s = \begin{pmatrix} Q^\top & 0 \\ u^\top & 1 \end{pmatrix}$$

with $u \in \mathbb{R}^2$ and $Q \in \mathbf{SO}(2)$, and we have the right action given

$$x \cdot s = x^\top s = x^\top Q^\top + u^\top = (Qx + u)^\top \quad x \in \mathbb{R}^2,$$

which corresponds to the matrix equation

$$\begin{pmatrix} y^\top & 1 \end{pmatrix} = \begin{pmatrix} x^\top & 1 \end{pmatrix} \begin{pmatrix} Q^\top & 0 \\ u^\top & 1 \end{pmatrix}.$$

The choice of this representation forces everything to be transposed. In particular, if we denote the matrix

$$\begin{pmatrix} Q^\top & 0 \\ u^\top & 1 \end{pmatrix}$$

by $(Q^\top, u^\top)$, since the product of the matrices

$$s = \begin{pmatrix} Q^\top & 0 \\ u^\top & 1 \end{pmatrix}, \quad t = \begin{pmatrix} R^\top & 0 \\ v^\top & 1 \end{pmatrix}$$

is

$$st = \begin{pmatrix} Q^\top R^\top & 0 \\ u^\top R^\top + v^\top & 1 \end{pmatrix},$$

the multiplication operation is given by

$$(Q^\top, u^\top)(R^\top, v^\top) = (Q^\top R^\top, v^\top + u^\top R^\top) = ((RQ)^\top, (v + Ru)^\top).$$

When $\mathbf{SE}(2,\mathbb{R})$ is viewed as the semi-direct product of $\mathbb{R}^2$ and $\mathbf{SO}(2)$, we use the representation (u, Q), and multiplication is given by

$$(v, R)(u, Q) = (v + Ru, RQ),$$

which corresponds to $(Q^\top, u^\top)(R^\top, v^\top)$ by transposition, but note the reversal of the arguments in the multiplication (which must take place since transposition of a product of matrices switches the order of the arguments).

For any matrix $s \in \mathbf{SE}(2,\mathbb{R})$ and any matrices $t_1, t_2 \in \mathbf{SO}(2)$, if we write

$$s = \begin{pmatrix} R^\top & 0 \\ u^\top & 1 \end{pmatrix}, \quad t_1 = \begin{pmatrix} Q_1^\top & 0 \\ 0 & 1 \end{pmatrix}, \quad t_2 = \begin{pmatrix} Q_2^\top & 0 \\ 0 & 1 \end{pmatrix}, \qquad Q_1, Q_2, R \in \mathbf{SO}(2),\ u \in \mathbb{R}^2,$$

then we have

$$t_1 s t_2 = \begin{pmatrix} Q_1^\top & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R^\top & 0 \\ u^\top & 1 \end{pmatrix} \begin{pmatrix} Q_2^\top & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} Q_1^\top R^\top & 0 \\ u^\top & 1 \end{pmatrix} \begin{pmatrix} Q_2^\top & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} Q_1^\top R^\top Q_2^\top & 0 \\ u^\top Q_2^\top & 1 \end{pmatrix}.$$

If we pick $Q_1 = (Q_2 R)^\top$ then we see that the matrix

$$\begin{pmatrix} I & 0 \\ u^\top Q_2^\top & 1 \end{pmatrix}$$

belong to the class KsK. Since $\mathbf{SO}(2)$ acts transitively on $\mathbb{R}^2$ it follows that the matrix

$$\begin{pmatrix} I & 0 \\ v^\top & 1 \end{pmatrix}$$

also belongs to double class KsK for any vector v such that $\|v\| = \|u\| = r^2$ $(r \geq 0)$. Therefore every double class KsK corresponds bijectively to some $r \in \mathbb{R}$ with $r \geq 0$. We can also view such a double class KsK as any vector (a, b) for which $a^2 + b^2 = r^2$ for a fixed $r \geq 0$ in $\mathbb{R}$. The functions in $\mathcal{C}(K\backslash G/K)$ are those of the form $\psi(r)$, where $\psi\colon [0, +\infty) \to \mathbb{C}$ is any continuous function. We showed in Corollary 10.11 that the continuous homomorphisms $\alpha\colon \mathbb{R}^2 \to \mathbb{C}^*$ are of the form

$$\alpha(a, b) = e^{\lambda a + \mu b}, \qquad \lambda, \mu \in \mathbb{C}.$$

For any $u \in K$ and any $s \in A$, we have

$$\begin{aligned} usu^{-1} &= \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ a & b & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a\cos\theta + b\sin\theta & -a\sin\theta + b\cos\theta & 1 \end{pmatrix}. \end{aligned}$$

For $a = r\cos\varphi$ and $b = r\sin\varphi$, so that $a^2 + b^2 = r^2$, we have

$$usu^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ r\cos(\varphi - \theta) & r\sin(\varphi - \theta) & 1 \end{pmatrix}.$$

Consequently, according to ($\dagger_5$), for

$$x = ts = t\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ r\cos\varphi & r\sin\varphi & 1 \end{pmatrix}, \qquad t \in K,$$

we have

$$\begin{aligned} \omega(x) &= \int_K \alpha(usu^{-1})\, d\lambda_K(u) \\ &= \frac{1}{2\pi}\int_0^{2\pi} e^{r(\lambda\cos(\varphi-\theta)+\mu\sin(\varphi-\theta))}\, d\theta = \frac{1}{2\pi}\int_0^{2\pi} e^{r(\lambda\cos\theta+\mu\sin\theta)}\, d\theta. \end{aligned}$$

It follows that the generalized spherical functions are the continuous functions on $[0, +\infty)$ given by

$$\psi(r) = \frac{1}{2\pi}\int_0^{2\pi} e^{r(\lambda\cos\theta+\mu\sin\theta)}\, d\theta,$$

for any $\lambda, \mu \in \mathbb{C}$. For λ and μ imaginary, these functions are bounded, hence they really are spherical functions. In the special case where $\lambda = 0$ and $\mu = i$, the function ψ is the *Bessel function* J_0. By a change of variable, if both λ and μ are imaginary, the function ψ becomes J_0.

The irreducible unitary representations of $\mathbf{SE}(2, \mathbb{R})$ are completely determined and can be expressed by means of Bessel functions; see Vilenkin [101].

In the general case where $G = \mathbf{SE}(n, \mathbb{R})$, with $K = \mathbf{SO}(n)$ and $A = \mathbb{R}^n$, with

$$\mathbf{SE}(n, \mathbb{R}) = \left\{ \begin{pmatrix} Q^\top & 0 \\ w^\top & 1 \end{pmatrix} \;\middle|\; Q \in \mathbf{SO}(n),\ w \in \mathbb{R}^n \right\},$$

one can also determine the generalized spherical functions, and they are now expressed in terms of the Bessel functions $J_{(n-2)/2}$; see Vilenkin [101].

In general, if G is a connected unimodular Lie group and K is a compact subgroup of G such that (G, K) is a Gelfand pair, it can be shown that the spherical functions are not only continuous but also *smooth*. The proof is not difficult but not that informative, so we omit it. This proof can be found in Dieudonné [20] (Chapter 12).

It is also possible to figure out how a differential operator on G that is invariant by left translations by G and invariant by right translations by K operates on spherical functions. In this case, the spherical functions are *eigenfunctions* of all such differential operators. If G is also semi-simple, then more can be said (there are *elliptic* operators, in particular, the *Casimir operator*), but will not go into this right now. These topics are discussed in Dieudonné [23] (Chapter XXIII, Sections 36 and 37).

17.7 The Fourier Transform

Again, let (G, K) be a Gelfand pair. Recall from Definition 17.7 that every spherical function $\omega \in \mathbf{S}(G/K)$ defines the character $\zeta_\omega \in \mathsf{X}_0(A)$ given by

$$\zeta_\omega(f) = (f, \omega) = \int_G f(x)\omega(x)\, d\lambda_G(x), \qquad f \in \mathrm{L}^1(K\backslash G/K),$$

where $A = \mathrm{L}^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$, a commutative, involutive, unital Banach algebra. By Theorem 17.7 the map $\omega \mapsto \zeta_\omega$ is a homeomorphism of $\mathbf{S}(G/K)$ equipped with the induced topology of Fréchet space of $\mathcal{C}(G)$ onto $\mathsf{X}_0(A)$ equipped with the topology induced by the weak*-topology of the dual A' of A.

It follows that the restriction of the Gelfand transform to $\mathsf{X}_0(A)$ of an element $f \in \mathrm{L}^1(K\backslash G/K)$ can be identified with the function $\overline{\mathcal{F}}f \colon \mathbf{S}(G/K) \to \mathbb{C}$ given by

$$(\overline{\mathcal{F}}f)(\omega) = (f, \omega) = \int_G f(x)\omega(x)\, d\lambda_G(x).$$

The above can be viewed as the Fourier cotransform of f. Thus we are led to the following definition.

Definition 17.20. Let (G, K) be a Gelfand pair (recall that G is unimodular). For every function $f \in \mathcal{L}^1(K\backslash G/K)$, the *Fourier cotransform* $\overline{\mathcal{F}}f$ of f is the function $\overline{\mathcal{F}}f \colon \mathbf{S}(G/K) \to \mathbb{C}$ given by

$$(\overline{\mathcal{F}}f)(\omega) = (f, \omega) = \int_G f(x)\omega(x)\, d\lambda_G(x), \qquad \omega \in \mathbf{S}(G/K),$$

and the *Fourier transform* $\mathcal{F}f$ of f is the function $\mathcal{F}f \colon \mathbf{S}(G/K) \to \mathbb{C}$ given by

$$\begin{aligned}(\mathcal{F}f)(\omega) = (\check{f}, \omega) &= \int_G f(x^{-1})\omega(x)\, d\lambda_G(x)\\ = (f, \check{\omega}) &= \int_G f(x)\omega(x^{-1})\, d\lambda_G(x), \qquad \omega \in \mathbf{S}(G/K).\end{aligned}$$

Observe that

$$\mathcal{F}f = \overline{\mathcal{F}}\check{f}.$$

It is also clear that $\mathcal{F}f$ and $\overline{\mathcal{F}}f$ depend only on the equivalence class $[f] \in \mathrm{L}^1(K\backslash G/K)$, so the Fourier transform and the Fourier cotransform are also defined on $\mathrm{L}^1(K\backslash G/K)$.

Definition 17.20 also applies to arbitrary functions $f \in \mathcal{L}^1(G)$. By $(**)$ of Section 17.2, namely

$$(f^\sharp, \psi) = \int_G f^\sharp(x)\psi(x)\, d\lambda_G(x) = \int_G f(x)\psi^\sharp(x)\, d\lambda_G(x) = (f, \psi^\sharp), \tag{$**$}$$

for all $f \in \mathcal{K}(G)$ and all $\psi \in \mathcal{C}(G)$, since $\omega^\sharp = \omega$, we obtain

$$\mathcal{F}f = \mathcal{F}(f^\sharp), \qquad \overline{\mathcal{F}}f = \overline{\mathcal{F}}(f^\sharp), \qquad \text{for all } f \in \mathcal{L}^1(G).$$

The Fourier transform and the Fourier cotransform are also defined on $\mathrm{L}^1(G)$.

As a consequence of the properties of the Gelfand transform, we have the following results.

Proposition 17.18. *Let (G, K) be a Gelfand pair. For every function $f \in \mathcal{L}^1(G)$, the Fourier transform $\mathcal{F}f$ and the Fourier cotransform $\overline{\mathcal{F}}f$ are continuous functions that tend to zero at infinity. For all $f, g \in \mathcal{L}^1(K\backslash G/K)$, we have*

$$\mathcal{F}(f * g) = (\mathcal{F}f)(\mathcal{F}g), \qquad \overline{\mathcal{F}}(f * g) = (\overline{\mathcal{F}}f)(\overline{\mathcal{F}}g). \tag{$*$}$$

For all $f \in \mathcal{L}^1(G)$, we have

$$\|\mathcal{F}f\| \leq \|f\|_1, \qquad \left\|\overline{\mathcal{F}}f\right\| \leq \|f\|_1.$$

Therefore $\mathcal{F}$ and $\overline{\mathcal{F}}$ are continuous linear maps from $\mathrm{L}^1(G)$ to $\mathcal{C}_0(\mathbf{S}(G/K); \mathbb{C})$.

Beware that the equations in $(*)$ *generally fail* if $f, g \in \mathcal{L}^1(G)$. Also, in general, even if $f \in \mathcal{K}_{\mathbb{C}}(G)$, the functions $\mathcal{F}f$ and $\overline{\mathcal{F}}f$ *do not* have compact support. However, we have the following properties (see Dieudonné [22] (Chapter XXII, Proposition 22.6.4.7).

Proposition 17.19. *Let f and g be two functions in $\mathcal{L}^1(G)$. If either $f \in \mathcal{L}^1(G/K)$ or $g \in \mathcal{L}^1(K\backslash G)$, then*

$$\mathcal{F}(f * g) = (\mathcal{F}f)(\mathcal{F}g), \qquad \textit{and} \qquad \overline{\mathcal{F}}(f * g) = (\overline{\mathcal{F}}f)(\overline{\mathcal{F}}g).$$

Proof. We prove the first equation assuming that $f \in \mathcal{L}^1(G/K)$, the proof of the other equations being similar. By left-invariance, we have

$$\begin{aligned}
\mathcal{F}(f * g)(\omega) &= \int_G \int_G \omega(x^{-1}) f(s) g(s^{-1}x)\, d\lambda_G(s)\, d\lambda_G(x) \\
&= \int_G \int_G \omega(x^{-1}s^{-1}) f(s) g(x)\, d\lambda_G(s)\, d\lambda_G(x),
\end{aligned}$$

and by right-invariance, since $f(st^{-1}) = f(s)$ for all $t \in K$, $\lambda_K(K) = 1$, the fact that $\mathcal{F}(f * g)(\omega)$ is independent of t, and by (s_1) of Theorem 17.6, we have

$$\begin{aligned}
\mathcal{F}(f * g)(\omega) &= \int_G \int_G \omega(x^{-1}s^{-1}) f(s) g(x)\, d\lambda_G(s)\, d\lambda_G(x) \\
&= \int_G \int_G \omega(x^{-1}ts^{-1}) f(st^{-1}) g(x)\, d\lambda_G(s)\, d\lambda_G(x) \\
&= \int_G \int_G \omega(x^{-1}ts^{-1}) f(s) g(x)\, d\lambda_G(s)\, d\lambda_G(x) \\
&= \int_G \int_G \int_K \omega(x^{-1}ts^{-1})\, d\lambda_K(t) f(s) g(x)\, d\lambda_G(s)\, d\lambda_G(x) \\
&= \int_G \int_G \omega(x^{-1}) \omega(s^{-1}) f(s) g(x)\, d\lambda_G(s)\, d\lambda_G(x) \\
&= \int_G \omega(s^{-1}) f(s)\, d\lambda_G(s) \int_G \omega(x^{-1}) g(x)\, d\lambda_G(x) = (\mathcal{F}f)(\omega)(\mathcal{F}g)(\omega),
\end{aligned}$$

as claimed. $\square$

In general, $\mathcal{F}(f * g) \neq (\mathcal{F}f)(\mathcal{F}g)$ if $f \notin \mathcal{L}^1(G/K)$ and $g \notin \mathcal{L}^1(K\backslash G)$.

The following properties also hold.

Proposition 17.20. *Let (G, K) be a Gelfand pair. For every function $f \in \mathcal{L}^1(K\backslash G)$, every $s \in G$, and every $\omega \in \mathbf{S}(G/K)$, we have*

$$\mathcal{F}(\lambda_s f)(\omega) = \omega(s^{-1})(\mathcal{F}f)(\omega), \qquad \overline{\mathcal{F}}(\lambda_s f)(\omega) = \omega(s)(\overline{\mathcal{F}}f)(\omega),$$

and for every function $f \in \mathcal{L}^1(G/K)$, every $s \in G$, and every $\omega \in \mathbf{S}(G/K)$, we have

$$\mathcal{F}(\rho_s f)(\omega) = \omega(s)(\mathcal{F}f)(\omega), \qquad \overline{\mathcal{F}}(\rho_s f)(\omega) = \omega(s^{-1})(\overline{\mathcal{F}}f)(\omega).$$

Proof. We prove that $\overline{\mathcal{F}}(\lambda_s f)(\omega) = \omega(s)(\overline{\mathcal{F}}f)(\omega)$, the proof for the other formulae being similar. Since $f(tx) = f(x)$ for all $t \in K$ and almost all $x \in G$, and since λ_G is left-invariant, we have

$$\begin{aligned}\overline{\mathcal{F}}(\lambda_s f)(\omega) &= \int_G f(s^{-1}x)\omega(x)\, d\lambda_G(x) = \int_G f(x)\omega(sx)\, d\lambda_G(x) \\ &= \int_G f(tx)\omega(stx)\, d\lambda_G(x) = \int_G f(x)\omega(stx)\, d\lambda_G(x).\end{aligned}$$

Then, since the rightmost integral above is independent of t because $\overline{\mathcal{F}}(\lambda_s f)(\omega)$ is independent of t, $\overline{\mathcal{F}}(\lambda_s f)(\omega)$ is independent of t, by (s_1) (from Theorem 17.6) and Fubini, we have

$$\begin{aligned}\int_G f(x)\omega(stx)\, d\lambda_G(x) &= \int_G \int_K f(x)\omega(stx)\, d\lambda_G(x)\, d\lambda_K(t) \\ &= \int_G f(x) \int_K \omega(stx)\, d\lambda_K(t)\, d\lambda_G(x) \\ &= \omega(s) \int_G f(x)\omega(x)\, d\lambda_G(x) = \omega(s)(\overline{\mathcal{F}}f)(\omega),\end{aligned}$$

as claimed. $\square$

In the next section we try to generalize Fourier inversion.

17.8 The Plancherel Transform

As in the previous section, let (G, K) be a Gelfand pair. If G is compact, then by Example 17.5 and Proposition 15.21, the spherical functions in $\mathbf{S}(G/K)$ are of positive type (recall Definition 12.14). However, when G is not compact, the spherical functions in $\mathbf{S}(G/K)$ are not necessarily of positive type. The subspace of spherical functions of positive type is deeply related to the measures of positive type (recall Definition 12.18) and is the domain of certain positive measures that yield a kind of Fourier inversion.

Definition 17.21. The subset of $\mathbf{S}(G/K)$ consisting of the *spherical functions of positive type* is denoted by $\mathbf{Z}(G/K)$. This space is equipped with the induced topology of Fréchet space of $\mathcal{C}(G)$.

In view of Theorem 12.19(b), the space $\mathbf{Z}(G/K)$ is closed in $\mathbf{S}(G/K)$, and thus it is locally compact.

Given a measure of positive type μ on G (see Definition 12.18), recall from Section 12.7 that the linear map $\varphi_\mu \colon \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C}$ given by

$$\varphi_\mu(f) = \int f(s)\, d\mu(s)$$

is a positive linear form in the sense of Definition 11.10. As in Section 12.5, the set

$$\mathfrak{n} = \{f \in \mathcal{K}_{\mathbb{C}}(G) \mid \varphi_\mu(f^* * f) = 0\}$$

is a left ideal in $\mathcal{K}_{\mathbb{C}}(G)$, and $H_0 = \mathcal{K}_{\mathbb{C}}(G)/\mathfrak{n}$ is a hermitian space with the hermitian inner product

$$\langle \pi(f), \pi(g)\rangle_\mu = \varphi_\mu(g^* * f) = \int (g^* * f)(s)\, d\mu(s), \tag{$\dagger_6$}$$

where $\pi \colon \mathcal{K}_{\mathbb{C}}(G) \to \mathcal{K}_{\mathbb{C}}(G)/\mathfrak{n}$ is the quotient map. Since

$$\begin{aligned}\int (g^* * f)(s)\, d\mu(s) &= \int\!\!\int \overline{g(t^{-1})} f(t^{-1}s)\, d\lambda(t)\, d\mu(s)\\ &= \int\!\!\int \overline{g(t)} f(ts)\, d\lambda(t)\, d\mu(s),\end{aligned}$$

we have

$$\langle \pi(f), \pi(g)\rangle_\mu = \varphi_\mu(g^* * f) = \int\!\!\int \overline{g(t)} f(ts)\, d\lambda(t)\, d\mu(s). \tag{$\dagger_7$}$$

The hermitian space $H_0 = \mathcal{K}_{\mathbb{C}}(G)/\mathfrak{n}$ is separable, and we let H_μ be the Hilbert space which is the completion of H_0. By Theorem 12.27, the measure of positive type μ defines a unitary representation $U_\mu \colon G \to \mathbf{U}(H_\mu)$, where $U_\mu(s) \in \mathbf{U}(H_\mu)$ is the extension of the map $U_\mu(s) \in \mathbf{U}(H_0)$ defined by

$$U_\mu(s)(\pi(f)) = \pi(\delta_s * f), \qquad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_{\mathbb{C}}(G).$$

By Theorem 12.14, the unitary representation $U_\mu \colon G \to \mathbf{U}(H_\mu)$ extends to a non-degenerate algebra representation $(U_\mu)_{\mathrm{ext}} \colon \mathrm{L}^1(G) \to \mathcal{L}(H_\mu)$.

The map γ defined on $\mathcal{K}(G) \times \mathcal{K}(G)$ by

$$\gamma(g, h) = \varphi_\mu(h^* * g)$$

satisfies the Conditions (U) and (N) of Section 11.8, and thus the restriction of γ to the involutive and commutative subalgebra $\mathcal{K}(K\backslash G/K)$ is a bitrace (see Definition 11.13), which

also satisfies Condition (U). Actually, this bitrace also satisfies Condition (N). This can be shown using a regularization argument that we omit. For details, see Dieudonné [22] (Chapter XXII, Section 7). Then $\mathcal{K}(K\backslash G/K)$ and $\pi(\mathcal{K}(K\backslash G/K))$ are commutative Hilbert algebras.

Let $\mathcal{H}_\mu$ be the closure of $\pi(\mathcal{K}(K\backslash G/K))$ in H_μ. Then the map $f \mapsto (U_\mu)_{\mathrm{ext}}(f)|\mathcal{H}_\mu$ is a representation of $\mathcal{K}(K\backslash G/K)$ in the separable Hilbert space $\mathcal{H}_\mu$ that we denote V_μ. Thus we have

$$V_\mu(f)(\pi(g)) = \pi(f * g) \qquad f, g \in \mathcal{K}(K\backslash G/K),$$

and by Proposition 12.13, we have

$$\|V_\mu(f)\| \leq \|f\|_1 .$$

This means that V_μ is a continuous algebra homomorphism from $\mathcal{K}(K\backslash G/K)$ (with the topology induced by the topology of $\mathrm{L}^1(K\backslash G/K)$) to the algebra $\mathcal{L}(\mathcal{H}_\mu)$ of continuous linear operators of $\mathcal{H}_\mu$. Since $\mathrm{L}^1(G)$ is a separable Banach algebra, we deduce that the closure $\mathcal{A}_\mu$ of $V_\mu(K\backslash G/K)$ is a C^* commutative separable subalgebra of $\mathcal{L}(\mathcal{H}_\mu)$.

Thus we see that the bitrace obtained by restricting the bitrace γ to $\mathcal{K}(K\backslash G/K)$ satisfies all the hypotheses of the Plancherel–Godement theorem (Theorem 11.41). To be more specific, in terms of the notations of Theorem 11.41, we have $A = \mathcal{K}(K\backslash G/K)$, $g = \gamma$, $U_g = V_\mu$, $H_g = \mathcal{H}_\mu$, and $\mathcal{A}_g = \mathcal{A}_\mu$. The Plancherel–Godement theorem yields the following result.

Theorem 17.21. *(Plancherel Transform Theorem) Let (G, K) be a Gelfand pair. For every measure of positive type μ on G, there is a unique (positive) Radon measure $\mu^\triangle$ defined on the locally compact space $\mathbf{Z}(G/K)$ of spherical functions of positive type, such that for every function $f \in \mathcal{K}(K\backslash G/K)$, the Fourier cotransform $\overline{\mathcal{F}}f$ belongs to $\mathcal{L}^2_{\mu^\triangle}(\mathbf{Z}(G/K);\mathbb{C})$, and for any two functions $f, g \in \mathcal{K}(K\backslash G/K)$, we have*

$$\int_G (g^* * f) d\mu = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}}f)(\omega)\overline{(\overline{\mathcal{F}}g)(\omega)}\, d\mu^\triangle(\omega).$$

The map $\Phi\colon f \mapsto [\overline{\mathcal{F}}f]$ from $\mathcal{K}(K\backslash G/K)$ to $\mathrm{L}^2_{\mu^\triangle}(\mathbf{Z}(G/K);\mathbb{C})$ factors as

$$\Phi = T_0 \circ \pi,$$

with $T_0\colon \pi(\mathcal{K}(K\backslash G/K)) \to \mathrm{L}^2_{\mu^\triangle}(\mathbf{Z}(G/K);\mathbb{C}) \cap \mathcal{C}_0(\mathbf{Z}(G/K);\mathbb{C})$, and T_0 extends to an isomorphism T between the Hilbert space $\mathcal{H}_\mu$ and the Hilbert space $\mathrm{L}^2_{\mu^\triangle}(\mathbf{Z}(G/K);\mathbb{C})$, as illustrated below:

$$\begin{tikzcd}
\mathcal{K}(K\backslash G/K) \arrow[r, "\pi"] \arrow[dr, "\Phi"'] & \pi(\mathcal{K}(K\backslash G/K)) \arrow[r] \arrow[d, "T_0"] & \mathcal{H}_\mu \arrow[d, "T"] \\
 & \mathrm{L}^2_{\mu^\triangle}(\mathbf{Z}(G/K);\mathbb{C}) \cap \mathcal{C}_0(\mathbf{Z}(G/K);\mathbb{C}) \arrow[r] & \mathrm{L}^2_{\mu^\triangle}(\mathbf{Z}(G/K);\mathbb{C}).
\end{tikzcd}$$

The only points which need clarification are the facts that the space S_g of Theorem 17.21 is homeomorphic to $\mathbf{Z}(G/K)$ and that the map $f \mapsto \widehat{f}$, with $f \in A = \mathcal{K}(K\backslash G/K)$, is simply the Fourier cotransform $\overline{\mathcal{F}}(f)$. The details require some knowledge of the proof of the Plancherel–Godement theorem and are omitted. The reader is referred to Dieudonné [22] (Chapter XXII, Section 7, Theorem 22.7.4).

Definition 17.22. Let (G, K) be a Gelfand pair. For every complex measure μ of positive type on G, the (positive) Radon measure $\mu^{\triangle}$ on $\mathbf{Z}(G/K)$ given by Theorem 17.21 is called the *Plancherel transform* of μ.

It is also useful to define the projection of a complex measure $\mu \in \mathbb{C}\mathcal{M}^1(G)$ onto the subspace $\mathbb{C}\mathcal{M}^1(K\backslash G/K)$ of complex measures invariant by left and right translations by elements of K.

First, assume that μ is a positive finite measure. We define the linear functional $\Phi^{\sharp}_{\mu}$ by

$$\Phi^{\sharp}_{\mu}(f) = \int_G f^{\sharp}\, d\mu, \qquad f \in \mathcal{K}(G).$$

Since μ is a positive measure, the functional $\Phi^{\sharp}_{\mu}$ is positive, so by Radon–Riesz I (Theorem 7.8), there is a unique σ-Radon measure $\mu^{\sharp}$ such that

$$\int f\, d\mu^{\sharp} = \int f^{\sharp}\, d\mu, \qquad f \in \mathcal{K}(G).$$

Going back to an arbitrary complex measure μ and expressing it as $\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$, where the four measures on the right-hand side are positive, we obtain a complex measure $\mu^{\sharp}$ such that

$$\int f\, d\mu^{\sharp} = \int f^{\sharp}\, d\mu, \qquad f \in \mathcal{K}(G). \tag{$*_\sharp$}$$

Definition 17.23. Given any complex measure $\mu \in \mathbb{C}\mathcal{M}^1(G)$, the complex measure $\mu^{\sharp} \in \mathbb{C}\mathcal{M}^1(K\backslash G/K)$ defined by $(*_\sharp)$ is called the *projection* of μ.

Consequently, we see that

$$\lambda_t \mu^{\sharp} = \rho_t \mu^{\sharp} = \mu^{\sharp}, \qquad \text{for all } t \in K.$$

Conversely, the above equations imply that $\mu = \mu^{\sharp}$. *Thus the map $\mu \mapsto \mu^{\sharp}$ is a projection of $\mathbb{C}\mathcal{M}^1(G)$ onto the subspace $\mathbb{C}\mathcal{M}^1(K\backslash G/K)$.*

The following result is not hard to prove.

Proposition 17.22. *If μ is a measure of positive type on G, then for every $f \in \mathcal{K}(G)$, we have*

$$\int_G (f^* * f)^{\sharp}\, d\mu \geq 0.$$

Proposition 17.22 is proven in Dieudonné [22] (Chapter XXII, Section 7, Lemma 22.7.4.3). Using Proposition 17.22, we see that if μ is of positive type, then so is $\mu^\sharp$. Also, $\mu^\sharp = 0$ means that μ vanishes on the subspace $\mathcal{K}(K\backslash G/K)$ of $\mathcal{K}(G)$. Thus by the uniqueness clause in Theorem 17.21, we have the following result.

Proposition 17.23. *If μ is a measure of positive type on G, then $(\mu^\sharp)^\triangle = \mu^\triangle$. For any two measures of positive type μ and ν, we have $\mu^\triangle = \nu^\triangle$ iff $\mu^\sharp = \nu^\sharp$.*

Proposition 17.24. *For all $\omega \in \mathbf{Z}(G/K)$ and for every $f \in \mathcal{K}(K\backslash G/K)$, we have*

$$(\overline{\mathcal{F}}f)(\omega) = \overline{(\mathcal{F}\overline{f})(\omega)}.$$

Proof. By Theorem 12.19(3), if p is a function of positive type, then $\overline{\check{p}} = p$, so $\overline{\check{\omega}} = \omega$ for all $\omega \in \mathbf{Z}(G/K)$, and for every $f \in \mathcal{K}(K\backslash G/K)$, we have

$$\begin{aligned}(\overline{\mathcal{F}}f)(\omega) &= \int f(x)\omega(x)\, d\lambda_G(x) = \int f(x)\overline{\omega(x^{-1})}\, d\lambda_G(x)\\ &= \overline{\int \overline{f(x)}\omega(x^{-1})\, d\lambda_G(x)} = \overline{(\mathcal{F}\overline{f})(\omega)}.\end{aligned}$$

Therefore,

$$(\overline{\mathcal{F}}f)(\omega) = \overline{(\mathcal{F}\overline{f})(\omega)},$$

as claimed. ∎

In general, given a function $f \in \mathcal{K}(K\backslash G/K)$, by Theorem 17.21, $\overline{\mathcal{F}}f \in \mathrm{L}^2_{\mu^\triangle}(\mathbf{Z}(G/K))$, but $\overline{\mathcal{F}}f \notin \mathrm{L}^1_{\mu^\triangle}(\mathbf{Z}(G/K))$. However, if g is another function in $\mathcal{K}(K\backslash G/K)$, then Theorem 17.21 also shows that $\overline{\mathcal{F}}(f * g) = (\overline{\mathcal{F}}f)(\overline{\mathcal{F}}g) \in \mathrm{L}^1_{\mu^\triangle}(\mathbf{Z}(G/K))$.

Proposition 17.25. *For all $f, g \in \mathcal{K}(K\backslash G/K)$, for any measure μ of positive type, we have*

$$\int_G (f * g)\, d\mu = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}}f)(\omega)(\overline{\mathcal{F}}g)(\omega)\, d\mu^\triangle(\omega).$$

Consequently, $\overline{\mathcal{F}}(f * g) = (\overline{\mathcal{F}}f)(\overline{\mathcal{F}}g) \in \mathrm{L}^1_{\mu^\triangle}(\mathbf{Z}(G/K))$.

Proof. Since $\overline{\mathcal{F}}g = \overline{\mathcal{F}\overline{g}}$ and $\mathcal{F}g = \overline{\mathcal{F}}\check{g}$, we have $\overline{\overline{\mathcal{F}}\,\overline{g}} = \mathcal{F}g$ and $\mathcal{F}\check{g} = \overline{\mathcal{F}}g$, and so

$$\overline{\overline{\mathcal{F}}g^*} = \overline{\overline{\mathcal{F}}\,\overline{\check{g}}} = \mathcal{F}\check{g} = \overline{\mathcal{F}}g.$$

If we recall that $\mathcal{K}(K\backslash G/K)$ is a commutative algebra (under convolution), from Theorem 17.21 with g replaced by g^*, we deduce that

$$\int_G (f * g)\, d\mu = \int_G (g * f)\, d\mu = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}}f)(\omega)(\overline{\mathcal{F}}g)(\omega)\, d\mu^\triangle(\omega),$$

as claimed. ∎

Example 17.8. One of the main examples of Plancherel transform is the Dirac measure $\mu = \delta_e$. As a corollary of Proposition 12.25, the Dirac measure δ_e is of positive type.

Definition 17.24. The Plancherel transform $\delta_e^{\triangle}$ of the Dirac measure δ_e is called the *canonical measure* on $\mathbf{Z}(G/K)$ and is denoted $m_{\mathbf{Z}}$.

Because G is unimodular, we have

$$\int_G (g^* * f)\, d\delta_e = (g^* * f)(e) = \int_G f(s^{-1})\overline{g(s^{-1})}\, d\lambda_G(s) = \int_G f(s)\overline{g(s)}\, d\lambda_G(s),$$

and for any two functions $f, g \in \mathcal{K}(K\backslash G/K)$, by Theorem 17.21, we have

$$\int_G f(s)\overline{g(s)}\, d\lambda_G(s) = \int_G g^* * f\, d\delta_e = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f)(\omega)\overline{(\overline{\mathcal{F}} g)(\omega)}\, dm_{\mathbf{Z}}(\omega).$$

Write $f = f_1 + if_2$ and $g = g_1 + ig_2$, where f_1, f_2, g_1, g_2 are all real-valued. Then we have

$$\begin{aligned}\int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f)(\omega)\overline{(\overline{\mathcal{F}} g)(\omega)}\, dm_{\mathbf{Z}}(\omega) &= \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f_1)(\omega)\overline{(\overline{\mathcal{F}} g_1)(\omega)}\, dm_{\mathbf{Z}}(\omega)\\ &\quad + \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f_2)(\omega)\overline{(\overline{\mathcal{F}} g_2)(\omega)}\, dm_{\mathbf{Z}}(\omega)\\ &\quad - i\int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f_1)(\omega)\overline{(\overline{\mathcal{F}} g_2)(\omega)}\, dm_{\mathbf{Z}}(\omega)\\ &\quad + i\int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} g_1)(\omega)\overline{(\overline{\mathcal{F}} f_2)(\omega)}\, dm_{\mathbf{Z}}(\omega).\end{aligned}$$

Since $(\overline{\mathcal{F}} f)(\omega) = \overline{(\mathcal{F}\overline{f})(\omega)}$, and since f_1, f_2, g_1, g_2 are real-valued, we have

$$\begin{aligned}\int_G f_1 g_1\, d\lambda_G &= \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f_1)(\omega)\overline{(\overline{\mathcal{F}} g_1)(\omega)}\, dm_{\mathbf{Z}}(\omega) = \int_{\mathbf{Z}(G/K)} (\mathcal{F} f_1)(\omega)\overline{(\mathcal{F} g_1)(\omega)}\, dm_{\mathbf{Z}}(\omega)\\ \int_G f_2 g_2\, d\lambda_G &= \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f_2)(\omega)\overline{(\overline{\mathcal{F}} g_2)(\omega)}\, dm_{\mathbf{Z}}(\omega) = \int_{\mathbf{Z}(G/K)} (\mathcal{F} f_2)(\omega)\overline{(\mathcal{F} g_2)(\omega)}\, dm_{\mathbf{Z}}(\omega)\\ \int_G f_1 g_2\, d\lambda_G &= \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f_1)(\omega)\overline{(\overline{\mathcal{F}} g_2)(\omega)}\, dm_{\mathbf{Z}}(\omega) = \int_{\mathbf{Z}(G/K)} (\mathcal{F} f_1)(\omega)\overline{(\mathcal{F} g_2)(\omega)}\, dm_{\mathbf{Z}}(\omega)\\ \int_G g_1 f_2\, d\lambda_G &= \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} g_1)(\omega)\overline{(\overline{\mathcal{F}} f_2)(\omega)}\, dm_{\mathbf{Z}}(\omega) = \int_{\mathbf{Z}(G/K)} (\mathcal{F} g_1)(\omega)\overline{(\mathcal{F} f_2)(\omega)}\, dm_{\mathbf{Z}}(\omega),\end{aligned}$$

because the left integrals are real, and thus

$$\begin{aligned}\int_G f(s)\overline{g(s)}\, d\lambda_G(s) &= \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f)(\omega)\overline{(\overline{\mathcal{F}} g)(\omega)}\, dm_{\mathbf{Z}}(\omega) && (*_1)\\ &= \int_{\mathbf{Z}(G/K)} (\mathcal{F} f)(\omega)\overline{(\mathcal{F} g)(\omega)}\, dm_{\mathbf{Z}}(\omega). && (*_2)\end{aligned}$$

Observe that the above integrals are inner products. As a consequence, we have the following result.

Proposition 17.26. *For any two functions $f, g \in \mathcal{K}(K\backslash G/K)$, we have*

$$\begin{aligned}\int_G f(s)\overline{g(s)}\, d\lambda_G(s) &= \int_{\mathbf{Z}(G/K)} (\mathcal{F}f)(\omega)\overline{(\mathcal{F}g)(\omega)}\, dm_{\mathbf{Z}}(\omega) \\ &= \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}}f)(\omega)\overline{(\overline{\mathcal{F}}g)(\omega)}\, dm_{\mathbf{Z}}(\omega).\end{aligned}$$

The linear maps $f \mapsto \mathcal{F}f$ and $f \mapsto \overline{\mathcal{F}}f$, with $f, g \in \mathcal{K}(K\backslash G/K)$, are isometries, and by Theorem 17.21, these maps extend to isomorphisms from the Hilbert space $\mathrm{L}^2(K\backslash G/K)$ onto the Hilbert space $\mathrm{L}^2_{m_{\mathbf{Z}}}(\mathbf{Z}(G/K))$.

We can further extend these maps to linear maps of $\mathrm{L}^2(G)$ onto $\mathrm{L}^2_{m_{\mathbf{Z}}}(\mathbf{Z}(G/K))$, by setting $\mathcal{F}([f]) = \mathcal{F}([f^\sharp])$ and $\overline{\mathcal{F}}([f]) = \overline{\mathcal{F}}([f^\sharp])$. The equation $\overline{\mathcal{F}}([f]) = \overline{\mathcal{F}([\overline{f}])}$ holds. By abuse of notations, we write $\mathcal{F}f$ (resp. $\overline{\mathcal{F}}f$) for any function in the class $\mathcal{F}([f])$ (resp, $\overline{\mathcal{F}}([f])$). With these notation, $(*_1)$ and $(*_2)$ hold for $f, g \in \mathcal{L}^2(K\backslash G/K)$.

Proposition 17.26 is a generalization of the Plancherel theorem (Theorem 10.27), as we will see in Example 17.9.

Example 17.9. Another important example is the case where G is commutative and $K = \{e\}$. In this case, the functional equation (s_1) characterizing spherical functions reduces to

$$\omega(xy) = \omega(x)\omega(y),$$

so $\omega\colon G \to \mathbb{C}^*$ is a continuous homomorphism such that $\omega(e) = 1$ and $|\omega(x)| \leq 1$ for all $x \in G$. Since $\omega(x^{-1}) = \omega(x)^{-1}$, we conclude that $|\omega(x)| = 1$ for all $x \in G$, which means that ω is a *group character*. The function ω is of positive type. This is because $\omega(y^{-1}x) = \overline{\omega(y)}\omega(x)$, so for every function $f \in \mathcal{K}(G)$, have

$$\begin{aligned}\int_G (f^* * f)\omega\, d\lambda_G &= \int_G\int_G f^*(y)f(y^{-1}x)\omega(x)\, d\lambda_G(y)\, d\lambda_G(x) \\ &= \int_G\int_G \overline{f(y^{-1})}f(y^{-1}x)\omega(x)\, d\lambda_G(y)\, d\lambda_G(x) \\ &= \int_G\int_G \overline{f(y)}f(x)\omega(y^{-1}x)\, d\lambda_G(y)\, d\lambda_G(x) \\ &= \int_G\int_G \overline{f(y)\omega(y)}f(x)\omega(x)\, d\lambda_G(y)\, d\lambda_G(x) \\ &= \left|\int_G f(x)\omega(x)\, d\lambda_G(x)\right|^2,\end{aligned}$$

so

$$\int_G (f^* * f)\omega\, d\lambda_G = \left|\int_G f(x)\omega(x)\, d\lambda_G(x)\right|^2 \geq 0.$$

Consequently, $\mathbf{S}(G/\{e\}) = \mathbf{Z}(G/\{e\})$ is the space of group characters of G, and this space is homeomorphic to $\mathsf{X}_0(A)$, where $A = \mathrm{L}^1(G) \oplus \mathbb{C}\delta_e$.

The topological space $\widehat{G} = \mathbf{S}(G/\{e\})$ of group characters is a group, and it can be shown that the topology of $\mathbf{S}(G/\{e\})$ is compatible with the group structure; see Dieudonné [22] (Chapter XXII, Section 10, Lemma 22.10.2). Therefore, $\widehat{G}$ is a commutative topological group which is locally compact, metrizable and separable. Given a (left) Haar measure λ_G on G, it can be shown that the Plancherel transform $\lambda_{\widehat{G}} = m_{\mathbf{Z}} = \delta_e^{\triangle}$ is a (left) Haar measure on $\widehat{G}$; see Dieudonné [22] (Chapter XXII, Section 10, Lemma 22.10.5). The Haar measure $\lambda_{\widehat{G}}$ on $\widehat{G}$ and the Haar measure λ_G on G are said to be *associated*. If λ_G is replaced by $a\lambda_G$ with $a > 0$, then $\lambda_{\widehat{G}}$ is replaced by $a^{-1}\lambda_{\widehat{G}}$.

Proposition 17.26 shows that the equation

$$\int_G f(s)\overline{g(s)}\, d\lambda_G(s) = \int_{\widehat{G}} (\mathcal{F}f)(\omega)\overline{(\mathcal{F}g)(\omega)}\, d\lambda_{\widehat{G}}(\omega)$$

holds, and that $\mathcal{F}$ has an extension which is an isometry from $\mathrm{L}^2(G)$ to $\mathrm{L}^2(\widehat{G})$, providing another proof of the Plancherel theorem, Theorem 10.27.

We now retun to an arbitary locally compact group G (metrizable, separable, and unimodular). For any function $\omega \in \mathbf{Z}(G/K)$ of positive type, the measure $\omega\, d\lambda_G$ is a measure of positive type. It can be shown that

$$(\omega\, d\lambda_G)^{\triangle} = \delta_\omega,$$

the Dirac measure at ω; Dieudonné [22] (Chapter XXII, Section 7, Lemma 22.7.6.1). The result is a sort of Fourier inversion formula.

Proposition 17.27. *If μ is a measure of positive type on G, then for every function $f \in \mathcal{K}(K\backslash G/K)$, if the Fourier cotransform $\overline{\mathcal{F}}f$ belongs to $\mathrm{L}^1_{\mu^\triangle}(\mathbf{Z}(G/K))$, then*

$$\int_G f\, d\mu = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}}f)(\omega)\, d\mu^{\triangle}(\omega).$$

Proposition 17.27 is proven in Dieudonné [22] (Chapter XXII, Section 7, Lemma 22.7.8).

In general, given a function $f \in \mathcal{K}(K\backslash G/K)$, by Theorem 17.21, $\overline{\mathcal{F}}f \in \mathrm{L}^2_{\mu^\triangle}(\mathbf{Z}(G/K))$, but $\overline{\mathcal{F}}f \notin \mathrm{L}^1_{\mu^\triangle}(\mathbf{Z}(G/K))$. Using Proposition 17.25, the proof of Proposition 17.27 can be adapted to use the technique of regularization. If (g_n) is a sequence of positive functions in $\mathcal{K}(K\backslash G/K)$ having a compact support that tends to $\{e\}$, and such that $\int g_n\, d\lambda_G = 1$, then $\int (f * g_n) d\mu$ tends to $\int f d\mu$, and we have

$$\int_G f d\mu = \lim_{n\mapsto\infty} \int_G (f * g_n) d\mu = \lim_{n\mapsto\infty} \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}}f)(\omega)(\overline{\mathcal{F}}g_n)(\omega) d\mu^{\triangle}(\omega).$$

In particular, if we apply the above formula to $\mu = \delta_e$, if we compute $\int_G (\lambda_s f)\, d\delta_e = f(s^{-1})$, using Proposition 17.20, we find that for every $f \in \mathcal{K}(K\backslash G/K)$, we have

$$f(s) = \lim_{n \mapsto \infty} \int_{\mathbf{Z}(G/K)} \omega(s^{-1})(\overline{\mathcal{F}} f)(\omega)(\overline{\mathcal{F}} g_n)(\omega)\, dm_{\mathbf{Z}}(\omega).$$

The above process for the inversion of the Fourier cotransform is usually used when G is abelian and $K = \{e\}$.

We now take a closer look at the space $\mathcal{P}_+(K\backslash G/K)$ of functions in $\mathcal{C}(K\backslash G/K)$ which are of positive type.

Proposition 17.28. *The map $p \mapsto (p\,\lambda_G)^{\triangle}$ is a bijection between the space $\mathcal{P}_+(K\backslash G/K)$ of functions in $\mathcal{C}(K\backslash G/K)$ which are of positive type onto the space $\mathcal{M}^1_+(\mathbf{Z}(G/K))$ of bounded positive measures on $\mathbf{Z}(G/K)$. The inverse $\overline{\mathcal{F}}'$ of the above map ($p \mapsto (p\,\lambda_G)^{\triangle}$) is given by*

$$(\overline{\mathcal{F}}' \mu)(x) = \int_{\mathbf{Z}(G/K)} \omega(x)\, d\mu(\omega), \qquad \mu \in \mathcal{M}^1_+(\mathbf{Z}(G/K)).$$

For every $f \in \mathcal{L}^1(G)$, we have

$$\int_G f(x)(\overline{\mathcal{F}}' \mu)(x)\, d\lambda_G(x) = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f)(\omega)\, d\mu(\omega).$$

Proposition 17.28 is proven in Dieudonné [22] (Chapter XXII, Section 7, Lemma 22.7.10). The proof uses the Bochner–Godement theorem (Theorem 11.42).

In the special case where G is commutative and $K = \{e\}$, the space $\mathcal{P}_+(G)$ is the set of all continuous functions of positive type on G, and Proposition 17.28 implies that every function p of positive type can be written uniquely as

$$p(x) = \int_{\widehat{G}} \omega(x)\, d\mu(\omega)$$

for some positive measure μ on $\widehat{G}$, a result known as *Bochner's theorem*; see also Folland [33] Chapter 4, Theorem 4.18.

We conclude with three remarks.

1. If the Haar measure λ_G on G is replaced by $a\lambda_G$ with $a > 0$, then the space $\mathbf{S}(G/K)$ of spherical functions is unchanged. The Fourier transform and the Fourier cotransform are multiplied by a, and the Plancherel transform (see Definition 17.22) is multiplied by a^{-1}.

2. If (G, K) is a Gelfand pair and if G is compact, then by Example 17.5, Proposition 15.17, and Proposition 15.21, *all spherical functions are of positive type*. However, if G is not compact, this is generally false. For instance, the functions P_ρ of Example 17.6 do not satisfy the property $\overline{\check{\omega}} = \omega$ unless $\Re\rho = -1/2$. It can be shown that these functions are of positive type if $\Re\rho = -1/2$.

3. If G is compact, then we saw that the space $\mathbf{S}(G/K)$ is discrete and in bijection with the subset of $R(G)$ (of irreducible representations of G) consisting of those $\rho \in R(G)$ such that $(\rho : \sigma_0) = 1$. We can view the Fourier transform $\mathcal{F}f$ of a function $f \in \mathcal{L}^1(G)$ as the family

$$\left(c_\rho = \frac{1}{n_\rho}\langle f, m_{11}^{(\rho)}\rangle \right)_{\rho\in\mathbf{S}(G/K)}.$$

The Fourier transform and the Plancherel measures are discussed from a different point of view for symmetric spaces in Helgason [46] (Chapter 4) and Helgason [45] (Chapter 3, especially Section 12).

17.9 Extension of the Plancherel Transform; $\mathbf{P}(G)$ and $\mathbf{P}'(\mathbf{Z})$ ⊛

The purpose of this section is to extend the Plancherel transform to a bigger set of measures and to define the notion of Fourier transform of a measure. Let X be a locally compact space. Recall that the space of σ-Radon measures on X is denoted by $\mathcal{M}_\sigma^+(X)$ and the space of Radon measures on X is denoted $\mathcal{M}^+(X)$. The space of complex measures on X is denoted by $\mathbb{C}\mathcal{M}^1(X)$ and is called the space of *bounded measures*. The space of regular complex Borel measures is denoted by $\mathcal{M}^1(X)$ (which is an abbreviation for $\mathcal{M}^1_{\mathrm{reg},\mathbb{C}}(X)$). The space $\mathbb{C}\mathcal{M}^1(X)$ contains the space of positive bounded Borel measures, and of course, $\mathcal{M}^1(X) \subseteq \mathbb{C}\mathcal{M}^1(X)$.

Let G be a locally compact, metrizable, separable, and unimodular group.

Definition 17.25. The complex vector space spanned by the union of the complex measures and the σ-Radon measures is denoted by $\mathcal{M}_{\mathbb{C}}(G)$. Let $\mathbf{P}_+(G)$ be the set of measures of positive type, and let $\mathbf{P}(G)$ be the complex subspace of $\mathcal{M}_{\mathbb{C}}(G)$ spanned by $\mathbf{P}_+(G)$, which consists of all combinations $\mu_1 - \mu_2 + i\mu_3 - i\mu_4$, where the μ_i belong to $\mathbf{P}_+(G)$.

As a general rule, the subscript $+$ indicates that we are dealing with functions or measures of positive type, and the suppression of the subscript $+$ that we are considering the vector space spanned by that set.

It is easy to check that if $\mu \in \mathbf{P}(G)$, then $\mu^\sharp \in \mathbf{P}(G)$. The image of $\mathbf{P}(G)$ by the map $\mu \mapsto \mu^\sharp$ is $\mathbf{P}(G) \cap \mathcal{M}_{\mathbb{C}}(K\backslash G/K)$.

Let (G, K) be a Gelfand pair. The Plancherel transform $\mu \mapsto \mu^\triangle$ is a map from $\mathbf{P}_+(G)$ to $\mathcal{M}^+(\mathbf{Z}(G/K))$, the space of positive measures on $\mathbf{Z}(G/K)$. We have $(\mu+\nu)^\triangle = \mu^\triangle + \nu^\triangle$ and $(c\mu)^\triangle = c\mu^\triangle$, for any $c > 0$. From this, it is easy to show that for any combination $\mu = \mu_1 - \mu_2 + i\mu_3 - i\mu_4$ of measures $\mu_i \in \mathbf{P}_+(G)$, the sum $\mu_1^\triangle - \mu_2^\triangle + i\mu_3^\triangle - i\mu_4^\triangle$ is a measure on $\mathbf{Z}(G/K)$ that depends only on μ and not on its decomposition.

Definition 17.26. The $\mathbb{C}$-linear map $\mu \mapsto \mu^{\triangle}$ from $\mathbf{P}(G)$ to $\mathcal{M}_{\mathbb{C}}(\mathbf{Z}(G/K))$ is also called the *Plancherel transform*.

It is clear that Proposition 17.25 also applies to measures in $\mathbf{P}(G)$; that is, for any $\mu \in \mathbf{P}(G)$, we have

$$\int_G (f * g)\, d\mu = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f)(\omega)(\overline{\mathcal{F}} g)(\omega)\, d\mu^{\triangle}(\omega), \quad \text{for all } f, g \in \mathcal{K}(K\backslash G/K).$$

By regularization, it can be shown that for any two measures $\mu, \nu \in \mathbf{P}(G) \cap \mathcal{M}_{\mathbb{C}}(K\backslash G/K)$, if $\mu^{\triangle} = \nu^{\triangle}$, then $\mu = \nu$. But $\mu = \mu^{\sharp}$ and $\nu = \nu^{\sharp}$, so it follows that the kernel of the Plancherel transform is the subspace of measures $\mu \in \mathbf{P}(G)$ such that $\mu^{\sharp} = 0$.

Definition 17.27. The image of $\mathbf{P}(G)$ (or $\mathbf{P}(G) \cap \mathcal{M}_{\mathbb{C}}(K\backslash G/K)$) by the Plancherel transform is denoted by $\mathbf{P}'(\mathbf{Z})$. Let $\mathcal{P}(G)$ be the complex vector space spanned by the functions of positive type on G.

The space $\mathcal{P}(G)$ is a subspace of both $\mathrm{L}^{\infty}(G)$ and $\mathbf{P}(G)$, by viewing $f \in \mathcal{P}(G)$ as $f\, d\lambda_G$. By Proposition 17.28, the image of $\mathcal{P}(G)$ by the Plancherel transform is $\mathbb{C}\mathcal{M}^1(\mathbf{Z}(G/K))$, the space of bounded measures on $\mathbf{Z}(G/K)$. Consequently we obtain the following result.

Proposition 17.29. *Let $\mathcal{P}(K\backslash G/K)$ be the subspace of $\mathcal{P}(G)$ consisting of the functions invariant by left and right translations by elements of K. Then the map $f \mapsto (f\lambda_G)^{\triangle}$ is a linear bijection between $\mathcal{P}(K\backslash G/K)$ and $\mathbb{C}\mathcal{M}^1(\mathbf{Z}(G/K))$. The inverse map is denoted by $\overline{\mathcal{F}}'$ and is given by the formula*

$$(\overline{\mathcal{F}}' \mu')(x) = \int_{\mathbf{Z}(G/K)} \omega(x)\, d\mu'(\omega), \qquad \mu' \in \mathbb{C}\mathcal{M}^1(\mathbf{Z}(G/K)).$$

We also have

$$\left\| \overline{\mathcal{F}}' \mu' \right\| \leq \|\mu'\|,$$

since $|\omega(x)| \leq 1$ for all $x \in G$ and all $\omega \in \mathbf{Z}(G/K)$.

Thus the linear map $\overline{\mathcal{F}}'$ from $\mathbb{C}\mathcal{M}^1(\mathbf{Z}(G/K))$ to $\mathcal{P}(K\backslash G/K)$ is continuous for the topology of the Banach space $\mathcal{C}_b(G)$. However, in general, $\mathcal{P}(K\backslash G/K)$ is *not* closed in $\mathcal{C}_b(G)$.

Unfortunately, no convenient characterizations of the spaces $\mathbf{P}(G)$ and $\mathbf{P}'(\mathbf{Z})$ are known, besides their definition. There are necessary *or* sufficient conditions, but *no necessary and sufficient conditions* known. For example, a necessary condition for a measure μ' on $\mathbf{Z}(G/K)$ to belong to $\mathbf{P}'(\mathbf{Z})$ is that the Fourier cotransforms $\overline{\mathcal{F}} f$ of functions $f \in \mathcal{K}(K\backslash G/K)$ belong to $\mathrm{L}^2_{|\mu'|}(\mathbf{Z}(G/K))$, but there are counter-examples showing that this condition is not sufficient. Also, we showed that $\mathbb{C}\mathcal{M}^1(\mathbf{Z}(G/K)) \subseteq \mathbf{P}'(\mathbf{Z})$, but there are unbounded measures in $\mathbf{P}'(\mathbf{Z})$ (for example, the Haar measure on $\widehat{G}$, when G is a commutative noncompact locally compact group).

The following result holds.

Proposition 17.30. *For every measure $\mu' \in \mathbf{P}'(\mathbf{Z})$, and every function $g' \in \mathcal{L}^2_{|\mu'|}(\mathbf{Z}(G/K))$, we have $g'\, d\mu' \in \mathbf{P}'(\mathbf{Z})$. If $\nu \in \mathbf{P}(G)$ is a measure such that $\nu^{\triangle} = g'\, d\mu'$, then for every $f \in \mathcal{K}(K\backslash G/K)$ we have*

$$\int_G f\, d\nu = \int_{\mathbf{Z}(G/K)} (\overline{\mathcal{F}} f)(\omega) g'(\omega)\, d\mu'(\omega).$$

Proposition 17.30 is proven in Dieudonné [22] (Chapter XXII, Section 8, Lemma 22.8.3).

In particular, if we take μ' to be the canonical measure $m_{\mathbf{Z}}$, then we obtain the following corollary.

Proposition 17.31. *The space $\mathrm{L}^2_{m_{\mathbf{Z}}}(\mathbf{Z}(G/K))$ (viewed as a subspace of $\mathcal{M}_{\mathbb{C}}(\mathbf{Z}(G/K))$) using the embedding $\omega \mapsto \omega dm_{\mathbf{Z}}$ is contained in $\mathbf{P}'(\mathbf{Z})$. The restriction to $\mathrm{L}^2(K\backslash G/K) \subseteq \mathbf{P}(G)$ of the Plancherel transform is identical to the extension of the Fourier transform $\mathcal{F}$ from $\mathrm{L}^2(K\backslash G/K)$ to $\mathrm{L}^2_{m_{\mathbf{Z}}}(\mathbf{Z}(G/K))$ given by Proposition 17.26. In particular, for any $f \in \mathrm{L}^2(K\backslash G/K)$, we have $(f\, d\lambda_G)^{\triangle} = (\mathcal{F} f)\, dm_{\mathbf{Z}}$.*

Proposition 17.31 is proven in Dieudonné [22] (Chapter XXII, Section 8, Lemma 22.8.4).

Remark: It is possible that there is some positive measure $\mu' \in \mathbf{P}'(\mathbf{Z})$, yet there are positive measures ν' with $0 \leq \nu' \leq \mu'$, and $\nu' \notin \mathbf{P}'(\mathbf{Z})$.

Here are more results about the space $\mathbb{C}\mathcal{M}^1(G)$ of bounded measures on G.

Proposition 17.32. *The space $\mathbb{C}\mathcal{M}^1(G)$ of bounded measures on G is contained in $\mathbf{P}(G)$. For every $\mu \in \mathbb{C}\mathcal{M}^1(G)$, we have*

$$\mu^{\triangle} = (\mathcal{F}\mu)\, dm_{\mathbf{Z}},$$

where $\mathcal{F}\mu$ is a continuous bounded function on $\mathbf{S}(G/K)$ given by

$$(\mathcal{F}\mu)(\omega) = \int_G \omega(x^{-1})\, d\mu(x).$$

Furthermore, for every function $f \in \mathcal{L}^1(K\backslash G/K)$ (resp. $f \in \mathcal{L}^2(K\backslash G/K)$), we have $\mu * f, f * \mu \in \mathcal{L}^1(G)$ (resp. $\mathcal{L}^2(G)$), and

$$\mathcal{F}(\mu * f) = \mathcal{F}(f * \mu) = (\mathcal{F} f)(\mathcal{F}\mu)$$

almost everywhere w.r.t. $m_{\mathbf{Z}}$.

Proposition 17.32 is proven in Dieudonné [22] (Chapter XXII, Section 8, Lemma 22.8.5). The following definition generalizes Definition 10.4.

Definition 17.28. If $\mu \in \mathbb{C}\mathcal{M}^1(G)$ is a bounded measure, then the function $\mathcal{F}\mu$ (defined on $\mathbf{S}(G/K)$) given by
$$(\mathcal{F}\mu)(\omega) = \int_G \omega(x^{-1})\, d\mu(x)$$
is called the *Fourier transform* of μ. We define the *Fourier cotransform* $\overline{\mathcal{F}}\mu$ of μ as $\mathcal{F}\check{\mu}$; that is,
$$(\overline{\mathcal{F}}\mu)(\omega) = \int_G \omega(x^{-1})\, d\check{\mu}(x) = \int_G \omega(x)\, d\mu(x).$$

For every function $f \in \mathcal{L}^1(G)$, we have $\mathcal{F}f = \mathcal{F}(f\, d\lambda_G)$, which justifies the terminology. We have
$$(\mathcal{F}\delta_x)(\omega) = \omega(x^{-1}), \qquad (\overline{\mathcal{F}}\delta_x)(\omega) = \omega(x).$$

It is clear that $\|\mathcal{F}\mu\| \leq \|\mu\|$, so $\mathcal{F}$ is a continuous linear map from the Banach space $\mathbb{C}\mathcal{M}^1(G)$ to the Banach space $\mathcal{C}_b(\mathbf{S}(G/K))$ of continuous bounded functions on $\mathbf{S}(G/K)$. However, in general, the bounded function $\mathcal{F}\mu$ does not tend to zero at infinity, as shown by $\mu = \delta_e$, for which $\mathcal{F}\delta_e$ is the constant 1.

As a corollary of Proposition 17.32, since $\mathbb{C}\mathcal{M}^1(G)$ is contained in $\mathbf{P}(G)$, we have
$$\mathrm{L}^1(G) \cap \mathrm{L}^2(G) = \mathbb{C}\mathcal{M}^1(G) \cap \mathrm{L}^2(G),$$
and by Proposition 17.31, the class of the Fourier transform $\mathcal{F}f$ of a function $f \in \mathcal{L}^1(G) \cap \mathcal{L}^2(G)$ is identical to the class $\mathcal{F}[f]$ as in Definition 17.20.

Proposition 17.33. *For any two functions $f, g \in \mathcal{L}^2(K\backslash G/K)$, the bounded continuous function $f * g$ belongs to $\mathcal{P}(K\backslash G/K)$ and we have*
$$((f * g)d\lambda_G)^{\triangle} = (\mathcal{F}f)(\mathcal{F}g)\, dm_{\mathbf{Z}}.$$

Proposition 17.33 is proven in Dieudonné [22] (Chapter XXII, Section 8, Lemma 22.8.8).

In general, $f * g$ is *not* integrable for the measure λ_G so its Fourier transform is not definable by the formula of Definition 17.20.

Definition 17.29. Define the spaces $\mathcal{P}^1(K\backslash G/K)$ and $\mathcal{P}^2(K\backslash G/K)$ by
$$\begin{aligned}\mathcal{P}^1(K\backslash G/K) &= \mathcal{P}(K\backslash G/K) \cap \mathrm{L}^1(K\backslash G/K)\\ \mathcal{P}^2(K\backslash G/K) &= \mathcal{P}(K\backslash G/K) \cap \mathrm{L}^2(K\backslash G/K).\end{aligned}$$

Since for any function $f \in \mathcal{L}^\infty(G) \cap \mathcal{L}^1(G)$ we have $|f(x)|^2 \leq \|f\|_\infty |f(x)|$ almost everywhere, we conclude that $f \in \mathcal{L}^2(G)$. Since the functions in $\mathcal{P}(G)$ are bounded, we have the inclusions
$$\mathcal{P}^1(K\backslash G/K) \subseteq \mathcal{P}^2(K\backslash G/K) \subseteq \mathcal{P}(K\backslash G/K).$$

Proposition 17.34. *The image of $\mathcal{P}^2(K\backslash G/K)$ (as a subspace of $\mathbb{C}\mathcal{M}^1(K\backslash G/K)$) under the Plancherel transform is the subspace $\mathrm{L}^1_{m_{\mathbf{Z}}}(\mathbf{Z}(G/K)) \cap \mathrm{L}^2_{m_{\mathbf{Z}}}(\mathbf{Z}(G/K))$ of $\mathbb{C}\mathcal{M}^1(\mathbf{Z}(G/K))$. For every function $f \in \mathcal{P}^2(K\backslash G/K)$, we have the Fourier inversion formula*

$$f = \overline{\mathcal{F}}'((\mathcal{F}f)\, dm_{\mathbf{Z}}),$$

where $\overline{\mathcal{F}}'$ is defined in Proposition 17.29. Moreover, if we also have $f \in \mathcal{P}^1(K\backslash G/K)$, then we have the Fourier inversion formula

$$f(x) = \int_{\mathbf{Z}(G/K)} \omega(x) \left(\int_G f(y)\omega(y^{-1})\, d\lambda_G(y) \right) dm_{\mathbf{Z}}(\omega).$$

Proposition 17.34 is proven in Dieudonné [22] (Chapter XXII, Section 8, Lemma 22.8.10).

One should be cautious that in general, f is not integrable, so we can't use the formula of Definition 17.20 to define $\mathcal{F}f$. If $f \in \mathcal{P}^1(K\backslash G/K)$, then we have the formula above, but the two integrals cannot be replaced by the double integral

$$\iint_{G\times\mathbf{Z}(G/K)} \omega(x)\omega(y^{-1})f(y)\, d\lambda_G(y)\, dm_{\mathbf{Z}}(\omega),$$

because this integral is not defined in general.

Remark: The spaces $\mathcal{P}(K\backslash G/K), \mathcal{P}^1(K\backslash G/K), \mathcal{P}^2(K\backslash G/K)$ are generally not closed in $\mathcal{C}_b(G)$, but $\mathcal{P}^1(K\backslash G/K)$ is dense in $\mathrm{L}^1(K\backslash G/K)$, and $\mathcal{P}^2(K\backslash G/K)$ is dense in $\mathrm{L}^2(K\backslash G/K)$.

In the special case where G is commutative and $K = \{e\}$, we know that $\mathbf{S}(G/\{e\}) = \widehat{G}$, and by Pontrjagin duality (Theorem 10.30), $\widehat{\widehat{G}}$ and G can be identifed, and then the transform $\overline{\mathcal{F}}'$ from $\mathbb{C}\mathcal{M}^1(\widehat{G})$ to $\mathcal{P}(G)$ defined in Proposition 17.29 is identified with the Fourier cotransform $\overline{\mathcal{F}}$ on $\mathcal{M}^1(\widehat{G})$. The Haar measure $\lambda_{\widehat{\widehat{G}}}$ is also identified with the Haar measure λ_G. Then Proposition 17.26 and Proposition 17.34 yields the Fourier inversion formula

$$f = \overline{\mathcal{F}}(\mathcal{F}f),$$

for every function $f \in \mathcal{P}^2(G)$, and since $\mathcal{P}^2(G)$ is dense in $\mathrm{L}^2(G)$, the above formula actually holds for all $f \in \mathrm{L}^2(G)$. This gives another proof of the inversion formula of the Pontrjagin duality theorem, Theorem 10.30.

By Proposition 17.31 and Proposition 17.32, for any $f \in \mathrm{L}^2(G)$ and any $\mu \in \mathbb{C}\mathcal{M}^1(G)$, we have

$$(f\, d\lambda_G)^{\triangle} = (\mathcal{F}f)\, d\lambda_{\widehat{G}}, \qquad \mu^{\triangle} = (\mathcal{F}\mu)\, d\lambda_{\widehat{G}}.$$

17.10 Spherical Functions of Positive Type and Irreducible Representations

Let (G, K) be a Gelfand pair (with G a locally compact, metrizable, separable, unimodular group). From Theorem 12.19, every spherical function of positive type $\omega \in \mathbf{Z}(G/K)$ induces a cyclic unitary representation $U_\omega\colon G \to \mathbf{U}(H_\omega)$ of G in a separable Hilbert space H_ω. Recall that the map φ_ω given by

$$\varphi_\omega(\mu) = \int_G \omega(s)\, d\mu, \qquad \mu \in \mathcal{M}^1(G)$$

is a positive linear form, and so is its restriction to the unital involutive subalgebra $A = \mathrm{L}^1(G) \oplus \mathbb{C}\delta_e$. If

$$\mathfrak{n} = \{\mu \in A \mid \varphi_\omega(\mu^* * \mu) = 0\},$$

then $\mathfrak{n}$ is a left ideal in A, and $H_0 = A/\mathfrak{n}$ is a hermitian space with the inner product

$$\langle \pi(\mu), \pi(\nu)\rangle = \varphi_\omega(\nu^* * \mu) = \int_G \omega(s)\, d(\nu^* * \mu)(s),$$

where $\pi\colon A \to A/\mathfrak{n}$ is the quotient map (and $\mu^* = \overline{\check{\mu}}$). If H_ω is the separable Hilbert space which is the completion of $H_0 = A/\mathfrak{n}$, then the unitary representation $U_\omega\colon G \to \mathbf{U}(H_\omega)$ is completely determined by

$$U_\omega(s)(\pi(\mu)) = \pi(\delta_s * \mu), \qquad \mu \in A, s \in G.$$

The unique unitary non-degenerate (algebra) representation $(U_\omega)_{\mathrm{ext}}\colon A \to \mathcal{L}(H_\omega)$ extending U_ω is completely determined by

$$((U_\omega)_{\mathrm{ext}}(\mu))(\pi(\nu)) = \pi(\mu * \nu), \qquad \mu, \nu \in A.$$

The vector $x_0 = \pi(\delta_e)$ is a cyclic vector for both representations.

Theorem 17.35. *Let (G, K) be a Gelfand pair.*

(1) For every spherical function of positive type $\omega \in \mathbf{Z}(G/K)$, the cyclic unitary representation $U_\omega\colon G \to \mathbf{U}(H_\omega)$ (with cyclic vector x_0) is irreducible, and its restriction to the compact group K contains the trivial representation of K, which means that

$$F = \{x \in H_\omega \mid U_\omega(t)(x) = x \text{ for all } t \in K\} \neq \{0\}.$$

In fact, the cyclic vector x_0 belongs to F.

(2) Conversely, every unitary representation $U\colon G \to \mathbf{U}(H)$ whose restriction to K contains the trivial representation of K is equivalent to one of the representations U_ω with $\omega \in \mathbf{Z}(G/K)$, and the multiplicity of the trivial representation of K in U is 1.

Theorem 17.35 is proven in Dieudonné [22] (Chapter XXII, Section 9, Lemma 22.9.2). To prove that U_ω is irreducible, it suffices to show that $P = (U_\omega)_{\mathrm{ext}}(\chi_K \lambda_G)$ is the orthogonal projection of H_ω onto the one-dimensional subspace $\mathbb{C}x_0$ Indeed, if F is a closed subspace of H_ω invariant under U_ω, and if F is not orthogonal to x_0, then $P(F) \subseteq F$ and $x_0 \in F$, so $F = H_\omega$ since x_0 is a cyclic vector. On the other hand, if F is orthogonal to x_0, then its orthogonal complement $F^\perp$ is also invariant under U_ω and contains x_0, and since x_0 is a cyclic vector $F^\perp = H_\omega$, and thus $F = (0)$.

The proof of Theorem 17.35 makes use of the following proposition.

Proposition 17.36. *For every irreducible representation $V\colon A \to \mathcal{L}(H)$ of an involutive commutative algebra A in a separable Hilbert space H, if $V(A)$ is separable, then V is a representation in a one-dimensional subspace (of H).*

Proposition 17.36 is proven in Dieudonné [22] (Chapter XXII, Section 9, Lemma 22.9.2.2).

Unlike the case where G is compact, there may not be any closed subspace F of $\mathrm{L}^2(G/K)$, invariant under the canonical representation (see Definition 15.13), and such that the subrepresentation of the canonical representation to F is equivalent to some representation of the form U_ω. However, we have the following results.

Proposition 17.37. *Given a linear map $f\colon E \to E$, if f has rank 1, which means that $\dim(f(E)) = 1$, then there is a linear form $\varphi \in E^*$ and some nonzero vector $u \in E$ such that*

$$f(x) = \varphi(x)u, \qquad \text{for all } x \in E.$$

Proof. This fact is immediately obtained by picking a basis $(e_\alpha)_{\alpha\in I}$ in E. Since f has rank 1, we can pick a nonzero vector $u \in f(E)$, and then $f(e_\alpha) = \lambda_\alpha u$ for some $\lambda_\alpha \in \mathbb{C}$, so we can let φ be the linear form given by $\varphi(e_\alpha) = \lambda_\alpha$. If u is replaced by cu with $c \neq 0$, then φ is replaced by $c^{-1}\varphi$. ∎

We define the trace of f as

$$\mathrm{tr}(f) = \varphi(u),$$

which is independent of the choice of u. If $g\colon E \to E$ is any other linear map, then it is easy to see that $f \circ g$ and $g \circ f$ have rank 1, and that $\mathrm{tr}(f \circ g) = \mathrm{tr}(g \circ f)$.

Proposition 17.38. *The following properties hold.*

(1) For every function $f \in \mathcal{L}^1(G/K)$, the linear map $(U_\omega)_{\mathrm{ext}}(f) \in \mathcal{L}(H_\omega)$ has rank 1, and we have

$$((U_\omega)_{\mathrm{ext}}(f))(z) = \langle z, x_0\rangle((U_\omega)_{\mathrm{ext}}(f))(x_0), \qquad z \in H_\omega,$$

where x_0 is the cyclic vector $x_0 = \pi(\delta_e)$. The trace of $(U_\omega)_{\mathrm{ext}}(f)$ is given by

$$\mathrm{tr}((U_\omega)_{\mathrm{ext}}(f)) = \overline{\mathcal{F}}f(\omega).$$

(2) For any two functions $f, g \in \mathcal{L}^1(G/K) \cap \mathcal{L}^2(G/K)$, we have

$$\mathrm{tr}((U_\omega)_{\mathrm{ext}}(f) \circ (U_\omega)_{\mathrm{ext}}(g)^*) = \overline{\mathcal{F}}(g^* * f)(\omega),$$

an integrable function for the canonical measure $m_{\mathbf{Z}}$ on $\mathbf{Z}(G/K)$, and

$$\int_G f(s)\overline{g(s)}\, d\lambda_G(s) = \int_{\mathbf{Z}(G/K)} \mathrm{tr}((U_\omega)_{\mathrm{ext}}(f) \circ (U_\omega)_{\mathrm{ext}}(g)^*)\, dm_{\mathbf{Z}}(\omega).$$

(3) For every continuous and bounded function $f \in \mathcal{L}^1(G/K)$, such that for all $s \in G$, the function $\overline{\mathcal{F}}(\delta_s * f)$ is integrable for $m_{\mathbf{Z}}$, we have

$$f(s) = \int_{\mathbf{Z}(G/K)} \overline{\mathcal{F}}(\delta_s * f)(\omega)\, dm_{\mathbf{Z}}(\omega) = \int_{\mathbf{Z}(G/K)} \mathrm{tr}(U_\omega(s) \circ (U_\omega)_{\mathrm{ext}}(f))\, dm_{\mathbf{Z}}(\omega).$$

Proposition 17.38 is proven in Dieudonné [22] (Chapter XXII, Section 9, Lemma 22.9.4).

Remark: Using Proposition 17.19 it can be shown that if $g, h \in \mathcal{K}(G/K)$, then $f = g * h$ satisfies the hypothesis of Proposition 17.38(3), namely, $\overline{\mathcal{F}}(\delta_s * f)$ is integrable for $m_{\mathbf{Z}}$. Indeed,

$$\overline{\mathcal{F}}(\delta_s * f) = \overline{\mathcal{F}}(\delta_s * (g * h)) = \overline{\mathcal{F}}((\delta_s * g) * h) = (\overline{\mathcal{F}}(\delta_s * g))(\overline{\mathcal{F}}h)$$

with both factors in $\mathcal{L}^2_{m_{\mathbf{Z}}}(\mathbf{Z}(G/K))$, so $\overline{\mathcal{F}}(\delta_s * f)$ is $m_{\mathbf{Z}}$-integrable. We also have $f = g * h \in \mathcal{K}(G/K)$.

In the special case where G is a commutative locally compact metrizable and separable group, we have the following results about unitary representations.

Proposition 17.39. *Let G be a commutative locally compact metrizable and separable group. Every unitary cyclic representation $U\colon G \to \mathbf{U}(H)$ of G in a separable Hilbert space H is equivalent to a representation $M\colon G \to \mathbf{U}(\mathrm{L}^2_\mu(\widehat{G}))$, where μ is a positive bounded measure on $\widehat{G}$ (the dual of G), and for every $s \in G$, the linear operator $M(s)$ is defined so that for every $g \in \mathcal{L}^2_\mu(\widehat{G})$, $M(s)(g)$ is the class of the function in $\mathcal{L}^2_\mu(\widehat{G})$ given by*

$$\chi \mapsto \chi(s)g(\chi), \qquad \chi \in \widehat{G}.$$

The proof of Proposition 17.39 is proven in Dieudonné [22] (Chapter XXII, Section 9, Lemma 22.15.1). The proof uses Bochner's theorem (see Proposition 17.28).

If $G = \mathbb{R}$, there is a more precise result due to Stone.

Theorem 17.40. *(Stone) Every unitary representation of the (additive) group $\mathbb{R}$ in a sepatable Hilbert space H is of the form*

$$t \mapsto e^{itA},$$

where A is a self-adjoint operator of H, not necessarily bounded. Conversely, for every self-adjoint not necessarily bounded operator A of H, the map $t \mapsto e^{itA}$ is a unitary representation of $\mathbb{R}$ in H.

Theorem 17.40 is proven in Dieudonné [22] (Chapter XXII, Section 9, Lemma 22.15.3).

Appendix A

Topology

A.1 Metric Spaces and Normed Vector Spaces

This chapter contains a review of basic topological concepts. First metric spaces are defined. Next normed vector spaces are defined. Closed and open sets are defined, and their basic properties are stated. The general concept of a topological space is defined. The closure and the interior of a subset are defined. The subspace topology and the product topology are defined. Continuous maps and homeomorphisms are defined. Limits of sequences are defined. Continuous linear maps and multilinear maps are defined and studied briefly.

Most spaces considered in this book have a topological structure given by a metric or a norm, and we first review these notions. We begin with metric spaces. Recall that $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$.

Definition A.1. A *metric space* is a set E together with a function $d\colon E \times E \to \mathbb{R}_+$, called a *metric, or distance*, assigning a nonnegative real number $d(x, y)$ to any two points $x, y \in E$, and satisfying the following conditions for all $x, y, z \in E$:

(D1) $d(x, y) = d(y, x)$. (symmetry)

(D2) $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$. (positivity)

(D3) $d(x, z) \leq d(x, y) + d(y, z)$. (triangle inequality)

Geometrically, Condition (D3) expresses the fact that in a triangle with vertices x, y, z, the length of any side is bounded by the sum of the lengths of the other two sides. From (D3), we immediately get

$$|d(x, y) - d(y, z)| \leq d(x, z).$$

Let us give some examples of metric spaces. Recall that the *absolute value* $|x|$ of a real number $x \in \mathbb{R}$ is defined such that $|x| = x$ if $x \geq 0$, $|x| = -x$ if $x < 0$, and for a complex number $x = a + ib$, by $|x| = \sqrt{a^2 + b^2}$.

Example A.1.

1. Let $E = \mathbb{R}$, and $d(x, y) = |x - y|$, the absolute value of $x - y$. This is the so-called natural metric on $\mathbb{R}$.

2. Let $E = \mathbb{R}^n$ (or $E = \mathbb{C}^n$). We have the *Euclidean metric*

$$d_2(x, y) = \left(|x_1 - y_1|^2 + \cdots + |x_n - y_n|^2\right)^{\frac{1}{2}},$$

   the distance between the points $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$.

3. For every set E, we can define the *discrete metric*, defined such that $d(x, y) = 1$ iff $x \neq y$, and $d(x, x) = 0$.

4. For any $a, b \in \mathbb{R}$ such that $a < b$, we define the following sets:

   $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$, (closed interval)

   $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$, (open interval)

   $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$, (interval closed on the left, open on the right)

   $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$, (interval open on the left, closed on the right)

   Let $E = [a, b]$, and $d(x, y) = |x - y|$. Then, $([a, b], d)$ is a metric space.

We will need to define the notion of proximity in order to define convergence of limits and continuity of functions. For this we introduce some standard "small neighborhoods."

Definition A.2. Given a metric space E with metric d, for every $a \in E$, for every $\rho \in \mathbb{R}$, with $\rho > 0$, the set

$$B(a, \rho) = \{x \in E \mid d(a, x) \leq \rho\}$$

is called the *closed ball of center a and radius ρ*, the set

$$B_0(a, \rho) = \{x \in E \mid d(a, x) < \rho\}$$

is called the *open ball of center a and radius ρ*, and the set

$$S(a, \rho) = \{x \in E \mid d(a, x) = \rho\}$$

is called the *sphere of center a and radius ρ*. It should be noted that ρ is finite (i.e., not $+\infty$). A subset X of a metric space E is *bounded* if there is a closed ball $B(a, \rho)$ such that $X \subseteq B(a, \rho)$.

Clearly, $B(a, \rho) = B_0(a, \rho) \cup S(a, \rho)$.

Example A.2.

1. In $E = \mathbb{R}$ with the distance $|x - y|$, an open ball of center a and radius ρ is the open interval $(a - \rho, a + \rho)$.

2. In $E = \mathbb{R}^2$ with the Euclidean metric, an open ball of center a and radius ρ is the set of points inside the disk of center a and radius ρ, excluding the boundary points on the circle.

3. In $E = \mathbb{R}^3$ with the Euclidean metric, an open ball of center a and radius ρ is the set of points inside the sphere of center a and radius ρ, excluding the boundary points on the sphere.

One should be aware that intuition can be misleading in forming a geometric image of a closed (or open) ball. For example, if d is the discrete metric, a closed ball of center a and radius $\rho < 1$ consists only of its center a, and a closed ball of center a and radius $\rho \geq 1$ consists of the entire space!

If $E = [a, b]$, and $d(x, y) = |x - y|$, as in Example A.1, an open ball $B_0(a, \rho)$, with $\rho < b - a$, is in fact the interval $[a, a + \rho)$, which is closed on the left.

We now consider a very important special case of metric spaces, normed vector spaces. Normed vector spaces have already been defined in Chapter B (Definition B.1) but for the reader's convenience we repeat the definition.

Definition A.3. Let E be a vector space over a field K, where K is either the field $\mathbb{R}$ of reals, or the field $\mathbb{C}$ of complex numbers. A *norm on E* is a function $\| \, \| \colon E \to \mathbb{R}_+$, assigning a nonnegative real number $\|u\|$ to any vector $u \in E$, and satisfying the following conditions for all $x, y \in E$:

(N1) $\|x\| \geq 0$, and $\|x\| = 0$ iff $x = 0$. (positivity)

(N2) $\|\lambda x\| = |\lambda| \, \|x\|$. (scaling)

(N3) $\|x + y\| \leq \|x\| + \|y\|$. (triangle inequality)

A vector space E together with a norm $\| \, \|$ is called a *normed vector space*. A function $\| \, \| \colon E \to \mathbb{R}_+$ satifying only properties (N2) and (N3) is called a *semi-norm*.

From (N3), we easily get

$$|\|x\| - \|y\|| \leq \|x - y\|.$$

Given a normed vector space E, if we define d such that

$$d(x, y) = \|x - y\|,$$

it is easily seen that d is a metric. Thus, every normed vector space is immediately a metric space. Note that the metric associated with a norm is invariant under translation, that is,

$$d(x+u,\, y+u) = d(x,\, y).$$

For this reason, we can restrict ourselves to open or closed balls of center 0.

If $\| \, \| \colon E \to \mathbb{R}_+$ is a semi-norm, then $\|x\| = 0$ does not necessarily imply that $x = 0$. However by setting $\lambda = 0$ and $x = 0$ in (N2), we see that $\|0\| = 0$. If we let $\mathcal{N} = \{x \in E \mid \|x\| = 0\}$, then $\mathcal{N}$ is a subspace of E. Indeed, $0 \in \mathcal{N}$, and if $\|x\| = \|y\| = 0$, then by (N2) and (N3) we have

$$\|\lambda x + \mu y\| \leq \|\lambda x\| + \|\mu y\| = |\lambda| \, \|x\| + |\mu| \, \|y\| = 0 + 0 = 0,$$

so $\lambda x + \mu y \in \mathcal{N}$. We can form the quotient space $E/\mathcal{N}$, and then it is easy to see that the semi-norm $\| \, \|$ induces a norm on $X/\mathcal{N}$.

Natural examples of semi-norms arise in integration theory; see Chapter 5.

Examples of normed vector spaces were given in Example B.1. We repeat the most important examples.

Example A.3. Let $E = \mathbb{R}^n$ (or $E = \mathbb{C}^n$). There are three standard norms. For every $(x_1, \ldots, x_n) \in E$, we have the norm $\|x\|_1$, defined such that,

$$\|x\|_1 = |x_1| + \cdots + |x_n|,$$

we have the *Euclidean norm* $\|x\|_2$, defined such that,

$$\|x\|_2 = \left(|x_1|^2 + \cdots + |x_n|^2\right)^{\frac{1}{2}},$$

and the *sup*-norm $\|x\|_\infty$, defined such that,

$$\|x\|_\infty = \max\{|x_i| \mid 1 \leq i \leq n\}.$$

More generally, we define the *ℓ^p-norm* (for $p \geq 1$) by

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

We proved in Proposition B.1 that the ℓ^p-norms are indeed norms. The closed unit balls centered at $(0, 0)$ for $\| \, \|_1$, $\| \, \|_2$, and $\| \, \|_\infty$, along with the containment relationships, are shown in Figures A.1 and A.2. Figures A.3 and A.4 illustrate the situation in $\mathbb{R}^3$.

In a normed vector space we define a closed ball or an open ball of radius ρ as a closed ball or an open ball of center 0. We may use the notation $B(\rho)$ and $B_0(\rho)$.

We will now define the crucial notions of open sets and closed sets, and of a topological space.

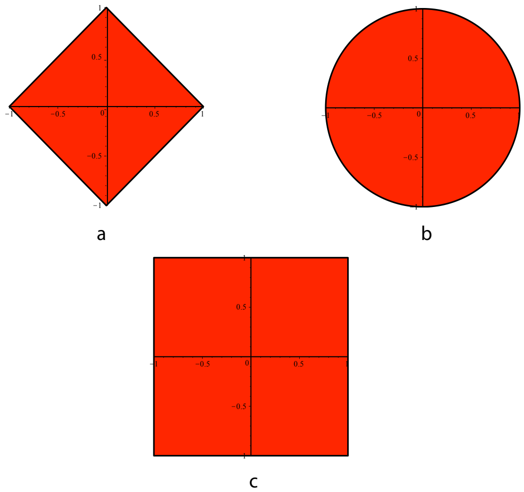


Figure A.1: Figure (a) shows the diamond shaped closed ball associated with $\| \|_1$. Figure (b) shows the closed unit disk associated with $\| \|_2$, while Figure (c) illustrates the closed unit ball associated with $\| \|_\infty$.

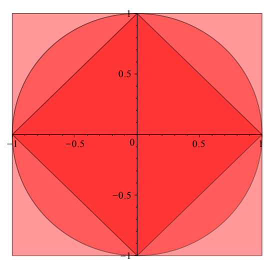


Figure A.2: The relationship between the closed unit balls centered at $(0, 0)$.

Definition A.4. Let (E, d) be a metric space. A subset $U \subseteq E$ is an *open set* in E if either $U = \emptyset$, or for every $a \in U$, there is some open ball $B_0(a, \rho)$ such that, $B_0(a, \rho) \subseteq U$.[1] A subset $F \subseteq E$ is a *closed set* in E if its complement $E - F$ is open in E. See Figure A.5.

The set E itself is open, since for every $a \in E$, every open ball of center a is contained in E. In $E = \mathbb{R}^n$, given n intervals $[a_i, b_i]$, with $a_i < b_i$, it is easy to show that the open n-cube

$$\{(x_1, \ldots, x_n) \in E \mid a_i < x_i < b_i,\ 1 \leq i \leq n\}$$

is an open set. In fact, it is possible to find a metric for which such open n-cubes are open balls! Similarly, we can define the closed n-cube

$$\{(x_1, \ldots, x_n) \in E \mid a_i \leq x_i \leq b_i,\ 1 \leq i \leq n\},$$

[1]Recall that $\rho > 0$.

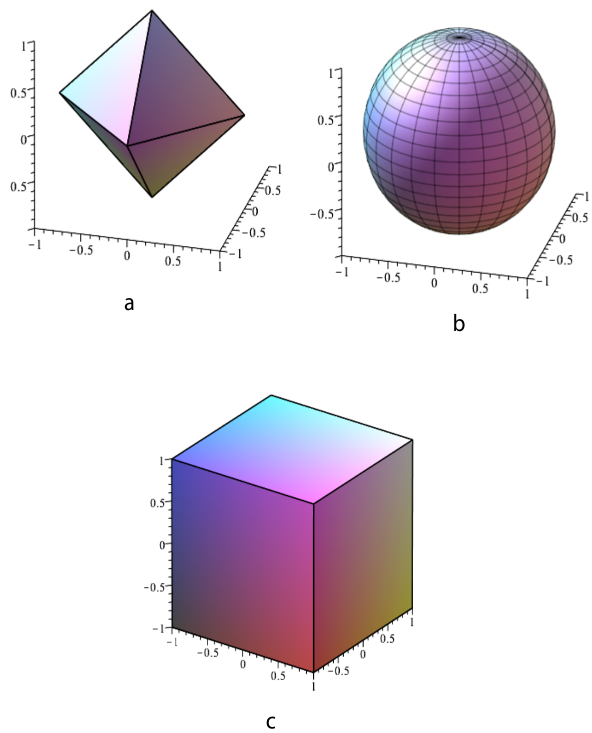


Figure A.3: Figure (a) shows the octahedral shaped closed ball associated with $\| \|_1$. Figure (b) shows the closed spherical associated with $\| \|_2$, while Figure (c) illustrates the closed unit ball associated with $\| \|_\infty$.

which is a closed set.

The open sets satisfy some important properties that lead to the definition of a topological space.

Proposition A.1. *Given a metric space E with metric d, the family $\mathcal{O}$ of all open sets defined in Definition A.4 satisfies the following properties:*

(O1) For every finite family $(U_i)_{1\leq i\leq n}$ of sets $U_i \in \mathcal{O}$, we have $U_1 \cap \cdots \cap U_n \in \mathcal{O}$, i.e., $\mathcal{O}$ is closed under finite intersections.

(O2) For every arbitrary family $(U_i)_{i\in I}$ of sets $U_i \in \mathcal{O}$, we have $\bigcup_{i\in I} U_i \in \mathcal{O}$, i.e., $\mathcal{O}$ is closed under arbitrary unions.

(O3) $\emptyset \in \mathcal{O}$, and $E \in \mathcal{O}$, i.e., $\emptyset$ and E belong to $\mathcal{O}$.

Furthermore, for any two distinct points $a \neq b$ in E, there exist two open sets U_a and U_b such that, $a \in U_a$, $b \in U_b$, and $U_a \cap U_b = \emptyset$.

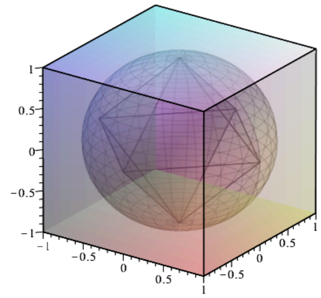


Figure A.4: The relationship between the closed unit balls centered at $(0, 0, 0)$.

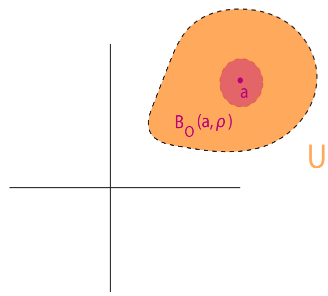


Figure A.5: An open set U in $E = \mathbb{R}^2$ under the standard Euclidean metric. Any point in the peach set U is surrounded by a small raspberry open set which lies within U.

Proof. It is straightforward. For the last point, letting $\rho = d(a, b)/3$ (in fact $\rho = d(a, b)/2$ works too), we can pick $U_a = B_0(a, \rho)$ and $U_b = B_0(b, \rho)$. By the triangle inequality, we must have $U_a \cap U_b = \emptyset$. $\square$

The above proposition leads to the very general concept of a topological space.

One should be careful that, in general, the family of open sets is not closed under infinite intersections. For example, in $\mathbb{R}$ under the metric $|x - y|$, letting $U_n = (-1/n, +1/n)$, each U_n is open, but $\bigcap_n U_n = \{0\}$, which is not open.

Later on, given any nonempty subset A of a metric space (E, d), we will need to know that certain special sets containing A are open.

Definition A.5. Let (E, d) be a metric space. For any nonempty subset A of E and any $x \in E$, let

$$d(x, A) = \inf_{a \in A} d(x, a).$$

Proposition A.2. *Let (E, d) be a metric space. For any nonempty subset A of E and for any two points $x, y \in E$, we have*

$$|d(x, A) - d(y, A)| \leq d(x, y).$$

Proof. For all $a \in A$ we have

$$d(x, a) \leq d(x, y) + d(y, a),$$

which implies

$$\begin{aligned} d(x, A) &= \inf_{a \in A} d(x, a) \\ &\leq \inf_{a \in A} (d(x, y) + d(y, a)) \\ &= d(x, y) + \inf_{a \in A} d(y, a) \\ &= d(x, y) + d(y, A). \end{aligned}$$

By symmetry, we also obtain $d(y, A) \leq d(x, y) + d(x, A)$, and thus

$$|d(x, A) - d(y, A)| \leq d(x, y),$$

as claimed. □

Definition A.6. Let (E, d) be a metric space. For any nonempty subset A of E, and any $r > 0$, let

$$V_r(A) = \{x \in E \mid d(x, A) < r\}.$$

Proposition A.3. *Let (E, d) be a metric space. For any nonempty subset A of E, and any $r > 0$, the set $V_r(A)$ is an open set containing A.*

Proof. For any $y \in E$ such that $d(x, y) < r - d(x, A)$, by Proposition A.2 we have

$$d(y, A) \leq d(x, A) + d(x, y) \leq d(x, A) + r - d(x, A) = r,$$

so $V_r(A)$ contains the open ball $B_0(x, r - d(x, A))$, which means that it is open. Obviously, $A \subseteq V_r(A)$. □

A.2 Topological Spaces

Motivated by Proposition A.1, a topological space is defined in terms of a family of sets satisfying the properties of open sets stated in that proposition.

Definition A.7. Given a set E, a *topology on E (or a topological structure on E)*, is defined as a family $\mathcal{O}$ of subsets of E called *open sets*, and satisfying the following three properties:

(1) For every finite family $(U_i)_{1\leq i\leq n}$ of sets $U_i \in \mathcal{O}$, we have $U_1 \cap \cdots \cap U_n \in \mathcal{O}$, i.e., $\mathcal{O}$ is closed under finite intersections.

(2) For every arbitrary family $(U_i)_{i\in I}$ of sets $U_i \in \mathcal{O}$, we have $\bigcup_{i\in I} U_i \in \mathcal{O}$, i.e., $\mathcal{O}$ is closed under arbitrary unions.

(3) $\emptyset \in \mathcal{O}$, and $E \in \mathcal{O}$, i.e., $\emptyset$ and E belong to $\mathcal{O}$.

A set E together with a topology $\mathcal{O}$ on E is called a *topological space*. Given a topological space $(E, \mathcal{O})$, a subset F of E is a *closed set* if $F = E - U$ for some open set $U \in \mathcal{O}$, i.e., F is the complement of some open set.

It is possible that an open set is also a closed set. For example, $\emptyset$ and E are both open and closed. When a topological space contains a proper nonempty subset U which is both open and closed, the space E is said to be *disconnected*.

Definition A.8. A topological space $(E, \mathcal{O})$ is said to satisfy the *Hausdorff separation axiom (or T_2-separation axiom)* if for any two distinct points $a \neq b$ in E, there exist two open sets U_a and U_b such that, $a \in U_a$, $b \in U_b$, and $U_a \cap U_b = \emptyset$. When the T_2-separation axiom is satisfied, we also say that $(E, \mathcal{O})$ is a *Hausdorff space*.

As shown by Proposition A.1, any metric space is a topological Hausdorff space, the family of open sets being in fact the family of arbitrary unions of open balls. Similarly, any normed vector space is a topological Hausdorff space, the family of open sets being the family of arbitrary unions of open balls. The topology $\mathcal{O}$ consisting of all subsets of E is called the *discrete topology*.

Remark: Most (if not all) spaces used in analysis are Hausdorff spaces. Intuitively, the Hausdorff separation axiom says that there are enough "small" open sets. Without this axiom, some counter-intuitive behaviors may arise. For example, a sequence may have more than one limit point (or a compact set may not be closed). Nevertheless, non-Hausdorff topological spaces arise naturally in algebraic geometry. But even there, some substitute for separation is used.

One of the reasons why topological spaces are important is that the definition of a topology only involves a certain family $\mathcal{O}$ of sets, and not **how** such family is generated from a metric or a norm. For example, different metrics or different norms can define the same family of open sets. Many topological properties only depend on the family $\mathcal{O}$ and not on the specific metric or norm. But the fact that a topology is definable from a metric or a norm is important, because it usually implies nice properties of a space. All our examples will be spaces whose topology is defined by a metric or a norm.

Definition A.9. A topological space $(E, \mathcal{O})$ is *metrizable* if there is a distance on E defining the topology $\mathcal{O}$.

Note that in a metric space (E, d), the metric d is *explicitly given*. However, in general, the topology of a metrizable space $(E, \mathcal{O})$ is not specified by an explicitly given metric, but *some metric* defining the topology $\mathcal{O}$ exists. Obviously, a metrizable topological space must be Hausdorff. Actually, a stronger separation property holds, a metrizable space is normal; see Definition A.30.

Remark: By taking complements we can state properties of the closed sets dual to those of Definition A.7. Thus, $\emptyset$ and E are closed sets, and the closed sets are closed under finite unions and arbitrary intersections.

It is also worth noting that the Hausdorff separation axiom implies that for every $a \in E$, the set $\{a\}$ is closed. Indeed, if $x \in E - \{a\}$, then $x \neq a$, and so there exist open sets U_a and U_x such that $a \in U_a$, $x \in U_x$, and $U_a \cap U_x = \emptyset$. See Figure A.6. Thus, for every $x \in E - \{a\}$, there is an open set U_x containing x and contained in $E - \{a\}$, showing by (O3) that $E - \{a\}$ is open, and thus that the set $\{a\}$ is closed.

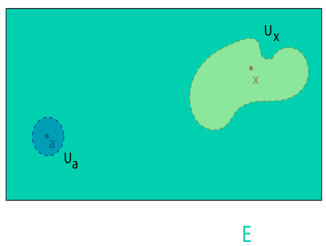


Figure A.6: A schematic illustration of the Hausdorff separation property

Given a topological space $(E, \mathcal{O})$, given any subset A of E, since $E \in \mathcal{O}$ and E is a closed set, the family $\mathcal{C}_A = \{F \mid A \subseteq F,\ F \text{ a closed set}\}$ of closed sets containing A is nonempty, and since any arbitrary intersection of closed sets is a closed set, the intersection $\bigcap \mathcal{C}_A$ of the sets in the family $\mathcal{C}_A$ is the smallest closed set containing A. By a similar reasoning, the union of all the open subsets contained in A is the largest open set contained in A.

Definition A.10. Given a topological space $(E, \mathcal{O})$, given any subset A of E, the smallest closed set containing A is denoted by $\overline{A}$, and is called the *closure, or adherence of* A. See Figure A.7. A subset A of E is *dense in* E if $\overline{A} = E$. The largest open set contained in A is denoted by $\mathring{A}$, and is called the *interior of* A. See Figure A.8. The set $\mathrm{Fr}\, A = \overline{A} \cap \overline{E - A}$ is called the *boundary (or frontier) of* A. We also denote the boundary of A by ∂A. See Figure A.9.

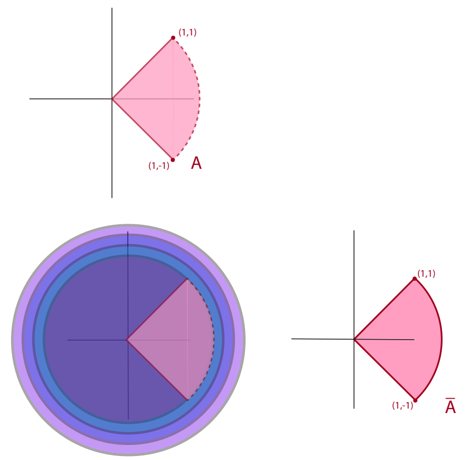


Figure A.7: The topological space $(E, \mathcal{O})$ is $\mathbb{R}^2$ with topology induced by the Euclidean metric. The subset A is the section $B_0(1)$ in the first and fourth quadrants bound by the lines $y = x$ and $y = -x$. The closure of A is obtained by the intersection of A with the closed unit ball.

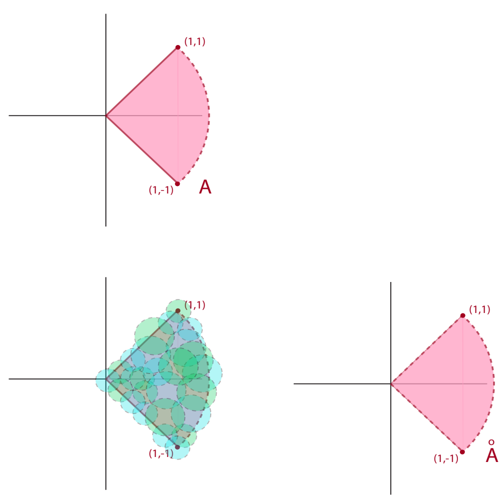


Figure A.8: The topological space $(E, \mathcal{O})$ is $\mathbb{R}^2$ with topology induced by the Euclidean metric. The subset A is the section $B_0(1)$ in the first and fourth quadrants bound by the lines $y = x$ and $y = -x$. The interior of A is obtained by the covering A with small open balls.

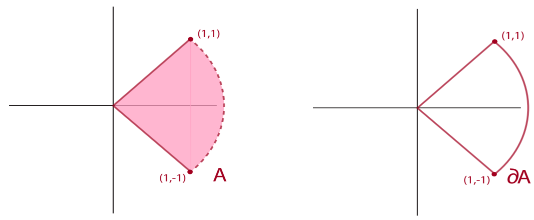


Figure A.9: The topological space $(E, \mathcal{O})$ is $\mathbb{R}^2$ with topology induced by the Euclidean metric. The subset A is the section $B_0(1)$ in the first and fourth quadrants bound by the lines $y = x$ and $y = -x$. The boundary of A is $\overline{A} - \overset{\circ}{A}$.

Remark: The notation $\overline{A}$ for the closure of a subset A of E is somewhat unfortunate, since $\overline{A}$ is often used to denote the set complement of A in E. Still, we prefer it to more cumbersome notations such as $\mathrm{clo}(A)$, and we denote the complement of A in E by $E - A$ (or sometimes, A^c).

By definition, it is clear that a subset A of E is closed iff $A = \overline{A}$. The set $\mathbb{Q}$ of rationals is dense in $\mathbb{R}$. It is easily shown that $\overline{A} = \overset{\circ}{A} \cup \partial A$ and $\overset{\circ}{A} \cap \partial A = \emptyset$. Another useful characterization of $\overline{A}$ is given by the following proposition.

Proposition A.4. *Given a topological space $(E, \mathcal{O})$, given any subset A of E, the closure $\overline{A}$ of A is the set of all points $x \in E$ such that for every open set U containing x, then $U \cap A \neq \emptyset$. See Figure A.10.*

Proof. If $A = \emptyset$, since $\emptyset$ is closed, the proposition holds trivially. Thus, assume that $A \neq \emptyset$. First assume that $x \in \overline{A}$. Let U be any open set such that $x \in U$. If $U \cap A = \emptyset$, since U is open, then $E - U$ is a closed set containing A, and since $\overline{A}$ is the intersection of all closed sets containing A, we must have $x \in E - U$, which is impossible. Conversely, assume that $x \in E$ is a point such that for every open set U containing x, then $U \cap A \neq \emptyset$. Let F be any closed subset containing A. If $x \notin F$, since F is closed, then $U = E - F$ is an open set such that $x \in U$, and $U \cap A = \emptyset$, a contradiction. Thus, we have $x \in F$ for every closed set containing A, that is, $x \in \overline{A}$. $\square$

Often it is necessary to consider a subset A of a topological space E, and to view the subset A as a topological space. The following proposition shows how to define a topology on a subset.

Proposition A.5. *Given a topological space $(E, \mathcal{O})$, given any subset A of E, let*

$$\mathcal{U} = \{U \cap A \mid U \in \mathcal{O}\}$$

be the family of all subsets of A obtained as the intersection of any open set in $\mathcal{O}$ with A. The following properties hold.

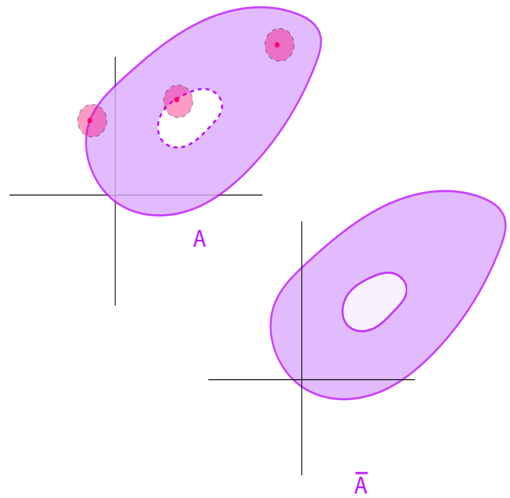


Figure A.10: The topological space $(E, \mathcal{O})$ is $\mathbb{R}^2$ with topology induced by the Euclidean metric. The purple subset A is illustrated with three red points, each in its closure since the open ball centered at each point has nontrivial intersection with A.

(1) *The space* $(A, \mathcal{U})$ *is a topological space.*

(2) *If* E *is a metric space with metric* d*, then the restriction* $d_A \colon A \times A \to \mathbb{R}_+$ *of the metric* d *to* A *defines a metric space. Furthermore, the topology induced by the metric* d_A *agrees with the topology defined by* $\mathcal{U}$*, as above.*

Proof. Left as an exercise. $\square$

Proposition A.5 suggests the following definition.

Definition A.11. Given a topological space $(E, \mathcal{O})$, given any subset A of E, the *subspace topology on* A *induced by* $\mathcal{O}$ is the family $\mathcal{U}$ of open sets defined such that

$$\mathcal{U} = \{U \cap A \mid U \in \mathcal{O}\}$$

is the family of all subsets of A obtained as the intersection of any open set in $\mathcal{O}$ with A. We say that $(A, \mathcal{U})$ has the *subspace topology*. If (E, d) is a metric space, the restriction $d_A \colon A \times A \to \mathbb{R}_+$ of the metric d to A is called the *subspace metric*.

For example, if $E = \mathbb{R}^n$ and d is the Euclidean metric, we obtain the subspace topology on the closed n-cube

$$\{(x_1, \ldots, x_n) \in E \mid a_i \leq x_i \leq b_i,\ 1 \leq i \leq n\}.$$

See Figure A.11,

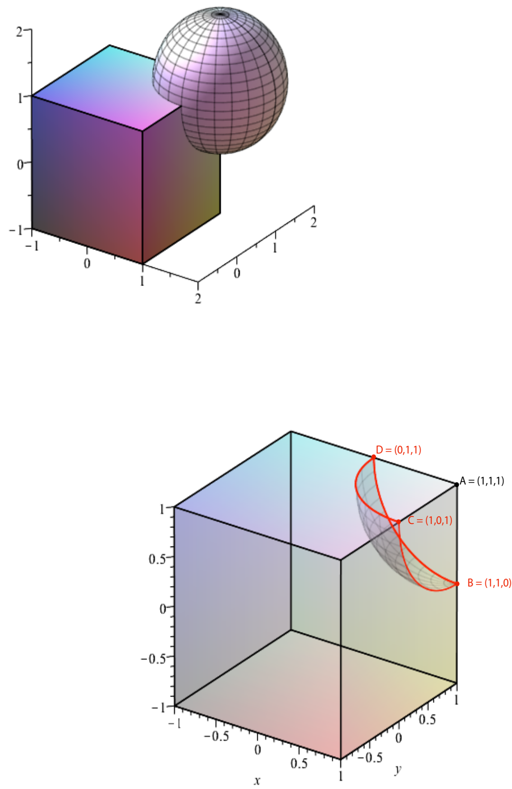


Figure A.11: An example of an open set in the subspace topology for $\{(x, y, z) \in \mathbb{R}^3 \mid -1 \leq x \leq 1, -1 \leq y \leq 1, -1 \leq z \leq 1\}$. The open set is the corner region $ABCD$ and is obtained by intersecting with the cube $B_0((1, 1, 1), 1)$.

One should realize that every open set $U \in \mathcal{O}$ which is entirely contained in A is also in the family $\mathcal{U}$, but $\mathcal{U}$ may contain open sets that are not in $\mathcal{O}$. For example, if $E = \mathbb{R}$ with $|x - y|$, and $A = [a, b]$, then sets of the form $[a, c)$, with $a < c < b$ belong to $\mathcal{U}$, but they are not open sets for $\mathbb{R}$ under $|x - y|$. However, there is agreement in the following situation.

Proposition A.6. *Given a topological space $(E, \mathcal{O})$, given any subset A of E, if $\mathcal{U}$ is the subspace topology, then the following properties hold.*

(1) If A is an open set $A \in \mathcal{O}$, then every open set $U \in \mathcal{U}$ is an open set $U \in \mathcal{O}$.

(2) If A is a closed set in E, then every closed set w.r.t. the subspace topology is a closed set w.r.t. $\mathcal{O}$.

Proof. Left as an exercise. $\square$

The concept of product topology is also useful. We have the following proposition.

Proposition A.7. *Given n topological spaces $(E_i, \mathcal{O}_i)$, let $\mathcal{B}$ be the family of subsets of $E_1 \times \cdots \times E_n$ defined as follows:*

$$\mathcal{B} = \{U_1 \times \cdots \times U_n \mid U_i \in \mathcal{O}_i,\ 1 \leq i \leq n\},$$

and let $\mathcal{P}$ be the family consisting of arbitrary unions of sets in $\mathcal{B}$, including $\emptyset$. Then $\mathcal{P}$ is a topology on $E_1 \times \cdots \times E_n$.

Proof. Left as an exercise. □

Definition A.12. Given n topological spaces $(E_i, \mathcal{O}_i)$, the *product topology on* $E_1 \times \cdots \times E_n$ is the family $\mathcal{P}$ of subsets of $E_1 \times \cdots \times E_n$ defined as follows: if

$$\mathcal{B} = \{U_1 \times \cdots \times U_n \mid U_i \in \mathcal{O}_i,\ 1 \leq i \leq n\},$$

then $\mathcal{P}$ is the family consisting of arbitrary unions of sets in $\mathcal{B}$, including $\emptyset$. See Figure A.12.

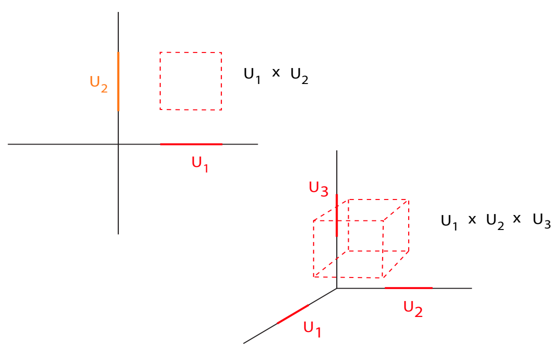


Figure A.12: Examples of open sets in the product topology for $\mathbb{R}^2$ and $\mathbb{R}^3$ induced by the Euclidean metric.

If each (E_i, d_{E_i}) is a metric space, there are three natural metrics that can be defined on $E_1 \times \cdots \times E_n$:

$$\begin{aligned}
d_1((x_1, \ldots, x_n), (y_1, \ldots, y_n)) &= d_{E_1}(x_1, y_1) + \cdots + d_{E_n}(x_n, y_n),\\
d_2((x_1, \ldots, x_n), (y_1, \ldots, y_n)) &= \left((d_{E_1}(x_1, y_1))^2 + \cdots + (d_{E_n}(x_n, y_n))^2\right)^{\frac{1}{2}},\\
d_\infty((x_1, \ldots, x_n), (y_1, \ldots, y_n)) &= \max\{d_{E_1}(x_1, y_1), \ldots, d_{E_n}(x_n, y_n)\}.
\end{aligned}$$

It is easy to show that

$$\begin{aligned}
d_\infty((x_1, \ldots, x_n), (y_1, \ldots, y_n)) \leq d_2((x_1, \ldots, x_n), (y_1, \ldots, y_n)) &\leq d_1((x_1, \ldots, x_n), (y_1, \ldots, y_n))\\
&\leq n d_\infty((x_1, \ldots, x_n), (y_1, \ldots, y_n)),
\end{aligned}$$

so these distances define the same topology, which is the product topology.

If each $(E_i, \| \|_{E_i})$ is a normed vector space, there are three natural norms that can be defined on $E_1 \times \cdots \times E_n$:

$$\begin{aligned}
\|(x_1, \ldots, x_n)\|_1 &= \|x_1\|_{E_1} + \cdots + \|x_n\|_{E_n}, \\
\|(x_1, \ldots, x_n)\|_2 &= \left(\|x_1\|_{E_1}^2 + \cdots + \|x_n\|_{E_n}^2\right)^{\frac{1}{2}}, \\
\|(x_1, \ldots, x_n)\|_\infty &= \max\left\{\|x_1\|_{E_1}, \ldots, \|x_n\|_{E_n}\right\}.
\end{aligned}$$

It is easy to show that

$$\|(x_1, \ldots, x_n)\|_\infty \leq \|(x_1, \ldots, x_n)\|_2 \leq \|(x_1, \ldots, x_n)\|_1 \leq n\|(x_1, \ldots, x_n)\|_\infty,$$

so these norms define the same topology, which is the product topology. It can also be verified that when $E_i = \mathbb{R}$, with the standard topology induced by $|x - y|$, the topology product on $\mathbb{R}^n$ is the standard topology induced by the Euclidean norm.

Definition A.13. Two metrics d and d' on a space E are *equivalent* if they induce the same topology $\mathcal{O}$ on E (i.e., they define the same family $\mathcal{O}$ of open sets). Similarly, two norms $\| \|$ and $\| \|'$ on a space E are *equivalent* if they induce the same topology $\mathcal{O}$ on E.

Given a topological space $(E, \mathcal{O})$, it is often useful, as in Proposition A.7, to define the topology $\mathcal{O}$ in terms of a subfamily $\mathcal{B}$ of subsets of E.

Definition A.14. We say that a family $\mathcal{B}$ of subsets of E is a *basis for the topology* $\mathcal{O}$, if $\mathcal{B}$ is a subset of $\mathcal{O}$, and if every open set U in $\mathcal{O}$ can be obtained as some union (possibly infinite) of sets in $\mathcal{B}$ (agreeing that the empty union is the empty set).

For example, given any metric space (E, d), $\mathcal{B} = \{B_0(a, \rho) \mid a \in E, \rho > 0\}$. In particular, if $d = \| \|_2$, the open intervals form a basis for $\mathbb{R}$, while the open disks form a basis for $\mathbb{R}^2$. The open rectangles also form a basis for $\mathbb{R}^2$ with the standard topology. See Figure A.13.

It is immediately verified that if a family $\mathcal{B} = (U_i)_{i \in I}$ is a basis for the topology of $(E, \mathcal{O})$, then $E = \bigcup_{i \in I} U_i$, and the intersection of any two sets $U_i, U_j \in \mathcal{B}$ is the union of some sets in the family $\mathcal{B}$ (again, agreeing that the empty union is the empty set). Conversely, a family $\mathcal{B}$ with these properties is the basis of the topology obtained by forming arbitrary unions of sets in $\mathcal{B}$.

Definition A.15. A *subbasis for* $\mathcal{O}$ is a family $\mathcal{S}$ of subsets of E, such that the family $\mathcal{B}$ of all finite intersections of sets in $\mathcal{S}$ (including E itself, in case of the empty intersection) is a basis of $\mathcal{O}$. See Figure A.13.

The following proposition gives useful criteria for determining whether a family of open subsets is a basis of a topological space.

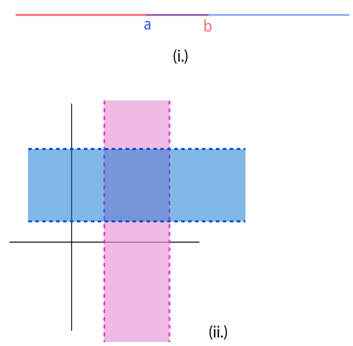


Figure A.13: Figure (*i.*) shows that the set of infinite open intervals forms a subbasis for $\mathbb{R}$. Figure (*ii.*) shows that the infinite open strips form a subbasis for $\mathbb{R}^2$.

Proposition A.8. *Given a topological space $(E, \mathcal{O})$ and a family $\mathcal{B}$ of open subsets in $\mathcal{O}$ the following properties hold:*

(1) The family $\mathcal{B}$ is a basis for the topology $\mathcal{O}$ iff for every open set $U \in \mathcal{O}$ and every $x \in U$, there is some $B \in \mathcal{B}$ such that $x \in B$ and $B \subseteq U$. See Figure A.14.

(2) The family $\mathcal{B}$ is a basis for the topology $\mathcal{O}$ iff

(a) For every $x \in E$, there is some $B \in \mathcal{B}$ such that $x \in B$.

(b) For any two open subsets, $B_1, B_2 \in \mathcal{B}$, for every $x \in E$, if $x \in B_1 \cap B_2$, then there is some $B_3 \in \mathcal{B}$ such that $x \in B_3$ and $B_3 \subseteq B_1 \cap B_2$. See Figure A.15.

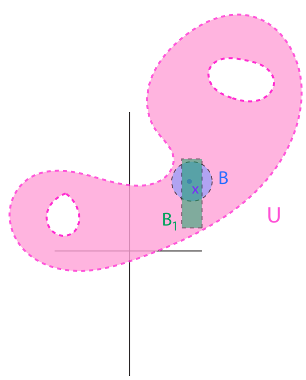


Figure A.14: Given an open subset U of $\mathbb{R}^2$ and $x \in U$, there exists an open ball B containing x with $B \subset U$. There also exists an open rectangle B_1 containing x with $B_1 \subset U$.

We now consider the fundamental property of continuity.

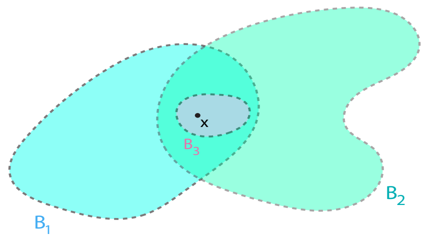


Figure A.15: A schematic illustration of Condition (b) in Proposition A.8.

A.3 Continuous Functions, Limits

Definition A.16. Let $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ be topological spaces, and let $f: E \to F$ be a function. For every $a \in E$, we say that f *is continuous at* a, if for every open set $V \in \mathcal{O}_F$ containing $f(a)$, there is some open set $U \in \mathcal{O}_E$ containing a, such that, $f(U) \subseteq V$. See Figure A.16. We say that f *is continuous* if it is continuous at every $a \in E$.

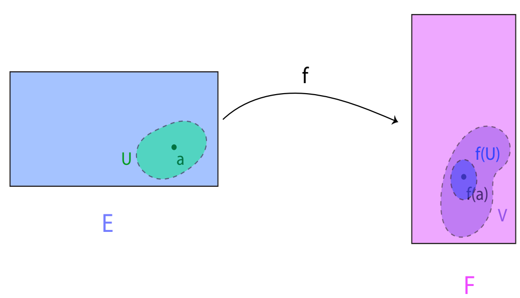


Figure A.16: A schematic illustration of Definition A.16.

Define a *neighborhood of* $a \in E$ as any subset N of E containing some open set $O \in \mathcal{O}$ such that $a \in O$. If f is continuous at a and N is any neighborhood of $f(a)$, there is some open set $V \subseteq N$ containing $f(a)$, and since f is continuous at a, there is some open set U containing a, such that $f(U) \subseteq V$. Since $V \subseteq N$, the open set U is a subset of $f^{-1}(N)$ containing a, and $f^{-1}(N)$ is a neighborhood of a. Conversely, if $f^{-1}(N)$ is a neighborhood of a whenever N is any neighborhood of $f(a)$, it is immediate that f is continuous at a. See Figure A.17.

It is easy to see that Definition A.16 is equivalent to the following statements.

Proposition A.9. *Let $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ be topological spaces, and let $f: E \to F$ be a function. For every $a \in E$, the function f is continuous at $a \in E$ iff for every neighborhood N of $f(a) \in F$, then $f^{-1}(N)$ is a neighborhood of a. The function f is continuous on E iff $f^{-1}(V)$ is an open set in $\mathcal{O}_E$ for every open set $V \in \mathcal{O}_F$.*

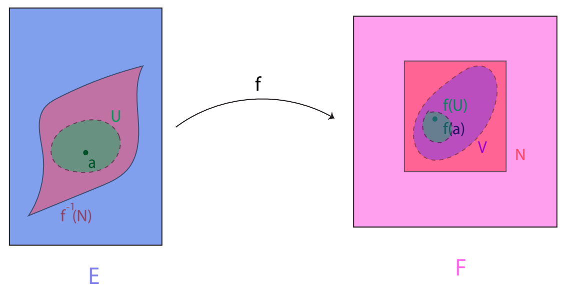


Figure A.17: A schematic illustration of the neighborhood condition.

If E and F are metric spaces defined by metrics d_E and d_F, we can show easily that f is continuous at a iff

for every $\epsilon > 0$, there is some $\eta > 0$, such that, for every $x \in E$,

$$\text{if } d_E(a, x) \leq \eta, \text{ then } d_F(f(a), f(x)) \leq \epsilon.$$

Similarly, if E and F are normed vector spaces defined by norms $\| \|_E$ and $\| \|_F$, we can show easily that f is continuous at a iff

for every $\epsilon > 0$, there is some $\eta > 0$, such that, for every $x \in E$,

$$\text{if } \|x - a\|_E \leq \eta, \text{ then } \|f(x) - f(a)\|_F \leq \epsilon.$$

It is worth noting that continuity is a topological notion, in the sense that equivalent metrics (or equivalent norms) define exactly the same notion of continuity.

Definition A.17. If $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ are topological spaces, and $f: E \to F$ is a function, for every nonempty subset $A \subseteq E$ of E, we say that f *is continuous on* A if the restriction of f to A is continuous with respect to $(A, \mathcal{U})$ and $(F, \mathcal{O}_F)$, where $\mathcal{U}$ is the subspace topology induced by $\mathcal{O}_E$ on A.

Given a product $E_1 \times \cdots \times E_n$ of topological spaces, as usual, we let $\pi_i: E_1 \times \cdots \times E_n \to E_i$ be the projection function such that, $\pi_i(x_1, \ldots, x_n) = x_i$. It is immediately verified that each π_i is continuous.

Given a topological space $(E, \mathcal{O})$, we say that a point $a \in E$ is *isolated* if $\{a\}$ is an open set in $\mathcal{O}$. Then if $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ are topological spaces, any function $f: E \to F$ is continuous at every isolated point $a \in E$. In the discrete topology, every point is isolated.

In a nontrivial normed vector space $(E, \| \|)$ (with $E \neq \{0\}$), no point is isolated. To show this, we show that every open ball $B_0(u, \rho,)$ contains some vectors different from u.

Indeed, since E is nontrivial, there is some $v \in E$ such that $v \neq 0$, and thus $\lambda = \|v\| > 0$ (by (N1)). Let

$$w = u + \frac{\rho}{\lambda + 1} v.$$

Since $v \neq 0$ and $\rho > 0$, we have $w \neq u$. Then,

$$\|w - u\| = \left\| \frac{\rho}{\lambda + 1} v \right\| = \frac{\rho\lambda}{\lambda + 1} < \rho,$$

which shows that $\|w - u\| < \rho$, for $w \neq u$.

The following proposition is easily shown.

Proposition A.10. *Given topological spaces $(E, \mathcal{O}_E)$, $(F, \mathcal{O}_F)$, and $(G, \mathcal{O}_G)$, and two functions $f\colon E \to F$ and $g\colon F \to G$, if f is continuous at $a \in E$ and g is continuous at $f(a) \in F$, then $g \circ f\colon E \to G$ is continuous at $a \in E$. Given n topological spaces $(F_i, \mathcal{O}_i)$, for every function $f\colon E \to F_1 \times \cdots \times F_n$, then f is continuous at $a \in E$ iff every $f_i\colon E \to F_i$ is continuous at a, where $f_i = \pi_i \circ f$.*

One can also show that in a metric space (E, d), the distance $d\colon E \times E \to \mathbb{R}$ is continuous, where $E \times E$ has the product topology. By the triangle inequality, we have

$$d(x, y) \leq d(x, x_0) + d(x_0, y_0) + d(y_0, y) = d(x_0, y_0) + d(x_0, x) + d(y_0, y)$$

and

$$d(x_0, y_0) \leq d(x_0, x) + d(x, y) + d(y, y_0) = d(x, y) + d(x_0, x) + d(y_0, y).$$

Consequently,

$$|d(x, y) - d(x_0, y_0)| \leq d(x_0, x) + d(y_0, y),$$

which proves that d is continuous at (x_0, y_0). In fact this shows that d is uniformly continuous; see Definition A.45.

Given any nonempty subset A of E, by Proposition A.2, the map $x \mapsto d(x, A)$ is continuous (in fact, uniformy continuous).

Similarly, for a normed vector space $(E, \| \, \|)$, the norm $\| \, \|\colon E \to \mathbb{R}$ is (uniformly) continuous.

Given a function $f\colon E_1 \times \cdots \times E_n \to F$, we can fix $n - 1$ of the arguments, say $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$, and view f as a function of the remaining argument,

$$x_i \mapsto f(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n),$$

where $x_i \in E_i$. If f is continuous, it is clear that each f_i is continuous.

One should be careful that the converse is false! For example, consider the function $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, defined such that,

$$f(x, y) = \frac{xy}{x^2 + y^2} \quad \text{if } (x, y) \neq (0, 0), \quad \text{and} \quad f(0, 0) = 0.$$

The function f is continuous on $\mathbb{R} \times \mathbb{R} - \{(0, 0)\}$, but on the line $y = mx$, with $m \neq 0$, we have $f(x, y) = \frac{m}{1+m^2} \neq 0$, and thus, on this line, $f(x, y)$ does not approach 0 when (x, y) approaches $(0, 0)$. See Figure A.18.

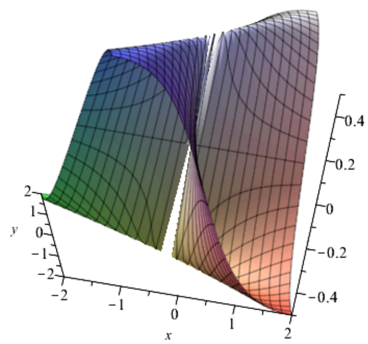


Figure A.18: The graph of $f(x, y) = \frac{xy}{x^2+y^2}$ for $(x, y) \neq (0, 0)$. The bottom of this graph, which shows the approach along the line $y = -x$, does not have a z value of 0.

The following proposition is useful for showing that real-valued functions are continuous.

Proposition A.11. *If E is a topological space, and $(\mathbb{R}, |x - y|)$ the reals under the standard topology, for any two functions $f\colon E \to \mathbb{R}$ and $g\colon E \to \mathbb{R}$, for any $a \in E$, for any $\lambda \in \mathbb{R}$, if f and g are continuous at a, then $f + g$, λf, $f \cdot g$, are continuous at a, and f/g is continuous at a if $g(a) \neq 0$.*

Proof. Left as an exercise. $\square$

Using Proposition A.11, we can show easily that every real polynomial function is continuous.

The notion of isomorphism of topological spaces is defined as follows.

Definition A.18. Let $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ be topological spaces, and let $f\colon E \to F$ be a function. We say that f *is a homeomorphism between E and F* if f is bijective, and both $f\colon E \to F$ and $f^{-1}\colon F \to E$ are continuous.

One should be careful that a bijective continuous function $f: E \to F$ is not necessarily a homeomorphism. For example, if $E = \mathbb{R}$ with the discrete topology, and $F = \mathbb{R}$ with the standard topology, the identity is not a homeomorphism. Another interesting example involving a parametric curve is given below. Let $L: \mathbb{R} \to \mathbb{R}^2$ be the function, defined such that,

$$L_1(t) = \frac{t(1+t^2)}{1+t^4},$$
$$L_2(t) = \frac{t(1-t^2)}{1+t^4}.$$

If we think of $(x(t), y(t)) = (L_1(t), L_2(t))$ as a geometric point in $\mathbb{R}^2$, the set of points $(x(t), y(t))$ obtained by letting t vary in $\mathbb{R}$ from $-\infty$ to $+\infty$, defines a curve having the shape of a "figure eight," with self-intersection at the origin, called the "lemniscate of Bernoulli." See Figure A.19. The map L is continuous, and in fact bijective, but its inverse L^{-1} is not continuous. Indeed, when we approach the origin on the branch of the curve in the upper left quadrant (i.e., points such that, $x \leq 0$, $y \geq 0$), then t goes to $-\infty$, and when we approach the origin on the branch of the curve in the lower right quadrant (i.e., points such that, $x \geq 0$, $y \leq 0$), then t goes to $+\infty$.

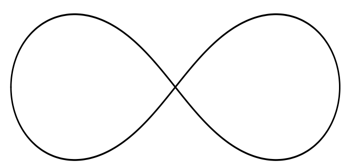


Figure A.19: The lemniscate of Bernoulli.

We also review the concept of limit of a sequence. Given any set E, a *sequence* is any function $x: \mathbb{N} \to E$, usually denoted by $(x_n)_{n \in \mathbb{N}}$, or $(x_n)_{n \geq 0}$, or even by (x_n).

Definition A.19. Given a topological space $(E, \mathcal{O})$, we say that *a sequence* $(x_n)_{n \in \mathbb{N}}$ *converges to some* $a \in E$ if for every open set U containing a, there is some $n_0 \geq 0$, such that, $x_n \in U$, for all $n \geq n_0$. We also say that *a is a limit of* $(x_n)_{n \in \mathbb{N}}$. See Figure A.20.

When E is a metric space with metric d, it is easy to show that this is equivalent to the fact that,

for every $\epsilon > 0$, there is some $n_0 \geq 0$, such that, $d(x_n, a) \leq \epsilon$, for all $n \geq n_0$.

When E is a normed vector space with norm $\| \, \|$, it is easy to show that this is equivalent to the fact that,

for every $\epsilon > 0$, there is some $n_0 \geq 0$, such that, $\|x_n - a\| \leq \epsilon$, for all $n \geq n_0$.

The following proposition shows the importance of the Hausdorff separation axiom.

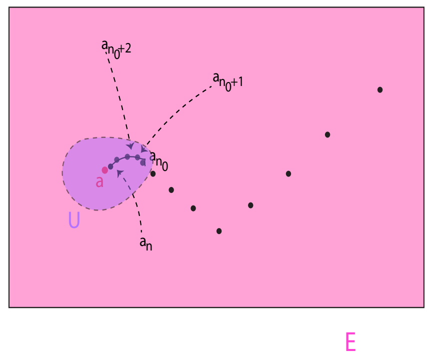


Figure A.20: A schematic illustration of Definition A.19.

Proposition A.12. *Given a topological space $(E, \mathcal{O})$, if the Hausdorff separation axiom holds, then every sequence has at most one limit.*

Proof. Left as an exercise. ☐

It is worth noting that the notion of limit is topological, in the sense that a sequence converge to a limit b iff it converges to the same limit b in any equivalent metric (and similarly for equivalent norms).

If E is a metric space and if A is a subset of E, there is a convenient way of showing that a point $x \in E$ belongs to the closure $\overline{A}$ of A in terms of sequences.

Proposition A.13. *Given any metric space (E, d), for any subset A of E and any point $x \in E$, we have $x \in \overline{A}$ iff there is a sequence (a_n) of points $a_n \in A$ converging to x.*

Proof. If the sequence (a_n) of points $a_n \in A$ converges to x, then for every open subset U of E containing x, there is some n_0 such that $a_n \in U$ for all $n \geq n_0$, so $U \cap A \neq \emptyset$, and Proposition A.4 implies that $x \in \overline{A}$.

Conversely, assume that $x \in \overline{A}$. Then for every $n \geq 1$, consider the open ball $B_0(x, 1/n)$. By Proposition A.4, we have $B_0(x, 1/n) \cap A \neq \emptyset$, so we can pick some $a_n \in B_0(x, 1/n) \cap A$. This, way, we define a sequence (a_n) of points in A, and by construction $d(x, a_n) < 1/n$ for all $n \geq 1$, so the sequence (a_n) converges to x. ☐

We still need one more concept of limit for functions.

Definition A.20. Let $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ be topological spaces, let A be some nonempty subset of E, and let $f \colon A \to F$ be a function. For any $a \in \overline{A}$ and any $b \in F$, we say that $f(x)$ *approaches b as x approaches a with values in A* if for every open set $V \in \mathcal{O}_F$ containing b, there is some open set $U \in \mathcal{O}_E$ containing a, such that, $f(U \cap A) \subseteq V$. See Figure A.21. This is denoted by

$$\lim_{x \to a, x \in A} f(x) = b.$$

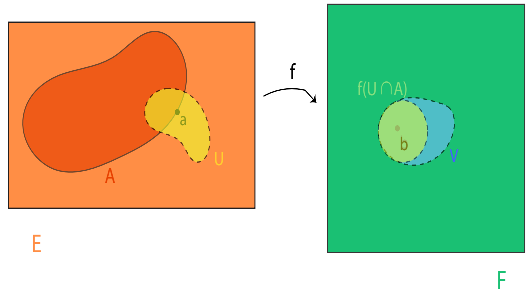


Figure A.21: A schematic illustration of Definition A.20.

First, note that by Proposition A.4, since $a \in \overline{A}$, for every open set U containing a, we have $U \cap A \neq \emptyset$, and the definition is nontrivial. Also, even if $a \in A$, the value $f(a)$ of f at a plays no role in this definition. When E and F are metric space with metrics d_E and d_F, it can be shown easily that the definition can be stated as follows:

For every $\epsilon > 0$, there is some $\eta > 0$, such that, for every $x \in A$,

$$\text{if } d_E(x, a) \leq \eta, \text{ then } d_F(f(x), b) \leq \epsilon.$$

When E and F are normed vector spaces with norms $\| \|_E$ and $\| \|_F$, it can be shown easily that the definition can be stated as follows:

For every $\epsilon > 0$, there is some $\eta > 0$, such that, for every $x \in A$,

$$\text{if } \|x - a\|_E \leq \eta, \text{ then } \|f(x) - b\|_F \leq \epsilon.$$

We have the following result relating continuity at a point and the previous notion.

Proposition A.14. *Let $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ be two topological spaces, and let $f \colon E \to F$ be a function. For any $a \in E$, the function f is continuous at a iff $f(x)$ approaches $f(a)$ when x approaches a (with values in E).*

Proof. Left as a trivial exercise. □

Another important proposition relating the notion of convergence of a sequence to continuity, is stated without proof.

Proposition A.15. *Let $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ be two topological spaces, and let $f \colon E \to F$ be a function.*

(1) If f is continuous, then for every sequence $(x_n)_{n \in \mathbb{N}}$ in E, if (x_n) converges to a, then $(f(x_n))$ converges to $f(a)$.

(2) If E is a metric space, and $(f(x_n))$ converges to $f(a)$ whenever (x_n) converges to a, for every sequence $(x_n)_{n\in\mathbb{N}}$ in E, then f is continuous.

A special case of Definition A.20 will be used when E and F are (nontrivial) normed vector spaces with norms $\| \|_E$ and $\| \|_F$. Let U be any nonempty open subset of E. We showed earlier that E has no isolated points and that every set $\{v\}$ is closed, for every $v \in E$. Since E is nontrivial, for every $v \in U$, there is a nontrivial open ball contained in U (an open ball not reduced to its center). Then, for every $v \in U$, $A = U - \{v\}$ is open and nonempty, and clearly, $v \in \overline{A}$. For any $v \in U$, if $f(x)$ approaches b when x approaches v with values in $A = U - \{v\}$, we say that $f(x)$ *approaches b when x approaches v with values $\neq v$ in U*. This is denoted by

$$\lim_{x\to v, x\in U, x\neq v} f(x) = b.$$

Remark: Variations of the above case show up in the following case: $E = \mathbb{R}$, and F is some arbitrary topological space. Let A be some nonempty subset of $\mathbb{R}$, and let $f\colon A \to F$ be some function. For any $a \in A$, we say that f *is continuous on the right at a* if

$$\lim_{x\to a, x\in A\cap[a,+\infty)} f(x) = f(a).$$

We can define *continuity on the left* at a in a similar fashion.

Let us consider another variation. Let A be some nonempty subset of $\mathbb{R}$, and let $f\colon A \to F$ be some function. For any $a \in A$, we say that f *has a discontinuity of the first kind at a* if

$$\lim_{x\to a, x\in A\cap\,(-\infty,a)} f(x) = f(a_-)$$

and

$$\lim_{x\to a, x\in A\cap\,(a,+\infty)} f(x) = f(a_+)$$

both exist, and either $f(a_-) \neq f(a)$, or $f(a_+) \neq f(a)$.

Note that it is possible that $f(a_-) = f(a_+)$, but f is still discontinuous at a if this common value differs from $f(a)$. Functions defined on a nonempty subset of $\mathbb{R}$, and that are continuous, except for some points of discontinuity of the first kind, play an important role in analysis.

We now turn to connectivity properties of topological spaces.

A.4 Connected Sets

Connectivity properties of topological spaces play a very important role in understanding the topology of surfaces. This section gathers the facts needed to have a good understanding of the classification theorem for compact surfaces (with boundary). The main references are Ahlfors and Sario [1] and Massey [70, 71]. For general background on topology, geometry, and algebraic topology, we also highly recommend Bredon [15] and Fulton [35].

Definition A.21. A topological space $(E, \mathcal{O})$ is *connected* if the only subsets of E that are both open and closed are the empty set and E itself. Equivalently, $(E, \mathcal{O})$ is connected if E cannot be written as the union $E = U \cup V$ of two disjoint nonempty open sets U, V, or if E cannot be written as the union $E = U \cup V$ of two disjoint nonempty closed sets. A subset, $S \subseteq E$, is *connected* if it is connected in the subspace topology on S induced by $(E, \mathcal{O})$. See Figure A.22. A connected open set is called a *region* and a closed set is a *closed region* if its interior is a connected (open) set.

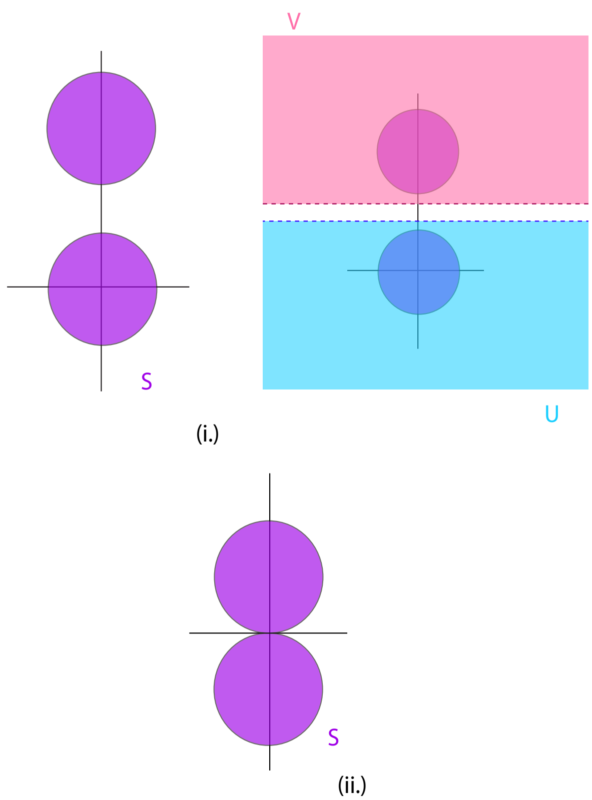


Figure A.22: Figure (i) shows that the union of two disjoint disks in $\mathbb{R}^2$ is a disconnected set since each circle can be separated by open half regions. Figure (ii) is an example of a connected subset of $\mathbb{R}^2$ since the two disks can not separated by open sets.

The definition of connectivity is meant to capture the fact that a connected space S is "one piece." Given the metric space $(\mathbb{R}^n, \| \, \|_2)$, the quintessential examples of connected spaces are $B_0(a, \rho)$ and $B(a, \rho)$. In particular, the following standard proposition characterizing the connected subsets of $\mathbb{R}$ can be found in most topology texts (for example, Munkres [75], Schwartz [84]). For the sake of completeness, we give a proof.

Proposition A.16. *A subset of the real line, $\mathbb{R}$, is connected iff it is an interval, i.e., of the form $[a, b]$, $(a, b]$, where $a = -\infty$ is possible, $[a, b)$, where $b = +\infty$ is possible, or (a, b),*

where $a = -\infty$ or $b = +\infty$ is possible.

Proof. Assume that A is a connected nonempty subset of $\mathbb{R}$. The cases where $A = \emptyset$ or A consists of a single point are trivial. Otherwise, we show that whenever $a, b \in A$, $a < b$, then the entire interval $[a, b]$ is a subset of A. Indeed, if this was not the case, there would be some $c \in (a, b)$ such that $c \notin A$, and then we could write $A = ((-\infty, c) \cap A) \cup ((c, +\infty) \cap A)$, where $(-\infty, c) \cap A$ and $(c, +\infty) \cap A$ are nonempty and disjoint open subsets of A, contradicting the fact that A is connected. It follows easily that A must be an interval.

Conversely, we show that an interval, I, must be connected. Let A be any nonempty subset of I which is both open and closed in I. We show that $I = A$. Fix any $x \in A$ and consider the set, R_x, of all y such that $[x, y] \subseteq A$. If the set R_x is unbounded, then $R_x = [x, +\infty)$. Otherwise, if this set is bounded, let b be its least upper bound. We claim that b is the right boundary of the interval I. Because A is closed in I, unless I is open on the right and b is its right boundary, we must have $b \in A$. In the first case, $A \cap [x, b) = I \cap [x, b) = [x, b)$. In the second case, because A is also open in I, unless b is the right boundary of the interval I (closed on the right), there is some open set $(b - \eta, b + \eta)$ contained in A, which implies that $[x, b + \eta/2] \subseteq A$, contradicting the fact that b is the least upper bound of the set R_x. Thus, b must be the right boundary of the interval I (closed on the right). A similar argument applies to the set, L_y, of all x such that $[x, y] \subseteq A$ and either L_y is unbounded, or its greatest lower bound a is the left boundary of I (open or closed on the left). In all cases, we showed that $A = I$, and the interval must be connected. $\square$

Intuitively, if a space is not connected, it is possible to define a continuous function which is constant on disjoint "connected components" and which takes possibly distinct values on disjoint components. This can be stated in terms of the concept of a locally constant function.

Definition A.22. Given two topological spaces X, Y, a function $f\colon X \to Y$ is *locally constant* if for every $x \in X$, there is an open set $U \subseteq X$ such that $x \in U$ and f is constant on U.

We claim that a locally constant function is continuous. In fact, we will prove that $f^{-1}(V)$ is open for every subset, $V \subseteq Y$ (not just for an open set V). It is enough to show that $f^{-1}(y)$ is open for every $y \in Y$, since for every subset $V \subseteq Y$,

$$f^{-1}(V) = \bigcup_{y \in V} f^{-1}(y),$$

and open sets are closed under arbitrary unions. However, either $f^{-1}(y) = \emptyset$ if $y \in Y - f(X)$ or f is constant on $U = f^{-1}(y)$ if $y \in f(X)$ (with value y), and since f is locally constant, for every $x \in U$, there is some open set, $W \subseteq X$, such that $x \in W$ and f is constant on W, which implies that $f(w) = y$ for all $w \in W$ and thus, that $W \subseteq U$, showing that U is a union of open sets and thus, is open. The following proposition shows that a space is connected iff every locally constant function is constant:

Proposition A.17. *A topological space is connected iff every locally constant function is constant. See Figure A.23.*

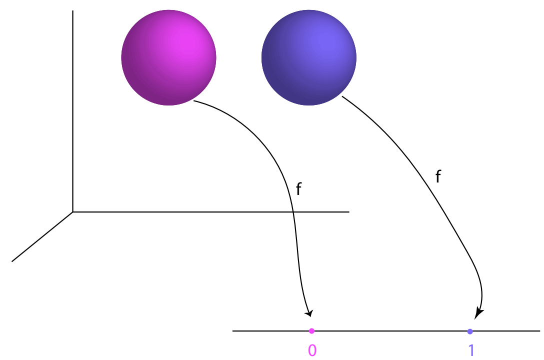


Figure A.23: An example of a locally constant, but not constant, real-valued function f over the disconnected set consisting of the disjoint union of the two solid balls. On the pink ball, f is 0, while on the purple ball, f is 1.

Proof. First, assume that X is connected. Let $f\colon X \to Y$ be a locally constant function to some space Y and assume that f is not constant. Pick any $y \in f(X)$. Since f is not constant, $U_1 = f^{-1}(y) \neq X$, and of course, $U_1 \neq \emptyset$. We proved just before Proposition A.17 that $f^{-1}(V)$ is open for every subset $V \subseteq Y$, and thus $U_1 = f^{-1}(y) = f^{-1}(\{y\})$ and $U_2 = f^{-1}(Y - \{y\})$ are both open, nonempty, and clearly $X = U_1 \cup U_2$ and U_1 and U_2 are disjoint. This contradicts the fact that X is connected and f must be constant.

Assume that every locally constant function $f\colon X \to Y$ is constant. If X is not connected, we can write $X = U_1 \cup U_2$, where both U_1, U_2 are open, disjoint, and nonempty. We can define the function, $f\colon X \to \mathbb{R}$, such that $f(x) = 1$ on U_1 and $f(x) = 0$ on U_2. Since U_1 and U_2 are open, the function f is locally constant, and yet not constant, a contradiction. $\square$

A characterization on the connected subsets of $\mathbb{R}^n$ is harder and requires the notion of arcwise connectedness. One of the most important properties of connected sets is that they are preserved by continuous maps.

Proposition A.18. *Given any continuous map, $f\colon E \to F$, if $A \subseteq E$ is connected, then $f(A)$ is connected.*

Proof. If $f(A)$ is not connected, then there exist some nonempty open sets, U, V, in F such that $f(A) \cap U$ and $f(A) \cap V$ are nonempty and disjoint, and

$$f(A) = (f(A) \cap U) \cup (f(A) \cap V).$$

Then, $f^{-1}(U)$ and $f^{-1}(V)$ are nonempty and open since f is continuous and

$$A = (A \cap f^{-1}(U)) \cup (A \cap f^{-1}(V)),$$

with $A \cap f^{-1}(U)$ and $A \cap f^{-1}(V)$ nonempty, disjoint, and open in A, contradicting the fact that A is connected. $\square$

An important corollary of Proposition A.18 is that for every continuous function, $f\colon E \to \mathbb{R}$, where E is a connected space, $f(E)$ is an interval. Indeed, this follows from Proposition A.16. Thus, if f takes the values a and b where $a < b$, then f takes all values $c \in [a, b]$. This is a very important property known as the intermediate value theorem.

Even if a topological space is not connected, it turns out that it is the disjoint union of maximal connected subsets and these connected components are closed in E. In order to obtain this result, we need a few lemmas.

Lemma A.19. *Given a topological space, E, for any family, $(A_i)_{i\in I}$, of (nonempty) connected subsets of E, if $A_i \cap A_j \neq \emptyset$ for all $i, j \in I$, then the union, $A = \bigcup_{i\in I} A_i$, of the family, $(A_i)_{i\in I}$, is also connected.*

Proof. Assume that $\bigcup_{i\in I} A_i$ is not connected. There exists two nonempty open subsets, U and V, of E such that $A \cap U$ and $A \cap V$ are disjoint and nonempty and such that

$$A = (A \cap U) \cup (A \cap V).$$

Now, for every $i \in I$, we can write

$$A_i = (A_i \cap U) \cup (A_i \cap V),$$

where $A_i \cap U$ and $A_i \cap V$ are disjoint, since $A_i \subseteq A$ and $A \cap U$ and $A \cap V$ are disjoint. Since A_i is connected, either $A_i \cap U = \emptyset$ or $A_i \cap V = \emptyset$. This implies that either $A_i \subseteq A \cap U$ or $A_i \subseteq A \cap V$. However, by assumption, $A_i \cap A_j \neq \emptyset$, for all $i, j \in I$, and thus, either both $A_i \subseteq A \cap U$ and $A_j \subseteq A \cap U$, or both $A_i \subseteq A \cap V$ and $A_j \subseteq A \cap V$, since $A \cap U$ and $A \cap V$ are disjoint. Thus, we conclude that either $A_i \subseteq A \cap U$ for all $i \in I$, or $A_i \subseteq A \cap V$ for all $i \in I$. But this proves that either

$$A = \bigcup_{i\in I} A_i \subseteq A \cap U,$$

or

$$A = \bigcup_{i\in I} A_i \subseteq A \cap V,$$

contradicting the fact that both $A \cap U$ and $A \cap V$ are disjoint and nonempty. Thus, A must be connected. $\square$

In particular, the above lemma applies when the connected sets in a family $(A_i)_{i\in I}$ have a point in common.

Lemma A.20. *If A is a connected subset of a topological space, E, then for every subset, B, such that $A \subseteq B \subseteq \overline{A}$, where $\overline{A}$ is the closure of A in E, the set B is connected.*

Proof. If B is not connected, then there are two nonempty open subsets, U, V, of E such that $B \cap U$ and $B \cap V$ are disjoint and nonempty, and

$$B = (B \cap U) \cup (B \cap V).$$

Since $A \subseteq B$, the above implies that

$$A = (A \cap U) \cup (A \cap V),$$

and since A is connected, either $A \cap U = \emptyset$, or $A \cap V = \emptyset$. Without loss of generality, assume that $A \cap V = \emptyset$, which implies that $A \subseteq A \cap U \subseteq B \cap U$. However, $B \cap U$ is closed in the subspace topology for B and since $B \subseteq \overline{A}$ and $\overline{A}$ is closed in E, the closure of A in B w.r.t. the subspace topology of B is clearly $B \cap \overline{A} = B$, which implies that $B \subseteq B \cap U$ (since the closure is the smallest closed set containing the given set). Thus, $B \cap V = \emptyset$, a contradiction. $\square$

In particular, Lemma A.20 shows that if A is a connected subset, then its closure, $\overline{A}$, is also connected. We are now ready to introduce the connected components of a space.

Definition A.23. Given a topological space, $(E, \mathcal{O})$, we say that two points, $a, b \in E$, are *connected* if there is some connected subset, A, of E such that $a \in A$ and $b \in A$.

It is immediately verified that the relation "a and b are connected in E" is an equivalence relation. Only transitivity is not obvious, but it follows immediately as a special case of Lemma A.19. Thus, the above equivalence relation defines a partition of E into nonempty disjoint *connected components*. The following proposition is easily proved using Lemma A.19 and Lemma A.20:

Proposition A.21. *Given any topological space, E, for any $a \in E$, the connected component containing a is the largest connected set containing a. The connected components of E are closed.*

The notion of a locally connected space is also useful.

Definition A.24. A topological space, $(E, \mathcal{O})$, is *locally connected* if for every $a \in E$, for every neighborhood, V, of a, there is a connected neighborhood, U, of a such that $U \subseteq V$. See Figure A.24.

As we shall see in a moment, it would be equivalent to require that E has a basis of connected open sets.

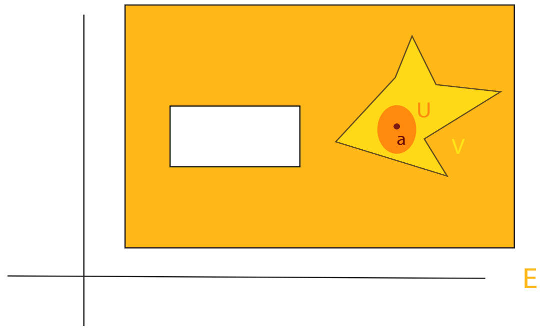


Figure A.24: The topological space E, which is homeomorphic to an annulus, is locally connected since each point is surrounded by a small disk contained in E.

There are connected spaces that are not locally connected and there are locally connected spaces that are not connected. The two properties are independent. For example, the subspace of $\mathbb{R}^2$ $S = \{(x, \sin(1/x)), \mid x > 0\} \cup \{(0, y) \mid -1 \leq y \leq 1\}$ is connected but not locally connected. See Figure A.25. The subspace S of $\mathbb{R}$ consisting $[0, 1] \cup [2, 3]$ is locally connected but not connected.

Proposition A.22. *A topological space, E, is locally connected iff for every open subset, A, of E, the connected components of A are open.*

Proof. Assume that E is locally connected. Let A be any open subset of E and let C be one of the connected components of A. For any $a \in C \subseteq A$, there is some connected neighborhood, U, of a such that $U \subseteq A$ and since C is a connected component of A containing a, we must have $U \subseteq C$. This shows that for every $a \in C$, there is some open subset containing a contained in C, so C is open.

Conversely, assume that for every open subset, A, of E, the connected components of A are open. Then, for every $a \in E$ and every neighborhood, U, of a, since U contains some open set A containing a, the interior, $\overset{\circ}{U}$, of U is an open set containing a and its connected components are open. In particular, the connected component C containing a is a connected open set containing a and contained in U. $\square$

Proposition A.22 shows that in a locally connected space, the connected open sets form a basis for the topology. It is easily seen that $\mathbb{R}^n$ is locally connected. Another very important property of surfaces and more generally, manifolds, is to be arcwise connected. The intuition is that any two points can be joined by a continuous arc of curve. This is formalized as follows.

Definition A.25. Given a topological space, $(E, \mathcal{O})$, an *arc (or path)* is a continuous map, $\gamma\colon [a, b] \to E$, where $[a, b]$ is a closed interval of the real line, $\mathbb{R}$. The point $\gamma(a)$ is the *initial*

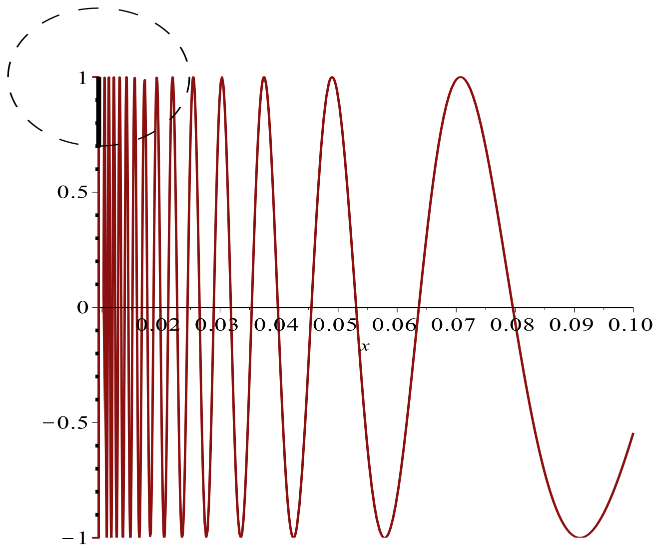


Figure A.25: Let S be the graph of $f(x) = \sin(1/x)$ union the y-axis between -1 and 1. This space is connected, but not locally connected.

point of the arc and the point $\gamma(b)$ is the *terminal point* of the arc. We say that γ *is an arc joining* $\gamma(a)$ *and* $\gamma(b)$. See Figure A.26. An arc is a *closed curve* if $\gamma(a) = \gamma(b)$. The set $\gamma([a, b])$ is the *trace* of the arc γ.

Typically, $a = 0$ and $b = 1$.

One should not confuse an arc, $\gamma\colon [a, b] \to E$, with its trace. For example, γ could be constant, and thus, its trace reduced to a single point.

An arc is a *Jordan arc* if γ is a homeomorphism onto its trace. An arc, $\gamma\colon [a, b] \to E$, is a *Jordan curve* if $\gamma(a) = \gamma(b)$ and γ is injective on $[a, b)$. Since $[a, b]$ is connected, by Proposition A.18, the trace $\gamma([a, b])$ of an arc is a connected subset of E.

Given two arcs $\gamma\colon [0, 1] \to E$ and $\delta\colon [0, 1] \to E$ such that $\gamma(1) = \delta(0)$, we can form a new arc defined as follows:

Definition A.26. Given two arcs, $\gamma\colon [0, 1] \to E$ and $\delta\colon [0, 1] \to E$, such that $\gamma(1) = \delta(0)$, we can form their *composition (or product)*, $\gamma\delta$,, defined such that

$$\gamma\delta(t) = \begin{cases} \gamma(2t) & \text{if } 0 \leq t \leq 1/2; \\ \delta(2t - 1) & \text{if } 1/2 \leq t \leq 1. \end{cases}$$

The *inverse*, γ^{-1}, *of the arc*, γ, is the arc defined such that $\gamma^{-1}(t) = \gamma(1-t)$, for all $t \in [0, 1]$.

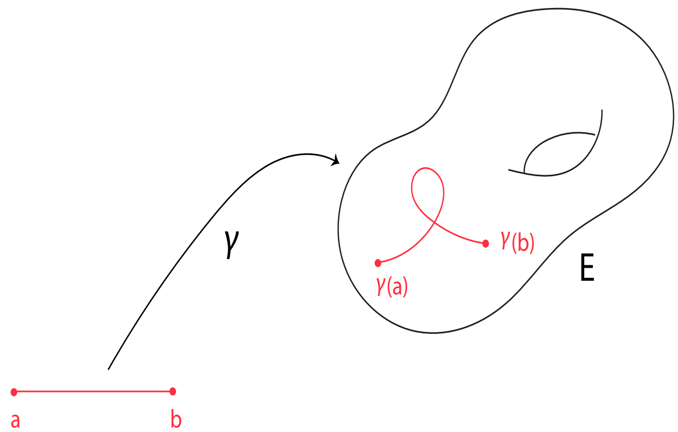


Figure A.26: Let E be the torus with subspace topology induced from $\mathbb{R}^3$ with red arc $\gamma([a,b])$. The torus is both arcwise connected and locally arcwise connected.

It is trivially verified that Definition A.26 yields continuous arcs.

Definition A.27. A topological space, E, is *arcwise connected* if for any two points, $a, b \in E$, there is an arc, $\gamma\colon [0,1] \to E$, joining a and b, i.e., such that $\gamma(0) = a$ and $\gamma(1) = b$. A topological space, E, is *locally arcwise connected* if for every $a \in E$, for every neighborhood, V, of a, there is an arcwise connected neighborhood, U, of a such that $U \subseteq V$. See Figure A.26.

The space $\mathbb{R}^n$ is locally arcwise connected, since for any open ball, any two points in this ball are joined by a line segment. Manifolds and surfaces are also locally arcwise connected. Proposition A.18 also applies to arcwise connectedness (this is a simple exercise). The following theorem is crucial to the theory of manifolds and surfaces:

Theorem A.23. *If a topological space, E, is arcwise connected, then it is connected. If a topological space, E, is connected and locally arcwise connected, then E is arcwise connected.*

Proof. First, assume that E is arcwise connected. Pick any point, a, in E. Since E is arcwise connected, for every $b \in E$, there is a path, $\gamma_b\colon [0,1] \to E$, from a to b and so,

$$E = \bigcup_{b \in E} \gamma_b([0,1])$$

a union of connected subsets all containing a. By Lemma A.19, E is connected.

Now assume that E is connected and locally arcwise connected. For any point $a \in E$, let F_a be the set of all points, b, such that there is an arc, $\gamma_b\colon [0,1] \to E$, from a to b. Clearly, F_a contains a. We show that F_a is both open and closed. For any $b \in F_a$, since E is locally

arcwise connected, there is an arcwise connected neighborhood U containing b (because E is a neighborhood of b). Thus, b can be joined to every point $c \in U$ by an arc, and since by the definition of F_a, there is an arc from a to b, the composition of these two arcs yields an arc from a to c, which shows that $c \in F_a$. But then $U \subseteq F_a$ and thus, F_a is open. See Figure A.27 $(i.)$. Now assume that b is in the complement of F_a. As in the previous case, there is some arcwise connected neighborhood U containing b. Thus, every point $c \in U$ can be joined to b by an arc. If there was an arc joining a to c, we would get an arc from a to b, contradicting the fact that b is in the complement of F_a. Thus, every point $c \in U$ is in the complement of F_a, which shows that U is contained in the complement of F_a, and thus, that the the complement of F_a is open. See Figure A.27 $(ii.)$. Consequently, we have shown that F_a is both open and closed and since it is nonempty, we must have $E = F_a$, which shows that E is arcwise connected. $\square$

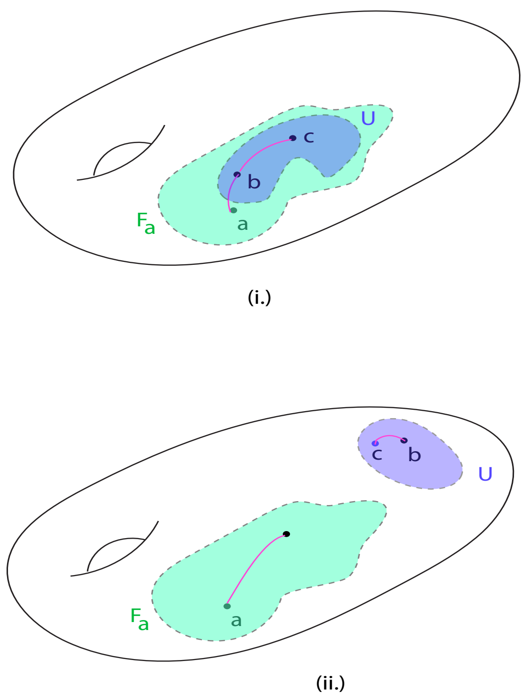


Figure A.27: Schematic illustrations of the proof techniques that show F_a is both open and closed.

If E is locally arcwise connected, the above argument shows that the connected components of E are arcwise connected.

It is not true that a connected space is arcwise connected. For example, the space consisting of the graph of the function

$$f(x) = \sin(1/x),$$

where $x > 0$, together with the portion of the y-axis, for which $-1 \leq y \leq 1$, is connected, but not arcwise connected. See Figure A.25.

A trivial modification of the proof of Theorem A.23 shows that in a normed vector space, E, a connected open set is arcwise connected by polygonal lines (i.e., arcs consisting of line segments). This is because in every open ball, any two points are connected by a line segment. Furthermore, if E is finite dimensional, these polygonal lines can be forced to be parallel to basis vectors.

We now consider compactness.

A.5 Compact Sets and Locally Compact Spaces

The property of compactness is very important in topology and analysis. We provide a quick review geared towards the study of manifolds, and for details, we refer the reader to Munkres [75], Schwartz [84]. In this section we will need to assume that the topological spaces are Hausdorff spaces. This is not a luxury, as many of the results are false otherwise.

We begin this section by providing the definition of compactness and describing a collection of compact spaces in $\mathbb{R}$. There are various equivalent ways of defining compactness. For our purposes, the most convenient way involves the notion of open cover.

Definition A.28. Given a topological space E, for any subset A of E, an *open cover* $(U_i)_{i\in I}$ *of* A is a family of open subsets of E such that $A \subseteq \bigcup_{i\in I} U_i$. An *open subcover* of an open cover $(U_i)_{i\in I}$ of A is any subfamily $(U_j)_{j\in J}$ which is an open cover of A, with $J \subseteq I$. An open cover $(U_i)_{i\in I}$ of A is *finite* if I is finite. See Figure A.28. The topological space E is *compact* if it is Hausdorff and for every open cover $(U_i)_{i\in I}$ of E, there is a finite open subcover $(U_j)_{j\in J}$ of E. Given any subset A of E, we say that A is *compact* if it is compact with respect to the subspace topology. We say that A is *relatively compact* if its closure $\overline{A}$ is compact.

It is immediately verified that a subset A of E is compact in the subspace topology relative to A iff for every open cover $(U_i)_{i\in I}$ of A by open subsets of E, there is a finite open subcover $(U_j)_{j\in J}$ of A. The property that every open cover contains a finite open subcover is often called the *Heine-Borel-Lebesgue* property. By considering complements, a Hausdorff space is compact iff for every family $(F_i)_{i\in I}$ of closed sets, if $\bigcap_{i\in I} F_i = \emptyset$, then $\bigcap_{j\in J} F_j = \emptyset$ for some finite subset J of I.

Definition A.28 requires that a compact space be Hausdorff. There are books in which a compact space is not necessarily required to be Hausdorff. Following Schwartz, we prefer calling such a space *quasi-compact*.

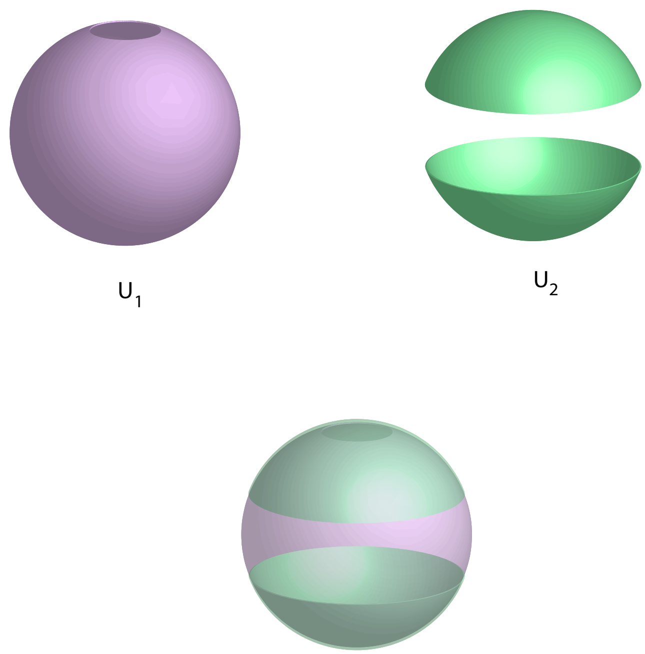


Figure A.28: An open cover of S^2 using two open sets induced by the Euclidean topology of $\mathbb{R}^3$.

Another equivalent and useful characterization can be given in terms of families having the finite intersection property.

Definition A.29. A family $(F_i)_{i\in I}$ of sets has the *finite intersection property* if $\bigcap_{j\in J} F_j \neq \emptyset$ for every finite subset J of I.

Proposition A.24. *A topological Hausdorff space E is compact iff for every family $(F_i)_{i\in I}$ of closed sets having the finite intersection property, then $\bigcap_{i\in I} F_i \neq \emptyset$.*

Proof. If E is compact and $(F_i)_{i\in I}$ is a family of closed sets having the finite intersection property, then $\bigcap_{i\in I} F_i$ cannot be empty, since otherwise we would have $\bigcap_{j\in J} F_j = \emptyset$ for some finite subset J of I, a contradiction. The converse is equally obvious. $\square$

Another useful consequence of compactness is as follows. For any family $(F_i)_{i\in I}$ of closed sets such that $F_{i+1} \subseteq F_i$ for all $i \in I$, if $\bigcap_{i\in I} F_i = \emptyset$, then $F_i = \emptyset$ for some $i \in I$. Indeed, there must be some finite subset J of I such that $\bigcap_{j\in J} F_j = \emptyset$, and since $F_{i+1} \subseteq F_i$ for all $i \in I$, we must have $F_j = \emptyset$ for the smallest F_j in $(F_j)_{j\in J}$. Using this fact, we note that $\mathbb{R}$ is *not* compact. Indeed, the family of closed sets, $([n, +\infty))_{n\geq 0}$, is decreasing and has an empty intersection.

It is immediately verified that every finite union of compact subsets is compact. Similarly, every finite union of relatively compact subsets is relatively compact (use the fact that $\overline{A \cup B} = \overline{A} \cap \overline{B}$).

Given a metric space, if we define a *bounded subset* to be a subset that can be enclosed in some closed ball (of finite radius), then any nonbounded subset of a metric space is not compact. However, a closed interval $[a, b]$ of the real line is compact.

Proposition A.25. *Every closed interval, $[a, b]$, of the real line is compact.*

Proof. We proceed by contradiction. Let $(U_i)_{i \in I}$ be any open cover of $[a, b]$ and assume that there is no finite open subcover. Let $c = (a + b)/2$. If both $[a, c]$ and $[c, b]$ had some finite open subcover, so would $[a, b]$, and thus, either $[a, c]$ does not have any finite subcover, or $[c, b]$ does not have any finite open subcover. Let $[a_1, b_1]$ be such a bad subinterval. The same argument applies and we split $[a_1, b_1]$ into two equal subintervals, one of which must be bad. Thus, having defined $[a_n, b_n]$ of length $(b - a)/2^n$ as an interval having no finite open subcover, splitting $[a_n, b_n]$ into two equal intervals, we know that at least one of the two has no finite open subcover and we denote such a bad interval by $[a_{n+1}, b_{n+1}]$. See Figure A.29. The sequence (a_n) is nondecreasing and bounded from above by b, and thus, by a fundamental property of the real line, it converges to its least upper bound, α. Similarly, the sequence (b_n) is nonincreasing and bounded from below by a and thus, it converges to its greatest lowest bound, β. Since $[a_n, b_n]$ has length $(b - a)/2^n$, we must have $\alpha = \beta$. However, the common limit $\alpha = \beta$ of the sequences (a_n) and (b_n) must belong to some open set, U_i, of the open cover and since U_i is open, it must contain some interval $[c, d]$ containing α. Then, because α is the common limit of the sequences (a_n) and (b_n), there is some N such that the intervals $[a_n, b_n]$ are all contained in the interval $[c, d]$ for all $n \geq N$, which contradicts the fact that none of the intervals $[a_n, b_n]$ has a finite open subcover. Thus, $[a, b]$ is indeed compact. $\square$

The argument of Proposition A.25 can be adapted to show that in $\mathbb{R}^m$, every closed set, $[a_1, b_1] \times \cdots \times [a_m, b_m]$, is compact. At every stage, we need to divide into 2^m subpieces instead of 2.

We next discuss some important properties of compact spaces. We begin with two separations axioms which only hold for Hausdorff spaces:

Proposition A.26. *Given a topological Hausdorff space, E, for every compact subset, A, and every point, b, not in A, there exist disjoint open sets, U and V, such that $A \subseteq U$ and $b \in V$. See Figure A.30. As a consequence, every compact subset is closed.*

Proof. Since E is Hausdorff, for every $a \in A$, there are some disjoint open sets, U_a and V_a, containing a and b respectively. Thus, the family, $(U_a)_{a \in A}$, forms an open cover of A. Since A is compact there is a finite open subcover, $(U_j)_{j \in J}$, of A, where $J \subseteq A$, and then $\bigcup_{j \in J} U_j$ is an open set containing A disjoint from the open set $\bigcap_{j \in J} V_j$ containing b. This shows that every point, b, in the complement of A belongs to some open set in this complement and thus, that the complement is open, i.e., that A is closed. See Figure A.31. $\square$

Actually, the proof of Proposition A.26 can be used to show the following useful property:

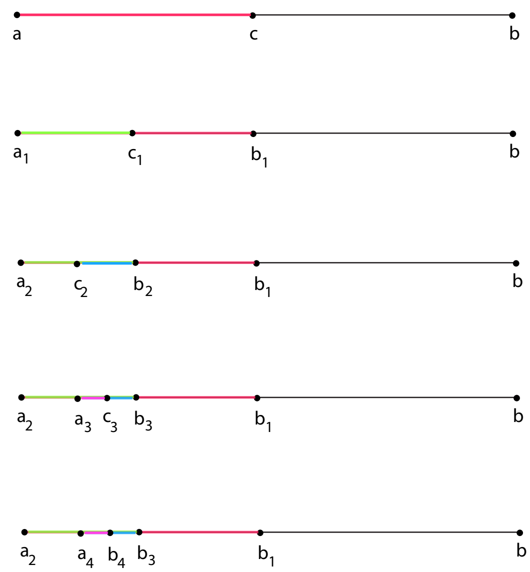


Figure A.29: The first four stages of the nested interval construction utilized in the proof of Proposition A.25.

Proposition A.27. *Given a topological Hausdorff space E, for every pair of compact disjoint subsets A and B, there exist disjoint open sets U and V, such that $A \subseteq U$ and $B \subseteq V$.*

Proof. We repeat the argument of Proposition A.26 with B playing the role of b and use Proposition A.26 to find disjoint open sets U_a containing $a \in A$, and V_a containing B. ☐

The following proposition shows that in a compact topological space, every closed set is compact:

Proposition A.28. *Given a compact topological space E, every closed set is compact.*

Proof. Since A is closed, $E - A$ is open and from any open cover, $(U_i)_{i \in I}$, of A, we can form an open cover of E by adding $E - A$ to $(U_i)_{i \in I}$ and, since E is compact, a finite subcover, $(U_j)_{j \in J} \cup \{E - A\}$, of E can be extracted such that $(U_j)_{j \in J}$ is a finite subcover of A. See Figure A.32. ☐

Remark: Proposition A.28 also holds for quasi-compact spaces, i.e., the Hausdorff separation property is not needed.

Putting Proposition A.27 and Proposition A.28 together, we note that if X is compact, then for every pair of disjoint closed sets A and B, there exist disjoint open sets U and V such that $A \subseteq U$ and $B \subseteq V$.

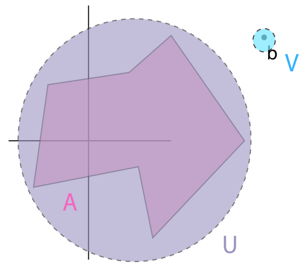


Figure A.30: The compact set of $\mathbb{R}^2$, A, is separated by any point in its complement.

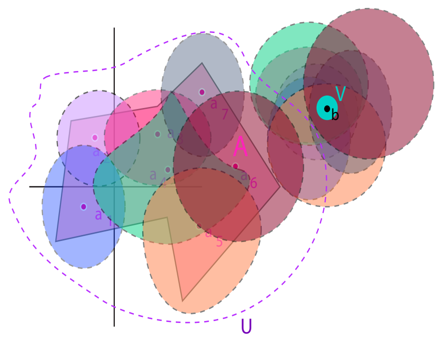


Figure A.31: For the pink compact set A, U is the union of the seven disks which cover A, while V is the intersection of the seven open sets containing b.

Definition A.30. A topological space E is *normal* if every one-point set is closed, and for every pair of disjoint closed sets A and B, there exist disjoint open sets U and V such that $A \subseteq U$ and $B \subseteq V$. A topological space E is *regular* if every one-point set is closed, and for every point $a \in E$ and every closed subset B of E, if $a \notin B$, then there exist disjoint open sets U and V such that $a \in U$ and $B \subseteq V$.

It is clear that a normal space is regular, and a regular space is Hausdorff. There are examples of Hausdorff spaces that are not regular, and of regular spaces that are not normal.

We just observed that a compact space is normal, and this is worth recording as a proposition.

Proposition A.29. *Every (Hausdorff) compact space is normal.*

An important property of metrizable spaces is that they are normal.

Proposition A.30. *Every metrizable space E is normal.*

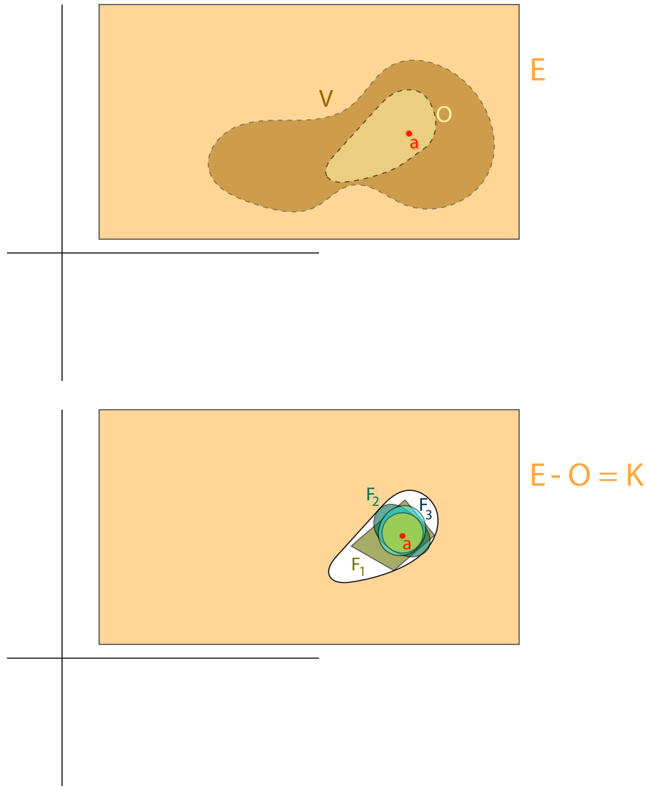


Figure A.32: An illustration of the proof of Proposition A.28. Both E and A are closed squares in $\mathbb{R}^2$. Note that an open cover of A, namely the green circles, when combined with the yellow square annulus $E - A$ covers all of the yellow square E.

Proof. Assume the topology of E is given by the metric d. Since B is closed and $A \cap B = \emptyset$, for every $a \in A$ since $a \notin \overline{B} = B$, there is some open ball $B_0(a, \epsilon_a)$ of radius $\epsilon_a > 0$ such that $B_0(a, \epsilon_a) \cap B = \emptyset$. Similarly, since A is closed and $A \cap B = \emptyset$, for every $b \in B$ there is some open ball $B_0(b, \epsilon_b)$ of radius $\epsilon_b > 0$ such that $B_0(b, \epsilon_b) \cap A = \emptyset$. Let

$$U = \bigcup_{a \in A} B_0(a, \epsilon_a/2), \quad V = \bigcup_{b \in B} B_0(b, \epsilon_b/2).$$

Then A and B are open sets such that $A \subseteq U$ and $B \subseteq V$, and we claim that $U \cap V = \emptyset$.

If not, then there is some $z \in U \cap V$, which implies that for some $a \in A$ and some $b \in B$, we have

$$z \in B_0(a, \epsilon_a/2) \cap B_0(b, \epsilon_b/2).$$

It follows that

$$d(a, b) \leq d(a, z) + d(z, b) < (\epsilon_a + \epsilon_b)/2.$$

If $\epsilon_a \leq \epsilon_b$,then $d(a, b) < \epsilon_b$, so $a \in B_0(b, \epsilon_b)$, contradicting the fact that $B_0(b, \epsilon_b) \cap A = \emptyset$. If $\epsilon_b \leq \epsilon_a$,then $d(a, b) < \epsilon_a$, so $b \in B_0(a, \epsilon_a)$, contradicting the fact that $B_0(a, \epsilon_a) \cap B = \emptyset$. $\square$

Normal spaces have a strong separation property regading disjoint closed subsets A and B. Actually, this separation property can be stated as the existence of a certain continuous function $f: E \to [0,1]$ taking the value 1 on A and the value 0 on B. This result is known as *Urysohn Lemma*. It is an important tool in topology and analysis.

Theorem A.31. *(Urysohn Lemma) Let E be a normal space. For any two closed disjoint subsets A and B, there is a continuous function $f: E \to [0,1]$ such that $f(x) = 1$ for all $x \in A$ and $f(x) = 0$ for all $x \in B$.*

A proof of Theorem A.31 can be found in Munkres [75] (Chapter 4, Section 33, Theorem 33.1). Theorem A.31 is one of the ingedients in the Urysohn metrization theorem (Theorem A.48).

Compact spaces also have the following property.

Proposition A.32. *Given a compact topological space, E, for every $a \in E$, for every neighborhood, V, of a, there exists a compact neighborhood, U, of a such that $U \subseteq V$. See Figure A.33.*

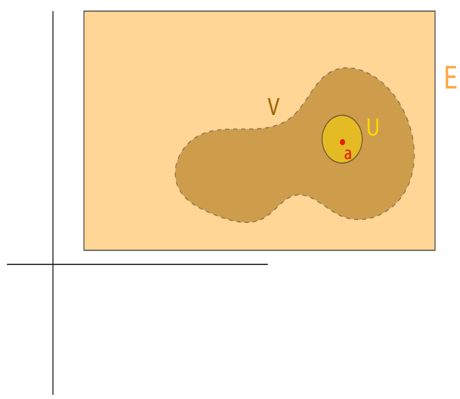


Figure A.33: Let E be the peach square of $\mathbb{R}^2$. Each point of E is contained in a compact neighborhood U, in this case the small closed yellow disk.

Proof. Since V is a neighborhood of a, there is some open subset, O, of V containing a. Then the complement, $K = E - O$, of O is closed and since E is compact, by Proposition A.28, K is compact. Now, if we consider the family of all closed sets of the form, $K \cap F$, where F is any closed neighborhood of a, since $a \notin K$, this family has an empty intersection and thus, there is a finite number of closed neighborhood, $F_1, \ldots, F_n$, of a, such that $K \cap F_1 \cap \cdots \cap F_n = \emptyset$. Then, $U = F_1 \cap \cdots \cap F_n$ is closed and hence by Proposition A.28, a compact neigborhood of a contained in $O \subseteq V$. See Figure A.34. $\square$

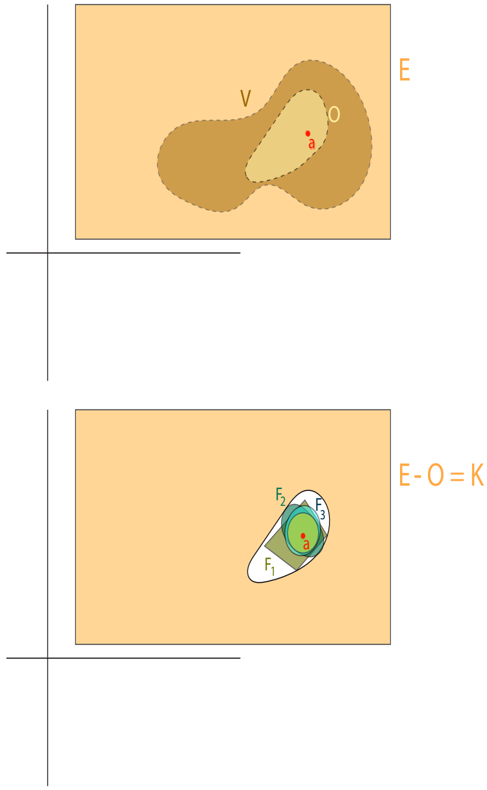


Figure A.34: Let E be the peach square of $\mathbb{R}^2$. The compact neighborhood of a, U, is the intersection of the closed sets F_1, F_2, F_3, each of which are contained in the complement of K.

It can be shown that in a normed vector space of finite dimension, a subset is compact iff it is closed and bounded. For $\mathbb{R}^n$ the proof is simple.

In a normed vector space of infinite dimension, there are closed and bounded sets that are not compact!

More could be said about compactness in metric spaces but we will only need the notion of Lebesgue number, which will be discussed a little later. Another crucial property of compactness is that it is preserved under continuity.

Proposition A.33. *Let E be a topological space and let F be a topological Hausdorff space. For every compact subset, A, of E, for every continuous map, $f\colon E \to F$, the subspace $f(A)$ is compact.*

Proof. Let $(U_i)_{i\in I}$ be an open cover of $f(A)$. We claim that $(f^{-1}(U_i))_{i\in I}$ is an open cover of A, which is easily checked. Since A is compact, there is a finite open subcover, $(f^{-1}(U_j))_{j\in J}$, of A, and thus, $(U_j)_{j\in J}$ is an open subcover of $f(A)$. $\square$

As a corollary of Proposition A.33, if E is compact, F is Hausdorff, and $f\colon E \to F$ is continuous and bijective, then f is a homeomorphism. Indeed, it is enough to show that f^{-1} is continuous, which is equivalent to showing that f maps closed sets to closed sets. However, closed sets are compact and Proposition A.33 shows that compact sets are mapped to compact sets, which, by Proposition A.26, are closed.

Another important corollary of Proposition A.33 is the following result.

Proposition A.34. *If E is a compact nonempty topological space and if $f\colon E \to \mathbb{R}$ is a continuous function, then there are points $a, b \in E$ such that $f(a)$ is the minimum of $f(E)$ and $f(b)$ is the maximum of $f(E)$.*

Proof. The set $f(E)$ is a compact subset of $\mathbb{R}$ and thus, a closed and bounded set which contains its greatest lower bound and its least upper bound. $\square$

The following property also holds.

Proposition A.35. *Let (E, d) be a metric space. For any nonempty subset A of E, if A is compact, then for every open subset U such that $A \subseteq U$, there is some $r > 0$ such that $V_r(A) \subseteq U$.*

Proof. The function $x \mapsto d(x, E - U)$ is continuous and $d(x, E - U) > 0$ for $x \in A$ (since $A \subseteq U$). By Proposition A.34, there is some $a \in A$ such that

$$d(a, E - U) = \inf_{x \in A} d(x, E - U).$$

But $d(a, E - U) = r > 0$, which implies that $V_r(A) \subseteq U$. $\square$

Another useful notion is that of local compactness. Indeed manifolds and surfaces are locally compact.

Definition A.31. A topological space E is *locally compact* if it is Hausdorff and for every $a \in E$, there is some compact neighborhood K of a. See Figure A.33.

From Proposition A.32, every compact space is locally compact but the converse is false. For example, $\mathbb{R}$ is locally compact but not compact. In fact it can be shown that a normed vector space of finite dimension is locally compact.

Proposition A.36. *Given a locally compact topological space, E, for every $a \in E$, for every neighborhood, N, of a, there exists a compact neighborhood, U, of a, such that $U \subseteq N$.*

Proof. For any $a \in E$, there is some compact neighborhood, V, of a. By Proposition A.32, every neigborhood of a relative to V contains some compact neighborhood U of a relative to V. But every neighborhood of a relative to V is a neighborhood of a relative to E and every neighborhood N of a in E yields a neighborhood, $V \cap N$, of a in V and thus, for every neighborhood, N, of a, there exists a compact neighborhood, U, of a such that $U \subseteq N$. $\square$

When E is a metric space, the subsets $V_r(A)$ defined in Definition A.6 have the following property.

Proposition A.37. *Let (E, d) be a metric space. If E is locally compact, then for any nonempty compact subset A of E, there is some $r > 0$ such that $\overline{V_r(A)}$ is compact.*

Proof. Since E is locally compact, for every $x \in A$, there is some compact subset V_x whose interior $\overset{\circ}{V}_x$ contains x. The family of open subsets $\overset{\circ}{V}_x$ is an open cover A, and since A is compact, it has a finite subcover $\{\overset{\circ}{V}_{x_1}, \ldots, \overset{\circ}{V}_{x_n}\}$. Then $U = V_{x_1} \cup \cdots \cup V_{x_n}$ is compact (as a finite union of compact subsets), and it contains an open subset containing A (the union of the $\overset{\circ}{V}_{x_i}$). By Proposition A.35, there is some $r > 0$ such that $V_r(A) \subseteq \overset{\circ}{U}$, and thus $\overline{V_r(A)} \subseteq U$. Since U is compact and $\overline{V_r(A)}$ is closed, $\overline{V_r(A)}$ is compact. $\square$

Another very important property of locally compact spaces is the Proposition A.39 below. This result implies the existence of continuous partitions of unity for a finite open cover of a compact subset. Such partitions of unity are used in proving that Radon functionals correspond to certain Borel measures. First we have the following proposition.

Proposition A.38. *Let E be a locally compact (Hausdorff) space. For every compact subset K and every open subset V, if $K \subseteq V$, then there is an open set W with compact closure such that $K \subseteq W \subseteq \overline{W} \subseteq V$.*

A proof of Proposition A.38 can be found in Rudin [79] (Chapter 2, Theorem 2.7). The following proposition shows the existence of continuous "bump functions" in a locally compact space. It is sometimes called Urysohn lemma (which is a bit confusing since there is already a Urysohn lemma (Proposition A.31).

Proposition A.39. *Let E be a locally compact (Hausdorff) space. For every compact subset K and every open subset V of E, if $K \subseteq V$, there is a continuous function $f \colon E \to [0, 1]$ such that $f(x) = 1$ for all $x \in K$, and such that* $\mathrm{supp}(f)$ *is compact and* $\mathrm{supp}(f) \subseteq V$*, where* $\mathrm{supp}(f)$ *is the closure of the subset $\{x \in E \mid f(x) \neq 0\}$, called the support of f.*

Proof. Theorem A.39 follows easily from the Urysohn lemma (Theorem A.31). Since E is locally compact, by Proposition A.38 we can find some open subset W with compact closure $\overline{W}$ such that $K \subseteq W \subseteq \overline{W} \subseteq V$. Since $\overline{W}$ is compact, it is normal, so we can apply Theorem A.31 to find a continuous function $f \colon \overline{W} \to [0, 1]$ such that $f(x) = 1$ for all $x \in K$ and $f(x) = 0$ for all $x \in \overline{W} - W$ (the boundary of W). Then we extend f to E by setting to 0 outside $\overline{W}$. Since the support of f is contained in $\overline{W}$, this function is continuous. $\square$

As a corollary of Proposition A.39 we obtain the existence of continuous partitions of unity for a finite open cover of a compact subset.

Proposition A.40. *Let E be a locally compact (Hausdorff) space. For any compact subset K of E and any finite open cover $(U_1, \ldots, U_n)$ of K (that is, $K \subseteq \bigcup_{i=1}^n U_i$), there exist n continuous functions $f_i \colon E \to [0,1]$ such that f_i has compact support* $\mathrm{supp}(f_i) \subseteq U_i$, *and*

$$\sum_{i=1}^n f_i(x) = 1 \quad \textit{for all } x \in K.$$

A proof of Proposition A.40 is not difficult. It can be found in Rudin [79] (Chapter 2, Theorem 2.13) and Lang [62] (Chapter IX, §2). A family $(f_1, \ldots, f_n)$ satisfying the properties of Proposition A.40 is called a *partition of unity on K subordinate to the cover* $(U_1, \ldots, U_n)$.

It is much harder to deal with noncompact manifolds than it is to deal with compact manifolds. However, manifolds are locally compact and it turns out that there are various ways of embedding a locally compact Hausdorff space into a compact Hausdorff space. The most economical construction consists in adding just one point. This construction, known as the *Alexandroff compactification*, is technically useful, and we now describe it and sketch the proof that it achieves its goal.

To help the reader's intuition, let us consider the case of the plane, $\mathbb{R}^2$. If we view the plane, $\mathbb{R}^2$, as embedded in 3-space, $\mathbb{R}^3$, say as the xy plane of equation $z = 0$, we can consider the sphere, Σ, of radius 1 centered on the z-axis at the point $(0,0,1)$ and tangent to the xOy plane at the origin (sphere of equation $x^2 + y^2 + (z-1)^2 = 1$). If N denotes the north pole on the sphere, i.e., the point of coordinates $(0,0,2)$, then any line, D, passing through the north pole and not tangent to the sphere (i.e., not parallel to the xOy plane) intersects the xOy plane in a unique point, M, and the sphere in a unique point, P, other than the north pole, N. This, way, we obtain a bijection between the xOy plane and the punctured sphere Σ, i.e., the sphere with the north pole N deleted. This bijection is called a *stereographic projection*. See Figure A.35.

The Alexandroff compactification of the plane puts the north pole back on the sphere, which amounts to adding a single point at infinity ∞ to the plane. Intuitively, as we travel away from the origin O towards infinity (in any direction!), we tend towards an ideal point at infinity ∞. Imagine that we "bend" the plane so that it gets wrapped around the sphere, according to stereographic projection. See Figure A.36. A simpler example takes a line and gets a circle as its compactification. The Alexandroff compactification is a generalization of these simple constructions.

Definition A.32. Let $(E, \mathcal{O})$ be a locally compact space. Let ω be any point not in E, and let $E_\omega = E \cup \{\omega\}$. Define the family, $\mathcal{O}_\omega$, as follows:

$$\mathcal{O}_\omega = \mathcal{O} \cup \{(E - K) \cup \{\omega\} \mid K \text{ compact in } E\}.$$

The pair, $(E_\omega, \mathcal{O}_\omega)$, is called the *Alexandroff compactification (or one point compactification) of* $(E, \mathcal{O})$. See Figure A.37.

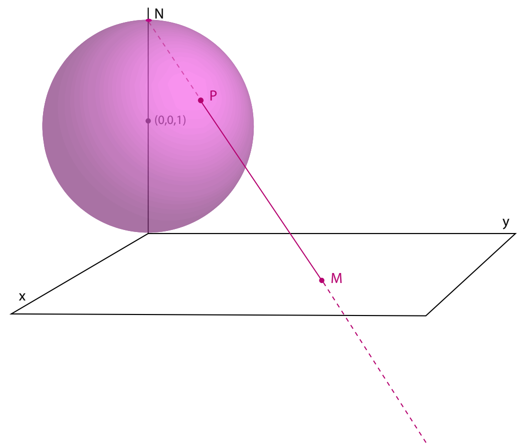


Figure A.35: The stereographic projections of $x^2 + y^2 + (z-1)^2 = 1$ onto the xy-plane.

The following theorem shows that $(E_\omega, \mathcal{O}_\omega)$ is indeed a topological space, and that it is compact.

Theorem A.41. *Let E be a locally compact topological space. The Alexandroff compactification, E_ω, of E is a compact space such that E is a subspace of E_ω and if E is not compact, then $\overline{E} = E_\omega$.*

Proof. The verification that $\mathcal{O}_\omega$ is a family of open sets is not difficult but a bit tedious. Details can be found in Munkres [75] or Schwartz [84]. Let us show that E_ω is compact. For every open cover, $(U_i)_{i \in I}$, of E_ω, since ω must be covered, there is some U_{i_0} of the form

$$U_{i_0} = (E - K_0) \cup \{\omega\}$$

where K_0 is compact in E. Consider the family, $(V_i)_{i \in I}$, defined as follows:

$$\begin{aligned} V_i &= U_i \quad \text{if} \quad U_i \in \mathcal{O}, \\ V_i &= E - K \quad \text{if} \quad U_i = (E - K) \cup \{\omega\}, \end{aligned}$$

where K is compact in E. Then, because each K is compact and thus closed in E (since E is Hausdorff), $E - K$ is open, and every V_i is an open subset of E. Furthermore, the family, $(V_i)_{i \in (I - \{i_0\})}$, is an open cover of K_0. Since K_0 is compact, there is a finite open subcover, $(V_j)_{j \in J}$, of K_0, and thus, $(U_j)_{j \in J \cup \{i_0\}}$ is a finite open cover of E_ω.

Let us show that E_ω is Hausdorff. Given any two points, $a, b \in E_\omega$, if both $a, b \in E$, since E is Hausdorff and every open set in $\mathcal{O}$ is an open set in $\mathcal{O}_\omega$, there exist disjoint open sets, U, V (in $\mathcal{O}$), such that $a \in U$ and $b \in V$. If $b = \omega$, since E is locally compact, there is some compact set, K, containing an open set, U, containing a and then, U and $V = (E - K) \cup \{\omega\}$ are disjoint open sets (in $\mathcal{O}_\omega$) such that $a \in U$ and $b \in V$.

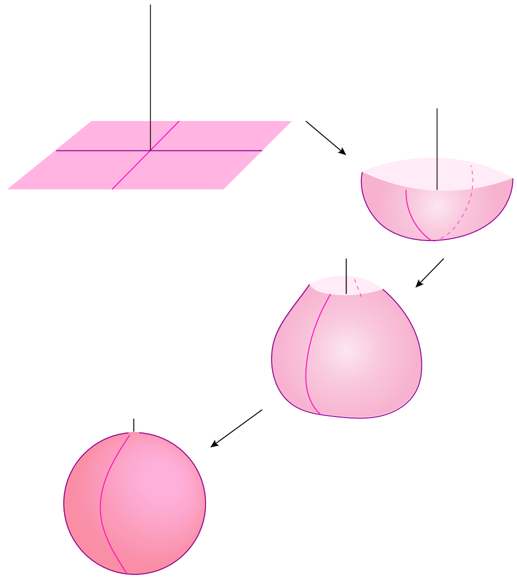


Figure A.36: A four stage illustration of how the xy-plane is wrapped around the unit sphere centered at $(0, 0, 1)$. When finished all of the sphere is covered except the point $(0, 0, 2)$.

The space E is a subspace of E_ω because for every open set, U, in $\mathcal{O}_\omega$, either $U \in \mathcal{O}$ and $E \cap U = U$ is open in E, or $U = (E - K) \cup \{\omega\}$, where K is compact in E, and thus, $U \cap E = E - K$, which is open in E, since K is compact in E and thus, closed (since E is Hausdorff). Finally, if E is not compact, for every compact subset, K, of E, $E - K$ is nonempty and thus, for every open set, $U = (E - K) \cup \{\omega\}$, containing ω, we have $U \cap E \neq \emptyset$, which shows that $\omega \in \overline{E}$ and thus, that $\overline{E} = E_\omega$. $\square$

A.6 Neighborhood Bases and Filters

When dealing with convolution we will need a notion of convergence more general than the notion of convergence of a sequence. There are two equivalent definitions of such a general notion of convergence. One in terms of nets, and the other in terms of filters. For our purposes, the definition in terms of filters is more convenient.

First let us review the notion of neighborhood and neighborhood base.

Definition A.33. Let X be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. For any subset $A \subseteq X$, a *neighborhood* of A is any subset N containing some open subset U containing A; in short, there is some $U \in \mathcal{O}$ such that $A \subseteq U \subseteq N$; see Figure A.38. If $A = \{x\}$, a neighborhood of $\{x\}$ is called simply a *neighborhood of* x.

A *neighborhood base* of a point x (resp. of a subset A) is a family $\mathcal{N}$ of neighborhoods of x (resp. of neighborhoods of A), such for every neighborhood V of x (resp. neighborhood of

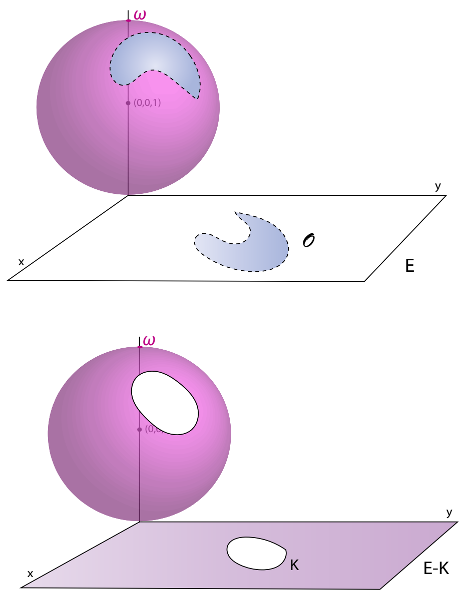


Figure A.37: The two types of open sets associated with the Alexandroff compactification of the xy-plane. The first type of open set does not include ω, i.e. the north pole, while the second type of open set contains ω.

A), there is some $N \in \mathcal{N}$ such that $N \subseteq V$; see Figure A.38.

In many cases a neighborhood base consists of open sets. Let us now define the notion of filter and filter base. This notion is defined for any set, not just for a topological space.

Definition A.34. Let X be any set. A *filter* $\mathcal{F}$ on X is a family of subsets of X satisfying the following properties.

(1) For any two subset A, B of X, if $A \in \mathcal{F}$ and if $A \subseteq B$, then $B \in \mathcal{F}$ ($\mathcal{F}$ is upward-closed).

(2) For any two subsets A, B of X, if $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$ (closure under intersection).

(3) We have $X \in \mathcal{F}$.

(4) The empty set *does not* belong to $\mathcal{F}$.

The axioms of a filter show that filters only exist on nonempty sets. In particular, Axiom (4) prevents $\mathcal{F} = 2^X$ from being a filter.

Example A.4.

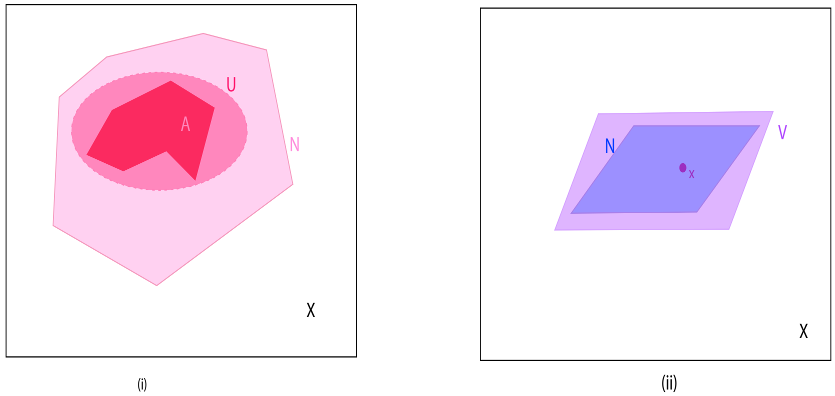


Figure A.38: Figure (i) illustrates a neighborhood of A, while Figure (ii) illustrates a neighborhood base of x.

1. If $X \neq \emptyset$, for any nonempty subset A of X, the family of all subsets of X containing A is a filter.

2. If $(X, \mathcal{O})$ is a topological space, then for any $x \in X$ (resp. any nonempty subset A of X), the family of neighborhoods of x (resp. A) is a filter; see Figure A.39.

3. If X is an infinite set, the family of complements of finite subsets of X is a filter. If $X = \mathbb{N}$, then the filter of complements of finite subsets of $\mathbb{N}$ is called the *Fréchet filter*.

4. Let $\mathcal{F}$ be a filter on X. For any $A \in \mathcal{F}$, let $S(A)$ be the family

$$S(A) = \{B \in \mathcal{F} \mid B \subseteq A\},$$

called a *section*. It is easy to check that the family of sections $S(A)$ (for all $A \in \mathcal{F}$) is a filter on the set $\mathcal{F}$, called the *filter of sections of* $\mathcal{F}$.

Filters are compared as follows.

Definition A.35. Let X be any nonempty set. Given two filters $\mathcal{F}$ and $\mathcal{F}'$ on X, we say that $\mathcal{F}'$ is *finer* than $\mathcal{F}$ if $\mathcal{F} \subseteq \mathcal{F}'$.

A convenient way to generate a filter is to use a filter base.

Definition A.36. Let X be any nonempty set. A *filter base* $\mathcal{B}$ on X is a family of subsets of X satisfying the following properties.

(1) For any two subsets A, B of X, if $A \in \mathcal{B}$ and $B \in \mathcal{B}$, then there is some $C \in \mathcal{B}$ such that $C \subseteq A \cap B$.

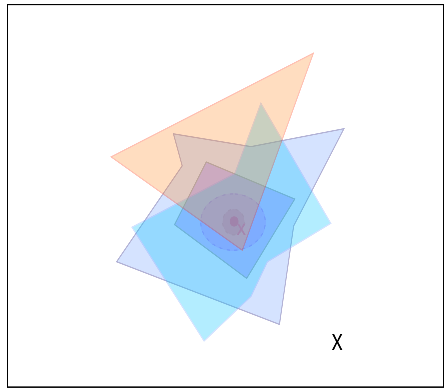


Figure A.39: An illustration of six elements of the (canonical) neighborhood filter of x described by Example A.4, (2).

(2) The family $\mathcal{B}$ is nonempty.

(3) The empty set *does not* belong to $\mathcal{B}$.

It is immediately verified that if $\mathcal{B}$ is a filter base on X, then the family of subsets of X containing some subset in $\mathcal{B}$ is a filter called the *filter generated by* $\mathcal{B}$.

If $(X, \mathcal{O})$ is a topological space, for any $x \in X$, the filter bases of neighborhoods of x are exactly the neighborhood bases of x.

The main reason for introducing filters is to define the following general notion of convergence.

Definition A.37. Let X be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. For any $x \in X$, a filter $\mathcal{F}$ *converges* to x, or x is a *limit* of the filter $\mathcal{F}$, if every neighborhood N of x belongs to $\mathcal{F}$; equivalently, the filter $\mathcal{B}(x)$ of neighborhoods of x is a subset of the filter $\mathcal{F}$; that, is the filter $\mathcal{F}$ is finer than the filter $\mathcal{B}(x)$. A filter base $\mathcal{B}$ *converges* to x if the filter generated by $\mathcal{B}$ converges to x.

The following proposition is an immediate consequence of the definition.

Proposition A.42. *Let X be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. For any $x \in X$, a filter base $\mathcal{B}$ converges to x iff every neighborhood base of x contains some set in $\mathcal{B}$.*

Intuitively, x is a limit of a filter base $\mathcal{B}$ if there are sets in $\mathcal{B}$ as close to x as desired; see Figure A.40.

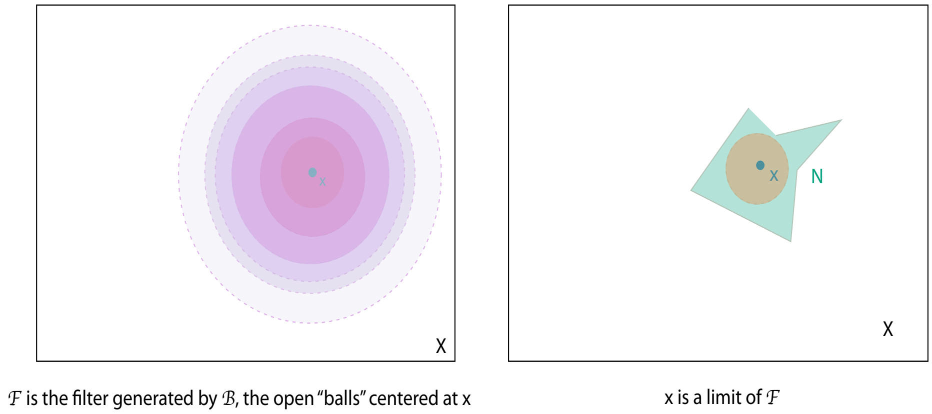


Figure A.40: Let X be a metric space, say $X = \mathbb{R}^2$. Let $\mathcal{F}$ be the filter generated $\mathcal{B}$, where an element of $\mathcal{B}$ is an open ball centered at x. Then by Proposition A.42, x is a limit of $\mathcal{F}$.

The limit of a sequence $(x_n)_{n\geq 0}$ of points $x_n \in X$ is a special case of Definition A.37; see Figure A.41. Indeed, if we define for every $n \geq 0$ the set S_n given by

$$S_n = \{x_p \mid p \geq n\},$$

then the family of sets S_n forms a filter base, and an element $y \in X$ is a limit of the sequence (x_n) iff the filter base $\{S_n\}$ converges to y. Indeed, by Proposition A.42, the filter base $\{S_n\}$ converges to y iff for every neighborhood V of y, there is some $S_n \subseteq V$, in other words, there is some $n \geq 1$ such that $x_p \in V$ for all $p \geq n$, which is the standard definition of convergence of a sequence.

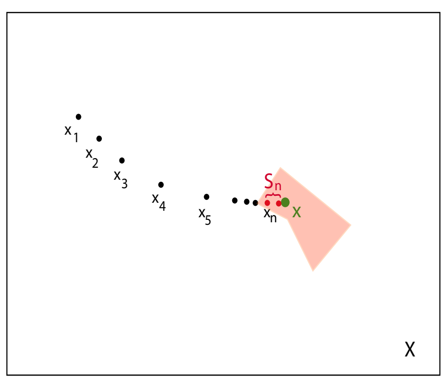


Figure A.41: The convergence of a sequence $(x_n)_{n\geq 0}$ reinterpreted in terms of Definition A.37.

We can also define the notion of limit of a function. Let $f\colon X \to Y$ be a function where

X is any nonempty set and Y is a topological space. Then if $\mathcal{F}$ is any filter on X, it is immediately verified that the family of sets $f(U)$, with $U \in \mathcal{F}$, is a filter base on Y.

Definition A.38. Let $f\colon X \to Y$ be a function where X is any nonempty set and Y is a topological space. For any filter $\mathcal{F}$ on X, and for any $y \in Y$, we say that y *is a limit of f according to $\mathcal{F}$* (or simply that y *is a limit of $\mathcal{F}$*) if the filter basis $f(\mathcal{F})$ (consisting of the subsets $f(U)$ of Y with $U \in \mathcal{F}$) converges to y. We write

$$\lim_{x,\mathcal{F}} f(x) = y.$$

If we view a sequence $(x_n)_{n\geq 0}$ of points in a topological space X as a function $x\colon \mathbb{N} \to X$, then (x_n) converges to y in the traditional sense iff x converges to y according to the Fréchet filter on $\mathbb{N}$ (the family of subsets of $\mathbb{N}$ having a finite complement); see Figure A.42.

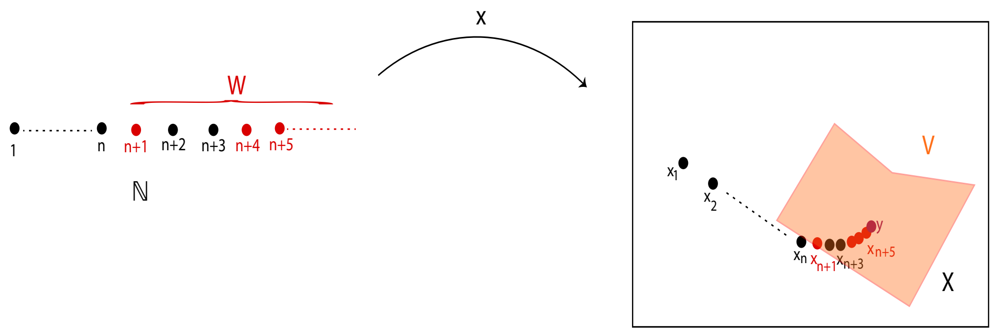


Figure A.42: The convergence of a sequence $(x_n)_{n\geq 0}$ reinterpreted in terms of Definition A.38 and the Fréchet filter on $\mathbb{N}$. For this illustration $W = \mathbb{N} - \{1, 2, \ldots, n, n+2, n+3\}$.

The following useful characterization of a limit of a filter is immediate from the definitions.

Proposition A.43. *Let $f\colon X \to Y$ be a function where X is any nonempty set and Y is a topological space. A point $y \in Y$ is a limit of a filter $\mathcal{F}$ on X if and only if for every neighborhood V of y, there is some $W \in \mathcal{F}$ such that $f(W) \subseteq V$, or equivalently, $f^{-1}(V) \in \mathcal{F}$ for every neighborhood V of y.*

Filters also provide a useful characterization of the notion of compactness. First we define ultrafilters.

Definition A.39. Let X be any nonempty set. A filter $\mathcal{F}$ on X is an *ultrafilter* if it is a maximal filter; that is, there is no filter different from $\mathcal{F}$ and finer than $\mathcal{F}$.

For example, for any $x \in X$, the filter of subsets containing x is an ultrafilter. The following important result shows that there are many ultrafilters, but they are very nonconstructive in nature. The proof uses Zorn's lemma.

Theorem A.44. *Let X be any nonempty set. Every filter $\mathcal{F}$ on X is contained in a finer ultrafilter.*

Observe that an ultrafilter $\mathcal{F}$ has the following completeness property: for any subset A of X, either $A \in \mathcal{F}$ or its complement $X - A \in \mathcal{F}$, but not both.

Indeed, if $A \notin \mathcal{F}$ and $X - A \notin \mathcal{F}$, then it is easy to see that the family $\mathcal{G}$ of subsets of X given by

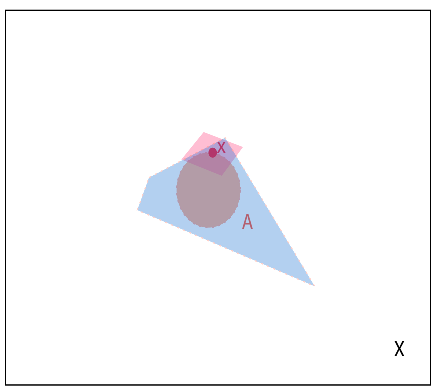
$$\mathcal{G} = \{B \subseteq X \mid A \cup B \in \mathcal{F}\}$$

is a filter finer than $\mathcal{F}$ and containing $X - A$, thus strictly finer than $\mathcal{F}$, contradicting the maximality of $\mathcal{F}$. This property of ultrafilters is used in logic to prove completeness results.

We also need to define a notion weaker than the notion of limit of a filter.

Definition A.40. Let X be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. A point $x \in X$ is a *cluster point* (or *cluster*) of a filter base $\mathcal{B}$ if every neighborhood of x has a nonempty intersection with every set in $\mathcal{B}$ (equivalently, if $x \in \bigcap_{V \in \mathcal{B}} \overline{V}$).

A limit x of a filter is a cluster point, but the converse is false in general; see Figure A.43.



$\mathcal{F}$ is the filter of neighborhoods of A

Figure A.43: Let A be an open disk and x any point on the boundary of X. Such an x is a cluster point of $\mathcal{F}$, the filter of neighborhoods of A, since any pink neighborhood of x has an intersection with any blue neighborhood of A. However this x is not a limit of $\mathcal{F}$ since $\mathcal{F}$ is not finer than the filter of neighborhood of x.

We see immediately that x is a cluster point of a filter $\mathcal{F}$ iff there is a filter $\mathcal{G}$ finer than $\mathcal{F}$ and the filter $\mathcal{G}$ converges to x. An ultrafilter $\mathcal{F}$ converges to a limit x iff x is a cluster point of $\mathcal{F}$.

Finally, we have the following characterizations of compactness.

Theorem A.45. *Let X be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. The following properties are equivalent.*

(1) Every filter $\mathcal{F}$ on X has some cluster point.

(2) Every ultrafilter $\mathcal{F}$ on X converges to some limit.

(3) Every open cover $(U_\alpha)_{\alpha\in I}$ of X contains some finite subcover; that is, if $\bigcup_{\alpha\in I} U_\alpha = X$, then there is a finite subset J of I such that $\bigcup_{\alpha\in J} U_\alpha = X$.

(4) For every family $(F_\alpha)_{\alpha\in I}$ of closed subsets of X, if $\bigcap_{\alpha\in I} F_\alpha = \emptyset$, then there is a finite subset J of I such that $\bigcap_{\alpha\in J} F_\alpha = \emptyset$.

Let us also mention that a topological space X is Hausdorff if and only if every filter has at most one limit.

The theory of filters and their use in topology is discussed quite extensively in Bourbaki [13] (Chapter 1).

A.7 Second-Countable and Separable Spaces

In studying surfaces and manifolds, an important property is the existence of a countable basis for the topology. Indeed this property, among other things, guarantees the existence of triangulations of manifolds, and the fact that a manifold is metrizable.

Definition A.41. A topological space E is called *second-countable* if there is a countable basis for its topology, i.e., if there is a countable family, $(U_i)_{i\geq 0}$, of open sets such that every open set of E is a union of open sets U_i.

It is easily seen that $\mathbb{R}^n$ is second-countable and more generally, that every normed vector space of finite dimension is second-countable. More generally, a metric space is second-countable if and only if it is separable, a very useful property that holds for all of the spaces that we will consider in practice.

Definition A.42. A topological space E is *separable* if it contains some countable subset S which is dense in X, that is, $\overline{S} = E$.

Observe that by Proposition A.4, a subset S of E is dense in E iff every nonempty open subset of E contains some element of S.

The (metric) space $\mathbb{R}$ is separable because $\mathbb{Q}$ is a countable dense subset of $\mathbb{R}$. Similarly, $\mathbb{C}$ is separable. In general, $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$, so $\mathbb{R}^n$ is separable, and similarly, every finite-dimensional normed vector space over $\mathbb{R}$ (or $\mathbb{C}$) is separable. For metric spaces, we have the following useful result.

Proposition A.46. *If E is a metric space, then E is second-countable if and only if E is separable.*

Proof. If $\mathcal{B} = (B_n)$ is a countable basis for the topology of E, then for any set S obtained by picking some point s_n in B_n, since every nonempty open subset U of E is the union of some of the B_n, the intersection $U \cap S$ is nonempty, and so S is dense in E.

Conversely, assume that there is a countable subset $S = (s_n)$ of E which is dense in E. We claim that the countable family $\mathcal{B}$ of open balls $B_0(s_n, 1/m)$ ($m \in \mathbb{N}, m > 0$) is a basis for the topology of E. For every $x \in E$ and every $r > 0$, there is some $m > 0$ such that $1/m < r/2$, and some n such that $s_n \in B_0(x, 1/m)$. It follows that $x \in B_0(s_n, 1/m)$. For all $y \in B_0(s_n, 1/m)$, we have

$$d(x, y) \leq d(x, s_n) + d(s_n, y) \leq 2/m < r,$$

thus $B_0(s_n, 1/m) \subseteq B_0(x, r)$, which by Proposition A.8(a) implies that $\mathcal{B}$ is a basis for the topology of E. $\square$

Proposition A.47. *If E is a compact metric space, then E is separable.*

Proof. For every $n > 0$, the family of open balls of radius $1/n$ forms an open cover of E, and since E is compact, there is a finite subset A_n of E such that $E = \bigcup_{a_i \in A_n} B_0(a_i, 1/n)$. It is easy to see that this is equivalent to the condition $d(x, A_n) < 1/n$ for all $x \in E$. Let $A = \bigcup_{n \geq 1} A_n$. Then A is countable, and for evey $x \in E$, we have

$$d(x, A) \leq d(x, A_n) < \frac{1}{n}, \quad \text{for all } n \geq 1,$$

which implies that $d(x, A) = 0$; that is, A is dense in E. $\square$

The following theorem due to Uryshon gives a very useful sufficient condition for a topological space to be metrizable.

Theorem A.48. *(Urysohn metrization theorem) If a topological space E is regular and second-countable, then it is metrizable.*

The proof of Theorem A.48 can be found in Munkres [75] (Chapter 4, Theorem 34.1). As a corollary of Theorem A.48, every (second-countable) manifold, and thus every Lie group is metrizable.

The following technical result shows that a locally compact metrizable space which is also separable can be expressed as the union of a countable monotonic sequence of compact subsets. This gives us a method for generalizing various properties of compact metric spaces to locally compact metric spaces of the above kind.

Proposition A.49. *Let E be a locally compact metrizable space. The following properties are equivalent:*

(1) There is a sequence $(U_n)_{n\geq 0}$ *of open subsets such that for all* $n \in \mathbb{N}$, $U_n \subseteq U_{n+1}$, $\overline{U_n}$ *is compact,* $\overline{U_n} \subseteq U_{n+1}$*, and* $E = \bigcup_{n\geq 0} U_n = \bigcup_{n\geq 0} \overline{U_n}$.

(2) The space E *is the union of a countable family of compact subsets of* E.

(3) The space E *is separable.*

Proof. We show (1) implies (2), (2) implies (3), and (3) implies (1). Obviously, (1) implies (2) since the $\overline{U_n}$ are compact.

If (2) holds, then $E = \bigcup_{n\geq 0} K_n$, for some compact subsets K_n. By Proposition A.47, each compact subset K_n is separable, so let S_n be a countable dense subset of K_n, Then $S = \bigcup_{n\geq 0} S_n$ is a countable dense subset of E, since

$$E = \bigcup_{n\geq 0} K_n \subseteq \bigcup_{n\geq 0} \overline{S_n} \subseteq \overline{S} \subseteq E.$$

Consequently (3) holds.

If (3) holds, let $S = \{s_n\}$ be a countable dense subset of E. By Proposition A.46, the space E has a countable basis $\mathcal{B}$ of open sets O_n. Since E is locally compact, for every $x \in E$, there is some compact neighborhood W_x containing x, and by Proposition A.8, there some index $n(x)$ such that $x \in O_{n(x)} \subseteq W_x$. Since W_x is a compact neighborhood, we deduce that $\overline{O_{n(x)}}$ is compact. Consequently, there is a subfamily of $\mathcal{B}$ consisting of open subsets O_i such that $\overline{O_i}$ is compact, which is a countable basis for the topology of E, so we may assume that we restrict our attention to this basis. We define the sequence $(U_n)_{n\geq 1}$ of open subsets of E by induction as follows: Set $U_1 = O_1$, and let

$$U_{n+1} = O_{n+1} \cup V_r(\overline{U_n}),$$

where $r > 0$ is chosen so that $\overline{V_r(\overline{U_n})}$ is compact, which is possible by Proposition A.37. We immediately check that the U_n satisfy (1) of Proposition A.49. □

Definition A.43. Given a topological space E, a subset A of E is *σ-compact* (or *countable at infinity*) if A is the union of countably many compact subsets.

Note that Proposition A.49 implies that a locally compact metrizable space is separable iff it is σ-compact.

It can also be shown that if E is a locally compact space that has a countable basis, then E_ω also has a countable basis (and in fact, is metrizable).

We also have the following property.

Proposition A.50. *Given a second-countable topological space* E*, every open cover* $(U_i)_{i\in I}$*, of* E *contains some countable subcover.*

Proof. Let $(O_n)_{n\geq 0}$ be a countable basis for the topology. Then all sets O_n contained in some U_i can be arranged into a countable subsequence, $(\Omega_m)_{m\geq 0}$, of $(O_n)_{n\geq 0}$ and for every Ω_m, there is some U_{i_m} such that $\Omega_m \subseteq U_{i_m}$. Furthermore, every U_i is some union of sets Ω_j, and thus, every $a \in E$ belongs to some Ω_j, which shows that $(\Omega_m)_{m\geq 0}$ is a countable open subcover of $(U_i)_{i\in I}$. ☐

As an immediate corollary of Proposition A.50, a locally connected second-countable space has countably many connected components.

A.8 Sequential Compactness

For a general topological Hausdorff space E, the definition of compactness relies on the existence of finite cover. However, when E has a countable basis or is a metric space, we may define the notion of compactness in terms of sequences. To understand how this is done, we need to first define accumulation points.

Definition A.44. Given a topological Hausdorff space, E, given any sequence, (x_n), of points in E, a point, $l \in E$, is an *accumulation point (or cluster point)* of the sequence (x_n) if every open set, U, containing l contains x_n for infinitely many n. See Figure A.44.

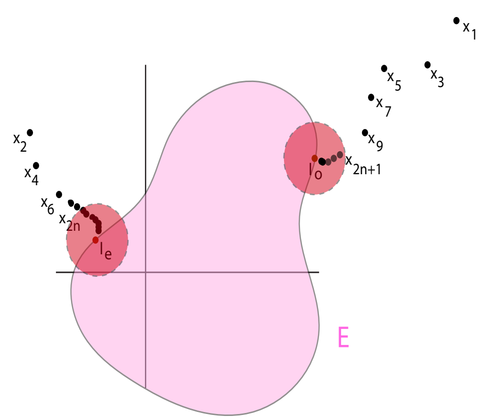


Figure A.44: The space E is the closed, bounded pink subset of $\mathbb{R}^2$. The sequence (x_n) has two accumulation points, one for the subsequence (x_{2n+1}) and one for (x_{2n}).

Clearly, if l is a limit of the sequence, (x_n), then it is an accumulation point, since every open set, U, containing a contains all x_n except for finitely many n.

For second-countable spaces we are able to give another characterization of accumulation points.

Proposition A.51. *Given a second-countable topological Hausdorff space, E, a point, l, is an accumulation point of the sequence, (x_n), iff l is the limit of some subsequence, (x_{n_k}), of (x_n).*

Proof. Clearly, if l is the limit of some subsequence (x_{n_k}) of (x_n), it is an accumulation point of (x_n).

Conversely, let $(U_k)_{k\geq 0}$ be the sequence of open sets containing l, where each U_k belongs to a countable basis of E, and let $V_k = U_1 \cap \cdots \cap U_k$. For every $k \geq 1$, we can find some $n_k > n_{k-1}$ such that $x_{n_k} \in V_k$, since l is an accumulation point of (x_n). Now, since every open set containing l contains some U_{k_0} and since $x_{n_k} \in U_{k_0}$ for all $k \geq 0$, the sequence (x_{n_k}) has limit l. $\square$

Remark: Proposition A.51 also holds for metric spaces.

As an illustration of Proposition A.51 let (x_n) be the sequence $(1, -1, 1, -1, \ldots)$. This sequence has two accumulation points, namely 1 and -1 since $(x_{2n+1}) = (1)$ and $(x_{2n}) = (-1)$.

In second-countable Hausdorff spaces, compactness can be characterized in terms of accumulation points (this is also true for metric spaces).

Proposition A.52. *A second-countable topological Hausdorff space, E, is compact iff every sequence, (x_n), of E has some accumulation point in E.*

Proof. Assume that every sequence, (x_n), has some accumulation point. Let $(U_i)_{i\in I}$ be some open cover of E. By Proposition A.50, there is a countable open subcover, $(O_n)_{n\geq 0}$, for E. Now, if E is not covered by any finite subcover of $(O_n)_{n\geq 0}$, we can define a sequence, (x_m), by induction as follows:

Let x_0 be arbitrary and for every $m \geq 1$, let x_m be some point in E not in $O_1 \cup \cdots \cup O_m$, which exists, since $O_1 \cup \cdots \cup O_m$ is not an open cover of E. We claim that the sequence, (x_m), does not have any accumulation point. Indeed, for every $l \in E$, since $(O_n)_{n\geq 0}$ is an open cover of E, there is some O_m such that $l \in O_m$, and by construction, every x_n with $n \geq m + 1$ does not belong to O_m, which means that $x_n \in O_m$ for only finitely many n and l is not an accumulation point. See Figure A.45.

Conversely, assume that E is compact, and let (x_n) be any sequence . If $l \in E$ is not an accumulation point of the sequence, then there is some open set, U_l, such that $l \in U_l$ and $x_n \in U_l$ for only finitely many n. Thus, if (x_n) does not have any accumulation point, the family, $(U_l)_{l\in E}$, is an open cover of E and since E is compact, it has some finite open subcover, $(U_l)_{l\in J}$, where J is a finite subset of E. But every U_l with $l \in J$ is such that $x_n \in U_l$ for only finitely many n, and since J is finite, $x_n \in \bigcup_{l\in J} U_l$ for only finitely many n, which contradicts the fact that $(U_l)_{l\in J}$ is an open cover of E, and thus contains all the x_n. Thus, (x_n) has some accumulation point. See Figure A.46. $\square$

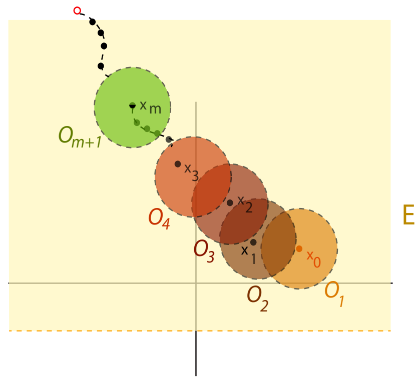


Figure A.45: The space E is the open half plane above the line $y = -1$. S ince E is not compact, we inductively build a sequence, (x_n) that will have no accumulation point in E. Note the y coordinate of x_n approaches infinity.

Remarks:

1. By combining Propositions A.51 and A.52, we have observe that a second-countable Hausdorff space E is compact iff every sequence (x_n) has a convergent subsequence (x_{n_k}). In other words, we say a second-countable Hausdorff space E is compact iff it is *sequentially compact*.

2. It should be noted that the proof showing that if E is compact, then every sequence has some accumulation point, holds for any arbitrary compact space (the proof does not use a countable basis for the topology). The converse also holds for metric spaces. We will prove this converse since it is a major property of metric spaces.

Given a metric space in which every sequence has some accumulation point, we first prove the existence of a *Lebesgue number*.

Lemma A.53. *Given a metric space, E, if every sequence, (x_n), has an accumulation point, for every open cover, $(U_i)_{i\in I}$, of E, there is some $\delta > 0$ (a Lebesgue number for $(U_i)_{i\in I}$) such that, for every open ball, $B_0(a,\epsilon)$, of radius $\epsilon \leq \delta$, there is some open subset, U_i, such that $B_0(a,\epsilon) \subseteq U_i$. See Figure A.47*

Proof. If there was no δ with the above property, then, for every natural number, n, there would be some open ball, $B_0(a_n, 1/n)$, which is not contained in any open set, U_i, of the open cover, $(U_i)_{i\in I}$. However, the sequence, (a_n), has some accumulation point, a, and since $(U_i)_{i\in I}$ is an open cover of E, there is some U_i such that $a \in U_i$. Since U_i is open, there is some open ball of center a and radius ϵ contained in U_i. Now, since a is an accumulation

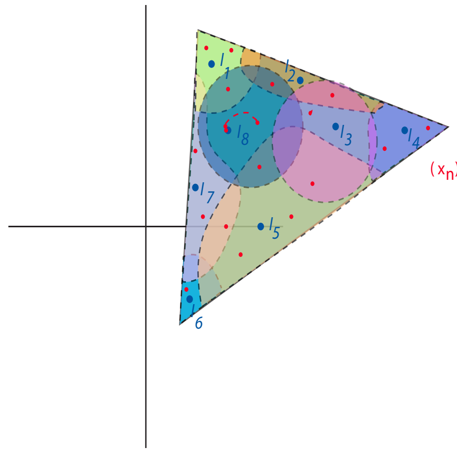


Figure A.46: The space E the closed triangular region of $\mathbb{R}^2$. Given a sequence (x_n) of red points in E, if the sequence has no accumulation points, then each l_i for $1 \leq i \leq 8$, is not an accumulation point. But as implied by the illustration, l_8 actually is an accumulation point of (x_n).

point of the sequence, (a_n), every open set containing a contains a_n for infinitely many n and thus, there is some n large enough so that

$$1/n \leq \epsilon/2 \quad \text{and} \quad a_n \in B_0(a, \epsilon/2),$$

which implies that

$$B_0(a_n, 1/n) \subseteq B_0(a, \epsilon) \subseteq U_i,$$

a contradiction. ☐

By a previous remark, since the proof of Proposition A.52 implies that in a compact topological space, every sequence has some accumulation point, by Lemma A.53, in a compact metric space, every open cover has a Lebesgue number. This fact can be used to prove another important property of compact metric spaces, the uniform continuity theorem.

Definition A.45. Given two metric spaces, (E, d_E) and (F, d_F), a function, $f: E \to F$, is *uniformly continuous* if for every $\epsilon > 0$, there is some $\eta > 0$, such that, for all $a, b \in E$,

$$\text{if} \quad d_E(a, b) \leq \eta \quad \text{then} \quad d_F(f(a), f(b)) \leq \epsilon.$$

See Figures A.48 and A.49.

As we saw earlier, the metric on a metric space is uniformly continuous, and the norm on a normed metric space is uniformly continuous.

The *uniform continuity theorem* can be stated as follows:

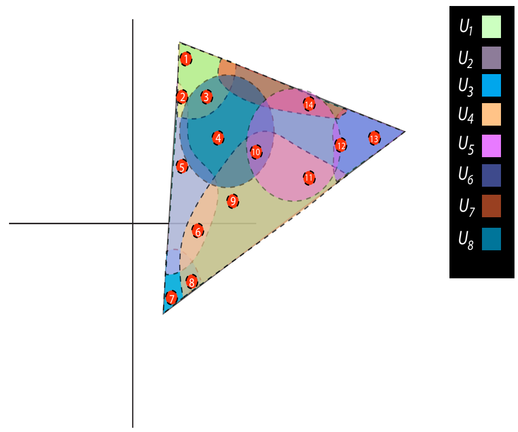


Figure A.47: The space E the closed triangular region of $\mathbb{R}^2$. It's open cover is $(U_i)_{i=1}^{8}$. The Lebesque number is the radius of the small orange balls labelled 1 through 14. Each open ball of this radius entirely contained within at least one U_i. For example, Ball 2 is contained in both U_1 and U_2.

Theorem A.54. *Given two metric spaces, (E, d_E) and (F, d_F), if E is compact and if $f\colon E \to F$ is a continuous function, then f is uniformly continuous.*

Proof. Consider any $\epsilon > 0$ and let $(B_0(y, \epsilon/2))_{y \in F}$ be the open cover of F consisting of open balls of radius $\epsilon/2$. Since f is continuous, the family,

$$(f^{-1}(B_0(y, \epsilon/2)))_{y \in F},$$

is an open cover of E. Since, E is compact, by Lemma A.53, there is a Lebesgue number, δ, such that for every open ball, $B_0(a, \eta)$, of radius $\eta \leq \delta$, then $B_0(a, \eta) \subseteq f^{-1}(B_0(y, \epsilon/2))$, for some $y \in F$. In particular, for any $a, b \in E$ such that $d_E(a, b) \leq \eta = \delta/2$, we have $a, b \in B_0(a, \delta)$ and thus, $a, b \in f^{-1}(B_0(y, \epsilon/2))$, which implies that $f(a), f(b) \in B_0(y, \epsilon/2)$. But then, $d_F(f(a), f(b)) \leq \epsilon$, as desired. $\square$

We now prove another lemma needed to obtain the characterization of compactness in metric spaces in terms of accumulation points.

Lemma A.55. *Given a metric space, E, if every sequence, (x_n), has an accumulation point, then for every $\epsilon > 0$, there is a finite open cover, $B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon)$, of E by open balls of radius ϵ.*

Proof. Let a_0 be any point in E. If $B_0(a_0, \epsilon) = E$, then the lemma is proved. Otherwise, assume that a sequence, $(a_0, a_1, \ldots, a_n)$, has been defined, such that $B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon)$ does not cover E. Then, there is some a_{n+1} not in $B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon)$ and either

$$B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_{n+1}, \epsilon) = E,$$

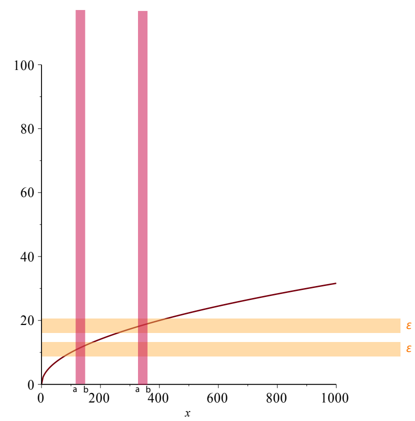


Figure A.48: The real valued function $f(x) = \sqrt{x}$ is uniformly continuous over $(0, \infty)$. Fix ϵ. If the x values lie within the rose colored η strip, the y values always lie within the peach ϵ strip.

in which case the lemma is proved, or we obtain a sequence, $(a_0, a_1, \ldots, a_{n+1})$, such that $B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_{n+1}, \epsilon)$ does not cover E. If this process goes on forever, we obtain an infinite sequence, (a_n), such that $d(a_m, a_n) > \epsilon$ for all $m \neq n$. Since every sequence in E has some accumulation point, the sequence, (a_n), has some accumulation point, a. Then, for infinitely many n, we must have $d(a_n, a) \leq \epsilon/3$ and thus, for at least two distinct natural numbers, p, q, we must have $d(a_p, a) \leq \epsilon/3$ and $d(a_q, a) \leq \epsilon/3$, which implies $d(a_p, a_q) \leq d(a_p, a) + d(a_q, a) \leq 2\epsilon/3$, contradicting the fact that $d(a_m, a_n) > \epsilon$ for all $m \neq n$. See Figure A.50. Thus, there must be some n such that

$$B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon) = E. \qquad \square$$

Definition A.46. A metric space E is said to be *precompact* (or *totally bounded*) if for every $\epsilon > 0$, there is a finite open cover, $B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon)$, of E by open balls of radius ϵ.

We now obtain the *Weierstrass–Bolzano* property.

Theorem A.56. *A metric space, E, is compact iff every sequence, (x_n), has an accumulation point.*

Proof. We already observed that the proof of Proposition A.52 shows that for any compact space (not necessarily metric), every sequence, (x_n), has an accumulation point. Conversely, let E be a metric space, and assume that every sequence, (x_n), has an accumulation point. Given any open cover, $(U_i)_{i \in I}$ for E, we must find a finite open subcover of E. By Lemma A.53, there is some $\delta > 0$ (a Lebesgue number for $(U_i)_{i \in I}$) such that, for every open ball, $B_0(a, \epsilon)$, of radius $\epsilon \leq \delta$, there is some open subset, U_j, such that $B_0(a, \epsilon) \subseteq U_j$. By Lemma

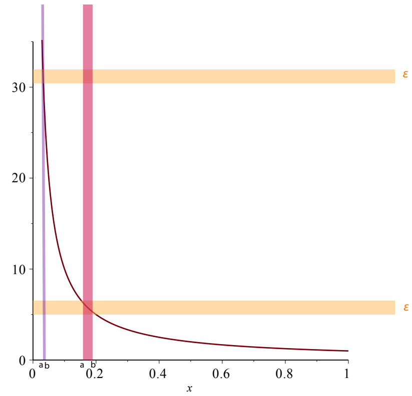


Figure A.49: The real valued function $f(x) = 1/x$ is not uniformly continuous over $(0, \infty)$. Fix ϵ. In order for the y values to lie within the peach epsilon strip, the widths of the eta strips decrease as $x \to 0$.

A.55, for every $\delta > 0$, there is a finite open cover, $B_0(a_0, \delta) \cup \cdots \cup B_0(a_n, \delta)$, of E by open balls of radius δ. But from the previous statement, every open ball, $B_0(a_i, \delta)$, is contained in some open set, U_{j_i}, and thus, $\{U_{j_1}, \ldots, U_{j_n}\}$ is an open cover of E. $\square$

A.9 Complete Metric Spaces and Compactness

Another very useful characterization of compact metric spaces is obtained in terms of Cauchy sequences. Such a characterization is quite useful in fractal geometry (and elsewhere). First recall the definition of a Cauchy sequence and of a complete metric space.

Definition A.47. Given a metric space, (E, d), a sequence, $(x_n)_{n \in \mathbb{N}}$, in E is a *Cauchy sequence* if the following condition holds: for every $\epsilon > 0$, there is some $p \geq 0$, such that, for all $m, n \geq p$, then $d(x_m, x_n) \leq \epsilon$.

If every Cauchy sequence in (E, d) converges we say that (E, d) is a *complete metric space*.

First let us show the following proposition:

Proposition A.57. *Given a metric space, E, if a Cauchy sequence, (x_n), has some accumulation point, a, then a is the limit of the sequence, (x_n).*

Proof. Since (x_n) is a Cauchy sequence, for every $\epsilon > 0$, there is some $p \geq 0$, such that, for all $m, n \geq p$, then $d(x_m, x_n) \leq \epsilon/2$. Since a is an accumulation point for (x_n), for infinitely

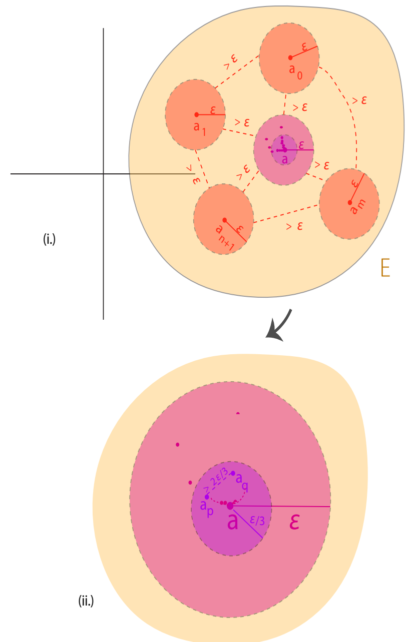


Figure A.50: Let E be the peach region of $\mathbb{R}^2$. If E is not covered by a finite collection of orange balls with radius ϵ, the points of the sequence (a_n) are separated by a distance of at least ϵ. This contradicts the fact that a is the accumulation point of a, as evidenced by the enlargement of the plum disk in Figure (ii).

many n, we have $d(x_n, a) \leq \epsilon/2$, and thus, for at least some $n \geq p$, we have $d(x_n, a) \leq \epsilon/2$. Then, for all $m \geq p$,

$$d(x_m, a) \leq d(x_m, x_n) + d(x_n, a) \leq \epsilon,$$

which shows that a is the limit of the sequence (x_n). $\square$

We can now prove the following theorem.

Theorem A.58. *A metric space, E, is compact iff it is precompact and complete.*

Proof. Let E be compact. For every $\epsilon > 0$, the family of all open balls of radius ϵ is an open cover for E and since E is compact, there is a finite subcover, $B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon)$, of E by open balls of radius ϵ. Thus E is precompact. Since E is compact, by Theorem A.56, every sequence, (x_n), has some accumulation point. Thus every Cauchy sequence, (x_n), has some accumulation point, a, and, by Proposition A.57, a is the limit of (x_n). Thus, E is complete.

Now assume that E is precompact and complete. We prove that every sequence, (x_n), has an accumulation point. By the other direction of Theorem A.56, this shows that E is

compact. Given any sequence, (x_n), we construct a Cauchy subsequence, (y_n), of (x_n) as follows: Since E is precompact, letting $\epsilon = 1$, there exists a finite cover, $\mathcal{U}_1$, of E by open balls of radius 1. Thus some open ball, B_o^0, in the cover, $\mathcal{U}_1$, contains infinitely many elements from the sequence (x_n). Let y_0 be any element of (x_n) in B_o^0. By induction, assume that a sequence of open balls, $(B_o^i)_{1 \leq i \leq m}$, has been defined, such that every ball, B_o^i, has radius $\frac{1}{2^i}$, contains infinitely many elements from the sequence (x_n) and contains some y_i from (x_n) such that

$$d(y_i, y_{i+1}) \leq \frac{1}{2^i},$$

for all i, $0 \leq i \leq m-1$. See Figure A.51. Then letting $\epsilon = \frac{1}{2^{m+1}}$, because E is precompact, there is some finite cover, $\mathcal{U}_{m+1}$, of E by open balls of radius ϵ and thus, of the open ball B_o^m. Thus, some open ball, B_o^{m+1}, in the cover, $\mathcal{U}_{m+1}$, contains infinitely many elements from the sequence, (x_n), and we let y_{m+1} be any element of (x_n) in B_o^{m+1}. Thus, we have defined by induction a sequence, (y_n), which is a subsequence of, (x_n), and such that

$$d(y_i, y_{i+1}) \leq \frac{1}{2^i},$$

for all i. However, for all $m, n \geq 1$, we have

$$d(y_m, y_n) \leq d(y_m, y_{m+1}) + \cdots + d(y_{n-1}, y_n) \leq \sum_{i=m}^{n} \frac{1}{2^i} \leq \frac{1}{2^{m-1}},$$

and thus, (y_n) is a Cauchy sequence Since E is complete, the sequence, (y_n), has a limit, and since it is a subsequence of (x_n), the sequence, (x_n), has some accumulation point. □

Another useful property of a complete metric space is that a subset is closed iff it is complete. This is shown in the following two propositions.

Proposition A.59. *Let (E, d) be a metric space, and let A be a subset of E. If A is complete (which means that every Cauchy sequence of elements in A converges to some point of A), then A is closed in E.*

Proof. Assume $x \in \overline{A}$. By Proposition A.13, there is some sequence (a_n) of points $a_n \in A$ which converges to x. Consequently (a_n) is a Cauchy sequence in E, and thus a Cauchy sequence in A (since $a_n \in A$ for all n). Since A is complete, the sequence (a_n) has a limit $a \in A$, but since E is a metric space it is Hausdorff, so $a = x$, which shows that $x \in A$; that is, A is closed. □

Proposition A.60. *Let (E, d) be a metric space, and let A be a subset of E. If E is complete and if A is closed in E, then A is complete.*

Proof. Let (a_n) be a Cauchy sequence in A. The sequence (a_n) is also a Cauchy sequence in E, and since E is complete, it has a limit $x \in E$. But $a_n \in A$ for all n, so by Proposition A.13 we must have $x \in \overline{A}$. Since A is closed, actually $x \in A$, which proves that A is complete. □

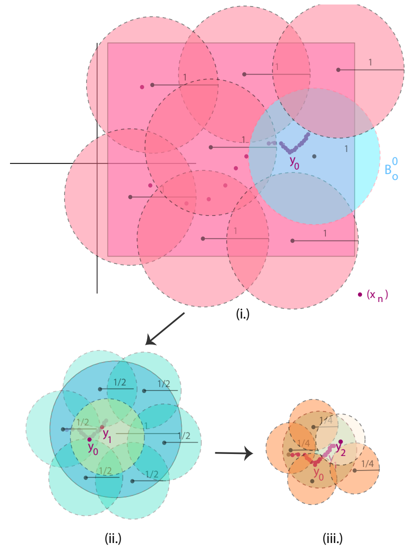


Figure A.51: The first three stages of the construction of the Cauchy sequence (y_n), where E is the pink square region of $\mathbb{R}^2$. The original sequence (x_n) is illustrated with plum colored dots. Figure $(i.)$ covers E with ball of radius 1 and shows the selection of B_o^0 and y_0. Figure $(ii.)$ covers B_o^0 with balls of radius 1/2 and selects the yellow ball as B_o^1 with point y_1. Figure $(iii.)$ covers B_o^1 with balls of radius 1/4 and selects the pale peach ball as B_o^2 with point y_2.

An arbitrary metric space (E, d) is not necessarily complete, but there is a construction of a metric space $(\widehat{E}, \widehat{d})$ such that $\widehat{E}$ is complete, and there is a continuous (injective) distance-preserving map $\varphi\colon E \to \widehat{E}$ such that $\varphi(E)$ is dense in $\widehat{E}$. This is a generalization of the construction of the set $\mathbb{R}$ of real numbers from the set $\mathbb{Q}$ of rational numbers in terms of Cauchy sequences. This construction can be immediately adapted to a normed vector space $(E, \| \, \|)$ to embed $(E, \| \, \|)$ into a complete normed vector space $(\widehat{E}, \| \, \|_{\widehat{E}})$ (a Banach space). This construction is used heavily in integration theory, where E is a set of functions.

A.10 Completion of a Metric Space

In order to prove a kind of uniqueness result for the completion $(\widehat{E}, \widehat{d})$ of a metric space (E, d), we need the following result about extending a uniformly continuous function.

Recall that E_0 is dense in E iff $\overline{E_0} = E$. Since E is a metric space, by Proposition A.13, this means that for every $x \in E$, there is some sequence (x_n) converging to x, with $x_n \in E_0$.

Theorem A.61. *Let E and F be two metric spaces, let E_0 be a dense subspace of E, and let $f_0\colon E_0 \to F$ be a continuous function. If f_0 is uniformly continuous and if F is complete, then there is a unique uniformly continuous function $f\colon E \to F$ extending f_0.*

Proof. We follow Schwartz's proof; see Schwartz [83] (Chapter XI, Section 3, Theorem 1).

Step 1. We begin by constructing a function $f\colon E \to F$ extending f_0. Since E_0 is dense in E, for every $x \in E$, there is some sequence (x_n) converging to x, with $x_n \in E_0$. Then the sequence (x_n) is a Cauchy sequence in E. We claim that $(f_0(x_n))$ is a Cauchy sequence in F.

Proof of the claim. For every $\epsilon > 0$, since f_0 is uniformly continuous, there is some $\eta > 0$ such that for all $(y, z) \in E_0$, if $d(y, z) \leq \eta$, then $d(f_0(y), f_0(z)) \leq \epsilon$. Since (x_n) is a Cauchy sequence with $x_n \in E_0$, there is some integer $p > 0$ such that if $m, n \geq p$, then $d(x_m, x_n) \leq \eta$, thus $d(f_0(x_m), f_0(x_n)) \leq \epsilon$, which proves that $(f_0(x_n))$ is a Cauchy sequence in F. ☐

Since F is complete and $(f_0(x_n))$ is a Cauchy sequence in F, the sequence $(f_0(x_n))$ converges to some element of F; denote this element by $f(x)$.

Step 2. Let us now show that $f(x)$ does not depend on the sequence (x_n) converging to x. Suppose that (x'_n) and (x''_n) are two sequences of elements in E_0 converging to x. Then the mixed sequence

$$x'_0, x''_0, x'_1, x''_1, \ldots, x'_n, x''_n, \ldots,$$

also converges to x. It follows that the sequence

$$f_0(x'_0), f_0(x''_0), f_0(x'_1), f_0(x''_1), \ldots, f_0(x'_n), f_0(x''_n), \ldots,$$

is a Cauchy sequence in F, and since F is complete, it converges to some element of F, which implies that the sequences $(f_0(x'_n))$ and $(f_0(x''_n))$ converge to the same limit.

As a summary, we have defined a function $f\colon E \to F$ by

$$f(x) = \lim_{n \mapsto \infty} f_0(x_n),$$

for any sequence (x_n) converging to x, with $x_n \in E_0$. See Figure A.52.

Step 3. The function f extends f_0. Since every element $x \in E_0$ is the limit of the constant sequence (x_n) with $x_n = x$ for all $n \geq 0$, by definition $f(x)$ is the limit of the sequence $(f_0(x_n))$, which is the constant sequence with value $f_0(x)$, so $f(x) = f_0(x)$; that is, f extends f_0.

Step 4. We now prove that f is uniformly continuous. Since f_0 is uniformly continuous, for every $\epsilon > 0$, there is some $\eta > 0$ such that if $a, b \in E_0$ and $d(a, b) \leq \eta$, then $d(f_0(a), f_0(b)) \leq \epsilon$. Consider any two points $x, y \in E$ such that $d(x, y) \leq \eta/2$. We claim that $d(f(x), f(y)) \leq \epsilon$, which shows that f is uniformly continuous.

Let (x_n) be a sequence of points in E_0 converging to x, and let (y_n) be a sequence of points in E_0 converging to y. By the triangle inequality,

$$d(x_n, y_n) \leq d(x_n, x) + d(x, y) + d(y, y_n) = d(x, y) + d(x_n, x) + d(y_n, y),$$

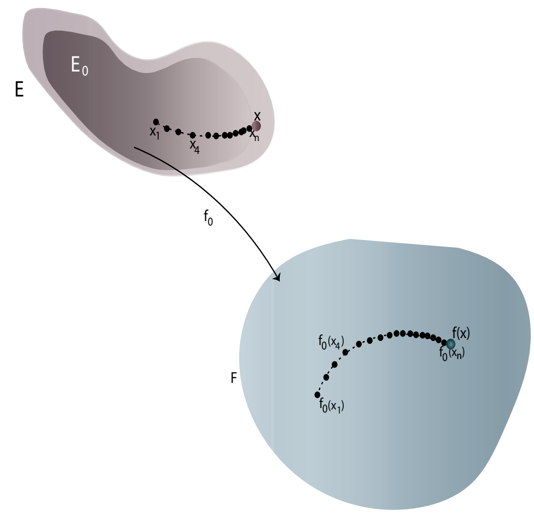


Figure A.52: A schematic illustration of the construction of $f \colon E \to F$ where $f(x) = \lim_{n \mapsto \infty} f_0(x_n)$ for any sequence (x_n) converging to x, with $x_n \in E_0$.

and since (x_n) converges to x and (y_n) converges to y, there is some integer $p > 0$ such that for all $n \geq p$, we have $d(x_n, x) \leq \eta/4$ and $d(y_n, y) \leq \eta/4$, and thus

$$d(x_n, y_n) \leq d(x, y) + \frac{\eta}{2}.$$

Since we assumed that $d(x, y) \leq \eta/2$, we get $d(x_n, y_n) \leq \eta$ for all $n \geq p$, and by uniform continuity of f_0, we get

$$d(f_0(x_n), f_0(y_n)) \leq \epsilon$$

for all $n \geq p$. Since the distance function on F is also continuous, and since $(f_0(x_n))$ converges to $f(x)$ and $(f_0(y_n))$ converges to $f(y)$, we deduce that the sequence $(d(f_0(x_n), f_0(y_n)))$ converges to $d(f(x), f(y))$. This implies that $d(f(x), f(y)) \leq \epsilon$, as desired.

Step 5. It remains to prove that f is unique. Since E_0 is dense in E, for every $x \in E$, there is some sequence (x_n) converging to x, with $x_n \in E_0$. Since f extends f_0 and since f is continuous, we get

$$f(x) = \lim_{n \mapsto \infty} f_0(x_n),$$

which only depends on f_0 and x and shows that f is unique. □

Remark: It can be shown that the theorem no longer holds if we either omit the hypothesis that F is complete or omit that f_0 is uniformly continuous.

For example, if $E_0 \neq E$ and if we let $F = E_0$ and f_0 be the identity function, it is easy to see that f_0 cannot be extended to a continuous function from E to E_0 (for any $x \in E - E_0$, any continuous extension f of f_0 would satisfy $f(x) = x$, which is absurd since $x \notin E_0$).

If f_0 is continuous but not uniformly continuous, a counter-example can be given by using $E = \overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ made into a metric space, $E_0 = \mathbb{R}$, $F = \mathbb{R}$, and f_0 the identity function; for details, see Schwartz [83] (Chapter XI, Section 3, page 134).

Definition A.48. If (E, d_E) and (F, d_F) are two metric spaces, then a function $f \colon E \to F$ is *distance-preserving*, or an *isometry*, if

$$d_F(f(x), f(y)) = d_E(x, y), \quad \text{for all for all } x, y \in E.$$

Observe that an isometry must be injective, because if $f(x) = f(y)$, then $d_F(f(x), f(y)) = 0$, and since $d_F(f(x), f(y)) = d_E(x, y)$, we get $d_E(x, y) = 0$, but $d_E(x, y) = 0$ implies that $x = y$. Also, an isometry is uniformly continuous (since we can pick $\eta = \epsilon$ to satisfy the condition of uniform continuity). However, an isometry is not necessarily surjective.

We now give a construction of the completion of a metric space. This construction is just a generalization of the classical construction of $\mathbb{R}$ from $\mathbb{Q}$ using Cauchy sequences.

Theorem A.62. *Let (E, d) be any metric space. There is a complete metric space $(\widehat{E}, \widehat{d})$ called a completion of (E, d), and a distance-preserving (uniformly continuous) map $\varphi \colon E \to \widehat{E}$ such that $\varphi(E)$ is dense in $\widehat{E}$, and the following extension property holds: for every complete metric space F and for every uniformly continuous function $f \colon E \to F$, there is a unique uniformly continuous function $\widehat{f} \colon \widehat{E} \to F$ such that*

$$f = \widehat{f} \circ \varphi,$$

as illustrated in the following diagram.

$$\begin{tikzcd} E \arrow[r, "\varphi"] \arrow[dr, "f"'] & \widehat{E} \arrow[d, "\widehat{f}"] \\ & F. \end{tikzcd}$$

As a consequence, for any two completions $(\widehat{E}_1, \widehat{d}_1)$ and $(\widehat{E}_2, \widehat{d}_2)$ of (E, d), there is a unique bijective isometry between $(\widehat{E}_1, \widehat{d}_1)$ and $(\widehat{E}_2, \widehat{d}_2)$.

Proof. Consider the set $\mathcal{E}$ of all Cauchy sequences (x_n) in E, and define the relation $\sim$ on $\mathcal{E}$ as follows:

$$(x_n) \sim (y_n) \quad \text{iff} \quad \lim_{n \mapsto \infty} d(x_n, y_n) = 0.$$

It is easy to check that $\sim$ is an equivalence relation on $\mathcal{E}$, and let $\widehat{E} = \mathcal{E}/\sim$ be the quotient set, that is, the set of equivalence classes modulo $\sim$. Our goal is to show that we can endow $\widehat{E}$ with a distance that makes it into a complete metric space satisfying the conditions of the theorem. We proceed in several steps.

Step 1. First let us construct the function $\varphi \colon E \to \widehat{E}$. For every $a \in E$, we have the constant sequence (a_n) such that $a_n = a$ for all $n \geq 0$, which is obviously a Cauchy sequence.

Let $\varphi(a) \in \widehat{E}$ be the equivalence class $[(a_n)]$ of the constant sequence (a_n) with $a_n = a$ for all n. By definition of $\sim$, the equivalence class $\varphi(a)$ is also the equivalence class of all sequences converging to a. The map $a \mapsto \varphi(a)$ is injective because a metric space is Hausdorff, so if $a \neq b$, then a sequence converging to a does not converge to b. After having defined a distance on $\widehat{E}$, we will check that φ is an isometry.

Step 2. Let us now define a distance on $\widehat{E}$. Let $\alpha = [(a_n)]$ and $\beta = [(b_n)]$ be two equivalence classes of Cauchy sequences in E. The triangle inequality implies that

$$d(a_m, b_m) \leq d(a_m, a_n) + d(a_n, b_n) + d(b_n, b_m) = d(a_n, b_n) + d(a_m, a_n) + d(b_m, b_n)$$

and

$$d(a_n, b_n) \leq d(a_n, a_m) + d(a_m, b_m) + d(b_m, b_n) = d(a_m, b_m) + d(a_m, a_n) + d(b_m, b_n),$$

which implies that

$$|d(a_m, b_m) - d(a_n, b_n)| \leq d(a_m, a_n) + d(b_m, b_n).$$

Since (a_n) and (b_n) are Cauchy sequences, the above inequality shows that $(d(a_n, b_n))$ is a Cauchy sequence of nonnegative reals. Since $\mathbb{R}$ is complete, the sequence $(d(a_n, b_n))$ has a limit, which we denote by $\widehat{d}(\alpha, \beta)$; that is, we set

$$\widehat{d}(\alpha, \beta) = \lim_{n \mapsto \infty} d(a_n, b_n), \quad \alpha = [(a_n)],\ \beta = [(b_n)].$$

See Figure A.53.

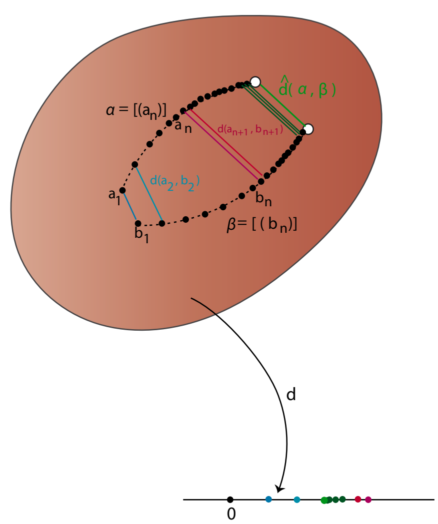


Figure A.53: A schematic illustration of $\widehat{d}(\alpha, \beta)$ from the Cauchy sequence $(d(a_n, b_n))$.

Step 3. Let us check that $\widehat{d}(\alpha, \beta)$ does not depend on the Cauchy sequences (a_n) and (b_n) chosen in the equivalence classes α and β.

If $(a_n) \sim (a'_n)$ and $(b_n) \sim (b'_n)$, then $\lim_{n\mapsto\infty} d(a_n, a'_n) = 0$ and $\lim_{n\mapsto\infty} d(b_n, b'_n) = 0$, and since

$$d(a'_n, b'_n) \leq d(a'_n, a_n) + d(a_n, b_n) + d(b_n, b'_n) = d(a_n, b_n) + d(a_n, a'_n) + d(b_n, b'_n),$$

and

$$d(a_n, b_n) \leq d(a_n, a'_n) + d(a'_n, b'_n) + d(b'_n, b_n) = d(a'_n, b'_n) + d(a_n, a'_n) + d(b_n, b'_n),$$

we have

$$|d(a_n, b_n) - d(a'_n, b'_n)| \leq d(a_n, a'_n) + d(b_n, b'_n),$$

so we have $\lim_{n\mapsto\infty} d(a'_n, b'_n) = \lim_{n\mapsto\infty} d(a_n, b_n) = \widehat{d}(\alpha, \beta)$. Therefore, $\widehat{d}(\alpha, \beta)$ is indeed well defined.

Step 4. Let us check that φ is indeed an isometry.

Given any two elements $\varphi(a)$ and $\varphi(b)$ in $\widehat{E}$, since they are the equivalence classes of the constant sequences (a_n) and (b_n) such that $a_n = a$ and $b_n = b$ for all n, the constant sequence $(d(a_n, b_n))$ with $d(a_n, b_n) = d(a, b)$ for all n converges to $d(a, b)$, so by definition $\widehat{d}(\varphi(a), \varphi(b)) = \lim_{n\mapsto\infty} d(a_n, b_n) = d(a, b)$, which shows that φ is an isometry.

Step 5. Let us verify that $\widehat{d}$ is a metric on $\widehat{E}$. By definition it is obvious that $\widehat{d}(\alpha, \beta) = \widehat{d}(\beta, \alpha)$. If α and β are two distinct equivalence classes, then for any Cauchy sequence (a_n) in the equivalence class α and for any Cauchy sequence (b_n) in the equivalence class β, the sequences (a_n) and (b_n) are inequivalent, which means that $\lim_{n\mapsto\infty} d(a_n, b_n) \neq 0$, that is, $\widehat{d}(\alpha, \beta) \neq 0$. Obviously, $\widehat{d}(\alpha, \alpha) = 0$.

For any equivalence classes $\alpha = [(a_n)]$, $\beta = [(b_n)]$, and $\gamma = [(c_n)]$, we have the triangle inequality

$$d(a_n, c_n) \leq d(a_n, b_n) + d(b_n, c_n),$$

so by continuity of the distance function, by passing to the limit, we obtain

$$\widehat{d}(\alpha, \gamma) \leq \widehat{d}(\alpha, \beta) + \widehat{d}(\beta, \gamma),$$

which is the triangle inequality for $\widehat{d}$. Therefore, $\widehat{d}$ is a distance on $\widehat{E}$.

Step 6. Let us prove that $\varphi(E)$ is dense in $\widehat{E}$. For any $\alpha = [(a_n)]$, let (x_n) be the constant sequence such that $x_k = a_n$ for all $k \geq 0$, so that $\varphi(a_n) = [(x_n)]$. Then we have

$$\widehat{d}(\alpha, \varphi(a_n)) = \lim_{m\mapsto\infty} d(a_m, a_n) \leq \sup_{p,q\geq n} d(a_p, a_q).$$

Since (a_n) is a Cauchy sequence, $\sup_{p,q\geq n} d(a_p, a_q)$ tends to 0 as n goes to infinity, so

$$\lim_{n\mapsto\infty} d(\alpha, \varphi(a_n)) = 0,$$

which means that the sequence $(\varphi(a_n))$ converge to α, and $\varphi(E)$ is indeed dense in $\widehat{E}$.

Step 7. Finally let us prove that the metric space $\widehat{E}$ is complete.

Let (α_n) be a Cauchy sequence in $\widehat{E}$. Since $\varphi(E)$ is dense in $\widehat{E}$, for every $n > 0$, there some $a_n \in E$ such that

$$\widehat{d}(\alpha_n, \varphi(a_n)) \leq \frac{1}{n}.$$

Since

$$\widehat{d}(\varphi(a_m), \varphi(a_n)) \leq \widehat{d}(\varphi(a_m), \alpha_m) + \widehat{d}(\alpha_m, \alpha_n) + \widehat{d}(\alpha_n, \varphi(a_n)) \leq \widehat{d}(\alpha_m, \alpha_n) + \frac{1}{m} + \frac{1}{n},$$

and since (α_m) is a Cauchy sequence, so is $(\varphi(a_n))$, and as φ is an isometry, the sequence (a_n) is a Cauchy sequence in E. Let $\alpha \in \widehat{E}$ be the equivalence class of (a_n). Since

$$\widehat{d}(\alpha, \varphi(a_n)) = \lim_{m \mapsto \infty} d(a_m, a_n)$$

and (a_n) is a Cauchy sequence, we deduce that the sequence $(\varphi(a_n))$ converges to α, and since $d(\alpha_n, \varphi(a_n)) \leq 1/n$ for all $n > 0$, the sequence (α_n) also converges to α.

Step 8. Let us prove the extension property. Let F be any complete metric space and let $f \colon E \to F$ be any uniformly continuous function. The function $\varphi \colon E \to \widehat{E}$ is an isometry and a bijection between E and its image $\varphi(E)$, so its inverse $\varphi^{-1} \colon \varphi(E) \to E$ is also an isometry, and thus is uniformly continuous. If we let $g = f \circ \varphi^{-1}$, then $g \colon \varphi(E) \to F$ is a uniformly continuous function, and $\varphi(E)$ is dense in $\widehat{E}$, so by Theorem A.61 there is a unique uniformly continuous function $\widehat{f} \colon \widehat{E} \to F$ extending $g = f \circ \varphi^{-1}$; see the diagram below:

$$\begin{tikzcd}
E \arrow[drr, "f"'] & \varphi(E) \arrow[l, "\varphi^{-1}"'] \arrow[dr, "g"] & \subseteq & \widehat{E} \arrow[dl, "\widehat{f}"] \\
& & F &
\end{tikzcd}.$$

This means that

$$\widehat{f}|\varphi(E) = f \circ \varphi^{-1},$$

which implies that

$$(\widehat{f}|\varphi(E)) \circ \varphi = f,$$

that is, $f = \widehat{f} \circ \varphi$, as illustrated in the diagram below:

$$\begin{tikzcd}
E \arrow[r, "\varphi"] \arrow[dr, "f"'] & \widehat{E} \arrow[d, "\widehat{f}"] \\
& F.
\end{tikzcd}$$

If $h \colon \widehat{E} \to F$ is any other uniformly continuous function such that $f = h \circ \varphi$, then $g = f \circ \varphi^{-1} = h|\varphi(E)$, so h is a uniformly continuous function extending g, and by Theorem A.61, we have have $h = \widehat{f}$, so $\widehat{f}$ is indeed unique.

Step 9. Uniqueness of the completion $(\widehat{E}, \widehat{d})$ up to a bijective isometry.

Let $(\widehat{E}_1, \widehat{d}_1)$ and $(\widehat{E}_2, \widehat{d}_2)$ be any two completions of (E, d). Then we have two uniformly continuous isometries $\varphi_1 \colon E \to \widehat{E}_1$ and $\varphi_2 \colon E \to \widehat{E}_2$, so by the unique extension property, there exist unique uniformly continuous maps $\widehat{\varphi_2} \colon \widehat{E}_1 \to \widehat{E}_2$ and $\widehat{\varphi_1} \colon \widehat{E}_2 \to \widehat{E}_1$ such that the following diagrams commute:

$$\begin{tikzcd} E \arrow[r, "\varphi_1"] \arrow[dr, "\varphi_2"'] & \widehat{E}_1 \arrow[d, "\widehat{\varphi_2}"] \\ & \widehat{E}_2 \end{tikzcd} \qquad \begin{tikzcd} E \arrow[r, "\varphi_2"] \arrow[dr, "\varphi_1"'] & \widehat{E}_2 \arrow[d, "\widehat{\varphi_1}"] \\ & \widehat{E}_1. \end{tikzcd}$$

Consequently we have the following commutative diagrams:

$$\begin{tikzcd} & \widehat{E}_2 \arrow[d, "\widehat{\varphi_1}"] \\ E \arrow[ur, "\varphi_2"] \arrow[r, "\varphi_1"] \arrow[dr, "\varphi_2"'] & \widehat{E}_1 \arrow[d, "\widehat{\varphi_2}"] \\ & \widehat{E}_2 \end{tikzcd} \qquad \begin{tikzcd} & \widehat{E}_1 \arrow[d, "\widehat{\varphi_2}"] \\ E \arrow[ur, "\varphi_1"] \arrow[r, "\varphi_2"] \arrow[dr, "\varphi_1"'] & \widehat{E}_2 \arrow[d, "\widehat{\varphi_1}"] \\ & \widehat{E}_1. \end{tikzcd}$$

However, $\mathrm{id}_{\widehat{E}_1}$ and $\mathrm{id}_{\widehat{E}_2}$ are uniformly continuous functions making the following diagrams commute

$$\begin{tikzcd} E \arrow[r, "\varphi_1"] \arrow[dr, "\varphi_1"'] & \widehat{E}_1 \arrow[d, "\mathrm{id}_{\widehat{E}_1}"] \\ & \widehat{E}_1 \end{tikzcd} \qquad \begin{tikzcd} E \arrow[r, "\varphi_2"] \arrow[dr, "\varphi_2"'] & \widehat{E}_2 \arrow[d, "\mathrm{id}_{\widehat{E}_2}"] \\ & \widehat{E}_2, \end{tikzcd}$$

so by the uniqueness of extensions we must have

$$\widehat{\varphi_1} \circ \widehat{\varphi_2} = \mathrm{id}_{\widehat{E}_1} \quad \text{and} \quad \widehat{\varphi_2} \circ \widehat{\varphi_1} = \mathrm{id}_{\widehat{E}_2}.$$

This proves that $\widehat{\varphi_1}$ and $\widehat{\varphi_2}$ are mutual inverses. Now since $\varphi_2 = \widehat{\varphi_2} \circ \varphi_1$, we have

$$\widehat{\varphi_2} | \varphi_1(E) = \varphi_2 \circ \varphi_1^{-1},$$

and since φ_1^{-1} and φ_2 are isometries, so is $\widehat{\varphi_2} | \varphi_1(E)$. But we showed in Step 8 that $\widehat{\varphi_2}$ is the uniform continuous extension of $\widehat{\varphi_2} | \varphi_1(E)$ and $\varphi_1(E)$ is dense in $\widehat{E}_1$, so for any two elements $\alpha, \beta \in \widehat{E}_1$, if (a_n) and (b_n) are sequences in $\varphi_1(E)$ converging to α and β, we have

$$\widehat{d}_2((\widehat{\varphi_2} | \varphi_1(E))(a_n), ((\widehat{\varphi_2} | \varphi_1(E))(b_n)) = \widehat{d}_1(a_n, b_n),$$

and by passing to the limit we get

$$\widehat{d}_2(\widehat{\varphi_2}(\alpha), \widehat{\varphi_2}(\beta)) = \widehat{d}_1(\alpha, \beta),$$

which shows that $\widehat{\varphi_2}$ is an isometry (similarly, $\widehat{\varphi_1}$ is an isometry). $\square$

Remarks:

1. Except for Step 8 and Step 9, the proof of Theorem A.62 is the proof given in Schwartz [83] (Chapter XI, Section 4, Theorem 1), and Kormogorov and Fomin [58] (Chapter 2, Section 7, Theorem 4).

2. The construction of $\widehat{E}$ relies on the completeness of $\mathbb{R}$, and so it cannot be used to construct $\mathbb{R}$ from $\mathbb{Q}$. However, this construction can be modified to yield a construction of $\mathbb{R}$ from $\mathbb{Q}$.

We show in Section A.13 that Theorem A.62 yields a construction of the completion of a normed vector space.

A.11 The Contraction Mapping Theorem

If (E, d) is a nonempty complete metric space, every map, $f \colon E \to E$, for which there is some k such that $0 \leq k < 1$ and

$$d(f(x), f(y)) \leq kd(x, y)$$

for all $x, y \in E$, has the very important property that it has a unique fixed point, that is, there is a unique, $a \in E$, such that $f(a) = a$. A map as above is called a *contraction mapping*. Furthermore, the fixed point of a contraction mapping can be computed as the limit of a fast converging sequence.

The fixed point property of contraction mappings is used to show some important theorems of analysis, such as the implicit function theorem and the existence of solutions to certain differential equations. It can also be used to show the existence of fractal sets defined in terms of iterated function systems. Since the proof is quite simple, we prove the fixed point property of contraction mappings. First, observe that a contraction mapping is (uniformly) continuous.

Proposition A.63. *If (E, d) is a nonempty complete metric space, every contraction mapping, $f \colon E \to E$, has a unique fixed point. Furthermore, for every $x_0 \in E$, defining the sequence, (x_n), such that $x_{n+1} = f(x_n)$, the sequence, (x_n), converges to the unique fixed point of f.*

Proof. First we prove that f has at most one fixed point. Indeed, if $f(a) = a$ and $f(b) = b$, since

$$d(a, b) = d(f(a), f(b)) \leq kd(a, b)$$

and $0 \leq k < 1$, we must have $d(a, b) = 0$, that is, $a = b$.

Next, we prove that (x_n) is a Cauchy sequence. Observe that

$$\begin{array}{rcl} d(x_2, x_1) & \leq & kd(x_1, x_0), \\ d(x_3, x_2) & \leq & kd(x_2, x_1) \leq k^2 d(x_1, x_0), \\ \vdots & & \vdots \\ d(x_{n+1}, x_n) & \leq & kd(x_n, x_{n-1}) \leq \cdots \leq k^n d(x_1, x_0). \end{array}$$

Thus, we have

$$\begin{array}{rcl} d(x_{n+p}, x_n) & \leq & d(x_{n+p}, x_{n+p-1}) + d(x_{n+p-1}, x_{n+p-2}) + \cdots + d(x_{n+1}, x_n) \\ & \leq & (k^{p-1} + k^{p-2} + \cdots + k + 1) k^n d(x_1, x_0) \\ & \leq & \dfrac{k^n}{1-k} d(x_1, x_0). \end{array}$$

We conclude that $d(x_{n+p}, x_n)$ converges to 0 when n goes to infinity, which shows that (x_n) is a Cauchy sequence. Since E is complete, the sequence (x_n) has a limit, a. Since f is continuous, the sequence $(f(x_n))$ converges to $f(a)$. But $x_{n+1} = f(x_n)$ converges to a and so $f(a) = a$, the unique fixed point of f. $\square$

Note that no matter how the starting point x_0 of the sequence (x_n) is chosen, (x_n) converges to the unique fixed point of f. Also, the convergence is fast, since

$$d(x_n, a) \leq \frac{k^n}{1-k} d(x_1, x_0).$$

The Hausdorff distance between compact subsets of a metric space provides a very nice illustration of some of the theorems on complete and compact metric spaces just presented.

Definition A.49. Given a metric space, (X, d), for any subset, $A \subseteq X$, for any, $\epsilon \geq 0$, define the *ϵ-hull* of A as the set

$$V_\epsilon(A) = \{x \in X,\ \exists a \in A \mid d(a, x) \leq \epsilon\}.$$

See Figure A.54. Given any two nonempty bounded subsets, A, B of X, define $D(A, B)$, *the Hausdorff distance between A and B*, by

$$D(A, B) = \inf\{\epsilon \geq 0 \mid A \subseteq V_\epsilon(B) \text{ and } B \subseteq V_\epsilon(A)\}.$$

Note that since we are considering nonempty bounded subsets, $D(A, B)$ is well defined (i.e., not infinite). However, D is not necessarily a distance function. It is a distance function if we restrict our attention to nonempty compact subsets of X (actually, it is also a metric on closed and bounded subsets). We let $\mathcal{K}(X)$ denote the set of all nonempty compact subsets of X. The remarkable fact is that D is a distance on $\mathcal{K}(X)$ and that if X is complete or compact, then so is $\mathcal{K}(X)$. The following theorem is taken from Edgar [31].

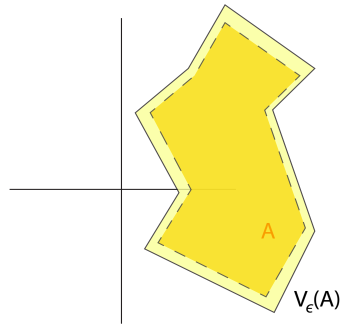


Figure A.54: The ϵ-hull of a polygonal region A of $\mathbb{R}^2$

Theorem A.64. *If (X, d) is a metric space, then the Hausdorff distance, D, on the set, $\mathcal{K}(X)$, of nonempty compact subsets of X is a distance. If (X, d) is complete, then $(\mathcal{K}(X), D)$ is complete and if (X, d) is compact, then $(\mathcal{K}(X), D)$ is compact.*

Proof. Since (nonempty) compact sets are bounded, $D(A, B)$ is well defined. Clearly D is symmetric. Assume that $D(A, B) = 0$. Then for every $\epsilon > 0$, $A \subseteq V_\epsilon(B)$, which means that for every $a \in A$, there is some $b \in B$ such that $d(a, b) \leq \epsilon$, and thus, that $A \subseteq \overline{B}$. Since Proposition A.26 implies that B is closed, $\overline{B} = B$, and we have $A \subseteq B$. Similarly, $B \subseteq A$, and thus, $A = B$. Clearly, if $A = B$, we have $D(A, B) = 0$. It remains to prove the triangle inequality. Assume that $D(A, B) \leq \epsilon_1$ and that $D(B, C) \leq \epsilon_2$. We must show that $D(A, C) \leq \epsilon_1 + \epsilon_2$. This will be accomplished if we can show that $C \subseteq V_{\epsilon_1+\epsilon_2}(A)$ and $A \subseteq V_{\epsilon_1+\epsilon_2}(C)$. By assumption and definition of D, $B \subseteq V_{\epsilon_1}(A)$ and $C \subseteq V_{\epsilon_2}(B)$. Then

$$V_{\epsilon_2}(B) \subseteq V_{\epsilon_2}(V_{\epsilon_1}(A)),$$

and since a basic application of the triangle inequality implies that

$$V_{\epsilon_2}(V_{\epsilon_1}(A)) \subseteq V_{\epsilon_1+\epsilon_2}(A),$$

we get

$$C \subseteq V_{\epsilon_2}(B) \subseteq V_{\epsilon_1+\epsilon_2}(A).$$

See Figure A.55.

Similarly, the conditions $(A, B) \leq \epsilon_1$ and $D(B, C) \leq \epsilon_2$ imply that

$$A \subseteq V_{\epsilon_1}(B), \qquad B \subseteq V_{\epsilon_2}(C).$$

Hence

$$A \subseteq V_{\epsilon_1}(B) \subseteq V_{\epsilon_1}(V_{\epsilon_2}(C)) \subseteq V_{\epsilon_1+\epsilon_2}(C),$$

and thus the triangle inequality follows.

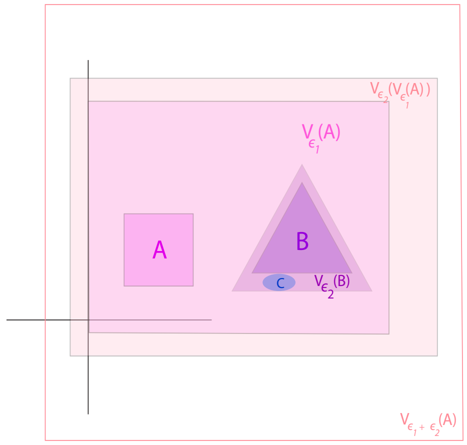


Figure A.55: Let A be the small pink square and B be the small purple triangle in $\mathbb{R}^2$. The periwinkle oval C is contained in $V_{\epsilon_1+\epsilon_2}(A)$.

Next we need to prove that if (X, d) is complete, then $(\mathcal{K}(X), D)$ is also complete. First we show that if (A_n) is a sequence of nonempty compact sets converging to a nonempty compact set A in the Hausdorff metric, then

$$A = \{x \in X \mid \text{there is a sequence, } (x_n), \text{ with } x_n \in A_n \text{ converging to } x\}.$$

Indeed, if (x_n) is a sequence with $x_n \in A_n$ converging to x and (A_n) converges to A then, for every $\epsilon > 0$, there is some x_n such that $d(x_n, x) \leq \epsilon/2$ and there is some $a_n \in A$ such that $d(a_n, x_n) \leq \epsilon/2$ and thus, $d(a_n, x) \leq \epsilon$, which shows that $x \in \overline{A}$. Since A is compact, it is closed, and $x \in A$. See Figure A.56.

Conversely, since (A_n) converges to A, for every $x \in A$, for every $n \geq 1$, there is some $x_n \in A_n$ such that $d(x_n, x) \leq 1/n$ and the sequence (x_n) converges to x.

Now let (A_n) be a Cauchy sequence in $\mathcal{K}(X)$. It can be proven that (A_n) converges to the set

$$A = \{x \in X \mid \text{there is a sequence, } (x_n), \text{ with } x_n \in A_n \text{ converging to } x\},$$

and that A is nonempty and compact. To prove that A is compact, one proves that it is totally bounded and complete. Details are given in Edgar [31].

Finally we need to prove that if (X, d) is compact, then $(\mathcal{K}(X), D)$ is compact. Since we already know that $(\mathcal{K}(X), D)$ is complete if (X, d) is, it is enough to prove that $(\mathcal{K}(X), D)$ is totally bounded if (X, d) is, which is not hard. $\square$

In view of Theorem A.64 and Theorem A.63, it is possible to define some nonempty compact subsets of X in terms of fixed points of contraction maps. This can be done in

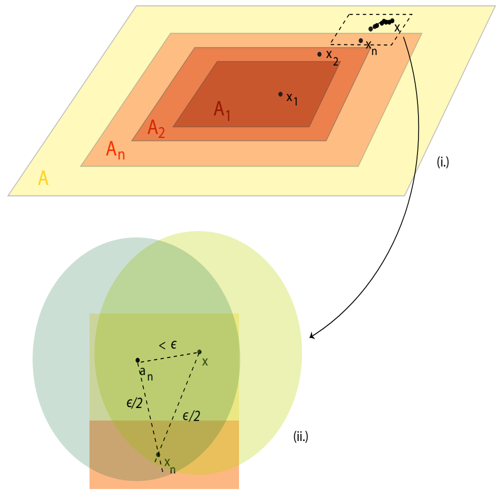


Figure A.56: Let (A_n) be the sequence of parallelograms converging to A, the large pale yellow parallelogram. Figure $(ii.)$ expands the dashed region and shows why $d(a_n, x) < \epsilon$.

terms of iterated function systems, yielding a large class of fractals. However, we will omit this topic and instead refer the reader to Edgar [31].

Before considering differentials, we need to look at the continuity of linear maps.

A.12 Continuous Linear and Multilinear Maps

If E and F are normed vector spaces, we first characterize when a linear map $f\colon E \to F$ is continuous.

Proposition A.65. *Given two normed vector spaces E and F, for any linear map $f\colon E \to F$, the following conditions are equivalent:*

(1) The function f is continuous at 0.

(2) There is a constant $k \geq 0$ such that,

$$\|f(u)\| \leq k, \ \textit{for every } u \in E \textit{ such that } \|u\| \leq 1.$$

(3) There is a constant $k \geq 0$ such that,

$$\|f(u)\| \leq k\|u\|, \ \textit{for every } u \in E.$$

(4) The function f is continuous at every point of E.

Proof. Assume (1). Then for every $\epsilon > 0$, there is some $\eta > 0$ such that, for every $u \in E$, if $\|u\| \leq \eta$, then $\|f(u)\| \leq \epsilon$. Pick $\epsilon = 1$, so that there is some $\eta > 0$ such that, if $\|u\| \leq \eta$, then $\|f(u)\| \leq 1$. If $\|u\| \leq 1$, then $\|\eta u\| \leq \eta\|u\| \leq \eta$, and so, $\|f(\eta u)\| \leq 1$, that is, $\eta\|f(u)\| \leq 1$, which implies $\|f(u)\| \leq \eta^{-1}$. Thus, (2) holds with $k = \eta^{-1}$.

Assume that (2) holds. If $u = 0$, then by linearity, $f(0) = 0$, and thus $\|f(0)\| \leq k\|0\|$ holds trivially for all $k \geq 0$. If $u \neq 0$, then $\|u\| > 0$, and since

$$\left\| \frac{u}{\|u\|} \right\| = 1,$$

we have

$$\left\| f\left(\frac{u}{\|u\|}\right) \right\| \leq k,$$

which implies that

$$\|f(u)\| \leq k\|u\|.$$

Thus, (3) holds.

If (3) holds, then for all $u, v \in E$, we have

$$\|f(v) - f(u)\| = \|f(v - u)\| \leq k\|v - u\|.$$

If $k = 0$, then f is the zero function, and continuity is obvious. Otherwise, if $k > 0$, for every $\epsilon > 0$, if $\|v - u\| \leq \frac{\epsilon}{k}$, then $\|f(v - u)\| \leq \epsilon$, which shows continuity at every $u \in E$. Finally, it is obvious that (4) implies (1). ☐

Among other things, Proposition A.65 shows that a linear map is continuous iff the image of the unit (closed) ball is bounded. Since a continuous linear map satisfies the condition $\|f(u)\| \leq k\|u\|$ (for some $k \geq 0$), it is also uniformly continuous.

If E and F are normed vector spaces, the set of all continuous linear maps $f\colon E \to F$ is denoted by $\mathcal{L}(E; F)$.

Using Proposition A.65, we can define a norm on $\mathcal{L}(E; F)$ which makes it into a normed vector space. This definition has already been given in Chapter B (Definition B.7) but for the reader's convenience, we repeat it here.

Definition A.50. Given two normed vector spaces E and F, for every continuous linear map $f\colon E \to F$, we define the *operator norm* $\|f\|$ *of* f as

$$\begin{aligned} \|f\| &= \inf\{k \geq 0 \mid \|f(x)\| \leq k\|x\|, \text{ for all } x \in E\} \\ &= \sup\{\|f(x)\| \mid \|x\| \leq 1\} \\ &= \sup\{\|f(x)\| \mid \|x\| = 1\}. \end{aligned}$$

From Definition A.50, for every continuous linear map $f \in \mathcal{L}(E; F)$, we have

$$\|f(x)\| \leq \|f\|\|x\|,$$

for every $x \in E$. It is easy to verify that $\mathcal{L}(E; F)$ is a normed vector space under the norm of Definition A.50. Furthermore, if E, F, G, are normed vector spaces, and $f\colon E \to F$ and $g\colon F \to G$ are continuous linear maps, we have

$$\|g \circ f\| \leq \|g\|\|f\|.$$

We can now show that when $E = \mathbb{R}^n$ or $E = \mathbb{C}^n$, with any of the norms $\| \, \|_1$, $\| \, \|_2$, or $\| \, \|_\infty$, then every linear map $f\colon E \to F$ is continuous.

Proposition A.66. *If $E = \mathbb{R}^n$ or $E = \mathbb{C}^n$, with any of the norms $\| \, \|_1$, $\| \, \|_2$, or $\| \, \|_\infty$, and F is any normed vector space, then every linear map $f\colon E \to F$ is continuous.*

Proof. Let $(e_1, \ldots, e_n)$ be the standard basis of $\mathbb{R}^n$ (a similar proof applies to $\mathbb{C}^n$). In view of Proposition B.2, it is enough to prove the proposition for the norm

$$\|x\|_\infty = \max\{|x_i| \mid 1 \leq i \leq n\}.$$

We have,

$$\|f(v) - f(u)\| = \|f(v - u)\| = \left\| f(\sum_{1 \leq i \leq n} (v_i - u_i)e_i) \right\| = \left\| \sum_{1 \leq i \leq n} (v_i - u_i) f(e_i) \right\|,$$

and so,

$$\|f(v) - f(u)\| \leq \Big(\sum_{1 \leq i \leq n} \|f(e_i)\| \Big) \max_{1 \leq i \leq n} |v_i - u_i| = \Big(\sum_{1 \leq i \leq n} \|f(e_i)\| \Big) \|v - u\|_\infty.$$

By the argument used in Proposition A.65 to prove that (3) implies (4), f is continuous. $\square$

Actually, we proved in Theorem B.3 that if E is a vector space of finite dimension, then any two norms are equivalent, so that they define the same topology. This fact together with Proposition A.66 prove the following:

Theorem A.67. *If E is a vector space of finite dimension (over $\mathbb{R}$ or $\mathbb{C}$), then all norms are equivalent (define the same topology). Furthermore, for any normed vector space F, every linear map $f\colon E \to F$ is continuous.*

If E is a normed vector space of infinite dimension, a linear map $f\colon E \to F$ may not be continuous. As an example, let E be the infinite vector space of all polynomials over $\mathbb{R}$. Let

$$\|P(X)\| = \max_{0 \leq x \leq 1} |P(x)|.$$

We leave as an exercise to show that this is indeed a norm. Let $F = \mathbb{R}$, and let $f\colon E \to F$ be the map defined such that, $f(P(X)) = P(3)$. It is clear that f is linear. Consider the sequence of polynomials

$$P_n(X) = \left(\frac{X}{2}\right)^n.$$

It is clear that $\|P_n\| = \left(\frac{1}{2}\right)^n$, and thus, the sequence P_n has the null polynomial as a limit. However, we have

$$f(P_n(X)) = P_n(3) = \left(\frac{3}{2}\right)^n,$$

and the sequence $f(P_n(X))$ diverges to $+\infty$. Consequently, in view of Proposition A.15 (1), f is not continuous.

We now consider the continuity of multilinear maps. We treat explicitly bilinear maps, the general case being a straightforward extension.

Proposition A.68. *Given normed vector spaces E, F and G, for any bilinear map $f\colon E \times F \to G$, the following conditions are equivalent:*

(1) The function f is continuous at $\langle 0, 0\rangle$.

2) There is a constant $k \geq 0$ such that,

$$\|f(u, v)\| \leq k, \text{ for all } u \in E, v \in F \text{ such that } \|u\|, \|v\| \leq 1.$$

3) There is a constant $k \geq 0$ such that,

$$\|f(u, v)\| \leq k\|u\|\|v\|, \text{ for all } u \in E, v \in F.$$

4) The function f is continuous at every point of $E \times F$.

Proof. It is similar to that of Proposition A.65, with a small subtlety in proving that (3) implies (4), namely that two different η's that are not independent are needed. $\square$

In contrast to continuous linear maps, which must be uniformly continuous, nonzero continuous bilinear maps are **not** uniformly continuous. Let $f\colon E \times F \to G$ be a continuous bilinear map such that $f(a, b) \neq 0$ for some $a \in E$ and some $b \in F$. Consider the sequences (u_n) and (v_n) (with $n \geq 1$) given by

$$\begin{aligned} u_n &= (x_n, y_n) = (na, nb) \\ v_n &= (x'_n, y'_n) = \left(\left(n + \frac{1}{n}\right)a, \left(n + \frac{1}{n}\right)b\right). \end{aligned}$$

Obviously

$$\|v_n - u_n\| \leq \frac{1}{n}(\|a\| + \|b\|),$$

so $\lim_{n\mapsto\infty} \|v_n - u_n\| = 0$. On the other hand

$$f(x_n', y_n') - f(x_n, y_n) = \left(2 + \frac{1}{n^2}\right) f(a, b),$$

and thus $\lim_{n\mapsto\infty} \|f(x_n', y_n') - f(x_n, y_n)\| = 2\,\|f(a,b)\| \neq 0$, which shows that f is not uniformly continuous, because if this was the case, this limit would be zero.

If E, F, and G, are normed vector spaces, we denote the set of all continuous bilinear maps $f\colon E \times F \to G$ by $\mathcal{L}_2(E, F; G)$. Using Proposition A.68, we can define a norm on $\mathcal{L}_2(E, F; G)$ which makes it into a normed vector space.

Definition A.51. Given normed vector spaces E, F, and G, for every continuous bilinear map $f\colon E \times F \to G$, we define the *norm* $\|f\|$ *of* f as

$$\begin{aligned}\|f\| &= \inf\left\{k \geq 0 \mid \|f(x,y)\| \leq k\|x\|\|y\|, \text{ for all } x \in E, y \in F\right\} \\ &= \sup\left\{\|f(x,y)\| \mid \|x\|, \|y\| \leq 1\right\}.\end{aligned}$$

From Definition A.50, for every continuous bilinear map $f \in \mathcal{L}_2(E, F; G)$, we have

$$\|f(x,y)\| \leq \|f\|\|x\|\|y\|,$$

for all $x \in E, y \in F$. It is easy to verify that $\mathcal{L}_2(E, F; G)$ is a normed vector space under the norm of Definition A.51.

Given a bilinear map $f\colon E \times F \to G$, for every $u \in E$, we obtain a linear map denoted $fu\colon F \to G$, defined such that, $fu(v) = f(u, v)$. Furthermore, since

$$\|f(x,y)\| \leq \|f\|\|x\|\|y\|,$$

it is clear that fu is continuous. We can then consider the map $\varphi\colon E \to \mathcal{L}(F; G)$, defined such that, $\varphi(u) = fu$, for any $u \in E$, or equivalently, such that,

$$\varphi(u)(v) = f(u, v).$$

Actually, it is easy to show that φ is linear and continuous, and that $\|\varphi\| = \|f\|$. Thus, $f \mapsto \varphi$ defines a map from $\mathcal{L}_2(E, F; G)$ to $\mathcal{L}(E; \mathcal{L}(F; G))$. We can also go back from $\mathcal{L}(E; \mathcal{L}(F; G))$ to $\mathcal{L}_2(E, F; G)$. We summarize all this in the following proposition.

Proposition A.69. *Let E, F, G be three normed vector spaces. The map $f \mapsto \varphi$, from $\mathcal{L}_2(E, F; G)$ to $\mathcal{L}(E; \mathcal{L}(F; G))$, defined such that, for every $f \in \mathcal{L}_2(E, F; G)$,*

$$\varphi(u)(v) = f(u, v),$$

is an isomorphism of vector spaces, and furthermore, $\|\varphi\| = \|f\|$.

As a corollary of Proposition A.69, we get the following proposition which will be useful when we define second-order derivatives.

Proposition A.70. *Let E, F be normed vector spaces. The map app from $\mathcal{L}(E;F) \times E$ to F, defined such that, for every $f \in \mathcal{L}(E;F)$, for every $u \in E$,*

$$app(f, u) = f(u),$$

is a continuous bilinear map.

Remark: If E and F are nontrivial, it can be shown that $\|\mathrm{app}\| = 1$. It can also be shown that composition

$$\circ \colon \mathcal{L}(E;F) \times \mathcal{L}(F;G) \to \mathcal{L}(E;G),$$

is bilinear and continuous.

The above propositions and definition generalize to arbitrary n-multilinear maps, with $n \geq 2$. Proposition A.68 extends in the obvious way to any n-multilinear map $f \colon E_1 \times \cdots \times E_n \to F$, but condition (3) becomes:

There is a constant $k \geq 0$ such that,

$$\|f(u_1, \ldots, u_n)\| \leq k\|u_1\| \cdots \|u_n\|, \text{ for all } u_1 \in E_1, \ldots, u_n \in E_n.$$

Definition A.51 also extends easily to

$$\begin{aligned}\|f\| &= \inf \{k \geq 0 \mid \|f(x_1, \ldots, x_n)\| \leq k\|x_1\| \cdots \|x_n\|, \text{ for all } x_i \in E_i, 1 \leq i \leq n\} \\ &= \sup \{\|f(x_1, \ldots, x_n)\| \mid \|x_n\|, \ldots, \|x_n\| \leq 1\}.\end{aligned}$$

Proposition A.69 is also easily extended, and we get an isomorphism between continuous n-multilinear maps in $\mathcal{L}_n(E_1, \ldots, E_n; F)$, and continuous linear maps in

$$\mathcal{L}(E_1; \mathcal{L}(E_2; \ldots; \mathcal{L}(E_n; F))).$$

An obvious extension of Proposition A.70 also holds.

Definition A.52. A normed vector space $(E, \| \, \|)$ over $\mathbb{R}$ (or $\mathbb{C}$) which is a complete metric space for the distance $d(u, v) = \|v - u\|$, is called a *Banach space*.

It can be shown that every normed vector space of finite dimension is a Banach space (is complete). This is because $\mathbb{R}$ (and $\mathbb{C}$) are complete. The following theorem is a key result of the theory of Banach spaces worth proving.

Theorem A.71. *If E and F are normed vector spaces, and if F is a Banach space, then $\mathcal{L}(E;F)$ is a Banach space (with the operator norm).*

Proof. Let $(f)_{n\geq 1}$ be a Cauchy sequence of continuous linear maps $f_n\colon E \to F$. We proceed in several steps.

Step 1. Define the pointwise limit $f\colon E \to F$ of the sequence $(f_n)_{n\geq 1}$.

Since $(f)_{n\geq 1}$ is a Cauchy sequence, for every $\epsilon > 0$, there is some $N > 0$ such that $\|f_m - f_n\| < \epsilon$ for all $m, n \geq N$. Since $\| \, \|$ is the operator norm, we deduce that for any $u \in E$, we have

$$\|f_m(u) - f_n(u)\| = \|(f_m - f_n)(u)\| \leq \|f_m - f_n\| \, \|u\| \leq \epsilon \, \|u\| \quad \text{for all } m, n \geq N,$$

that is,

$$\|f_m(u) - f_n(u)\| \leq \epsilon \, \|u\| \quad \text{for all } m, n \geq N. \tag{$*_1$}$$

If $u = 0$, then $f_m(0) = f_n(0) = 0$ for all m, n, so the sequence $(f_n(0))$ is a Cauchy sequence in F converging to 0. If $u \neq 0$, by replacing ϵ by $\epsilon/\|u\|$, we see that the sequence $(f_n(u))$ is a Cauchy sequence in F. Since F is complete, the sequence $(f_n(u))$ has a limit which we denote by $f(u)$. This defines our candidate limit function f by

$$f(u) = \lim_{n \mapsto \infty} f_n(u).$$

It remains to prove that

1. f is linear.
2. f is continous.
3. f is the limit of (f_n) for the operator norm.

Step 2. The function f is linear.

Recall that in a normed vector space, addition and multiplication by a fixed scalar are continuous (since $\|u + v\| \leq \|u\| + \|v\|$ and $\|\lambda u\| \leq |\lambda| \, \|u\|$). Thus by definition of f and since the f_n are linear we have

$$\begin{aligned}
f(u + v) &= \lim_{n \mapsto \infty} f_n(u + v) && \text{by definition of } f \\
&= \lim_{n \mapsto \infty} (f_n(u) + f_n(v)) && \text{by linearity of } f_n \\
&= \lim_{n \mapsto \infty} f_n(u) + \lim_{n \mapsto \infty} f_n(v) && \text{since + is continuous} \\
&= f(u) + f(v) && \text{by definition of } f.
\end{aligned}$$

Similarly,

$$\begin{aligned}
f(\lambda u) &= \lim_{n \mapsto \infty} f_n(\lambda u) && \text{by definition of } f \\
&= \lim_{n \mapsto \infty} \lambda f_n(u) && \text{by linearity of } f_n \\
&= \lambda \lim_{n \mapsto \infty} f_n(u) && \text{by continuity of scalar multiplication} \\
&= \lambda f(u) && \text{by definition of } f.
\end{aligned}$$

Therefore, f is linear.

Step 3. The function f is continuous.

Since $(f_n)_{n\geq 1}$ is a Cauchy sequence, for every $\epsilon > 0$, there is some $N > 0$ such that $\|f_m - f_n\| < \epsilon$ for all $m, n \geq N$. Since $f_m = f_n + f_m - f_n$, we get $\|f_m\| \leq \|f_n\| + \|f_m - f_n\|$, which implies that

$$\|f_m\| \leq \|f_n\| + \epsilon \quad \text{for all } m, n \geq N. \tag{$*_2$}$$

Using $(*_2)$, we also have

$$\|f_m(u)\| \leq \|f_m\| \, \|u\| \leq (\|f_n\| + \epsilon) \, \|u\| \quad \text{for all } m, n \geq N,$$

that is,

$$\|f_m(u)\| \leq (\|f_n\| + \epsilon) \, \|u\| \quad \text{for all } m, n \geq N. \tag{$*_3$}$$

Hold $n \geq N$ fixed and let m tend to $+\infty$ in $(*_3)$. Since the norm is continuous, we get

$$\|f(u)\| \leq (\|f_n\| + \epsilon) \, \|u\| \, ,$$

which shows that f is continuous.

Step 4. The function f is the limit of (f_n) for the operator norm.

Recall $(*_1)$:

$$\|f_m(u) - f_n(u)\| \leq \epsilon \, \|u\| \quad \text{for all } m, n \geq N. \tag{$*_1$}$$

Hold $n \geq N$ fixed but this time let m tend to $+\infty$ in $(*_1)$. By continuity of the norm we get

$$\|f(u) - f_n(u)\| = \|(f - f_n)(u)\| \leq \epsilon \, \|u\| \, .$$

By definition of the operator norm,

$$\|f - f_n\| = \sup\{\|(f - f_n)(u)\| \mid \|u\| = 1\} \leq \epsilon \quad \text{for all } n \geq N,$$

which proves that f_n converges to f for the operator norm. $\square$

As a special case of Theorem A.71, if we let $F = \mathbb{R}$ (or $F = \mathbb{C}$ in the case of complex vector spaces) we see that $E' = \mathcal{L}(E; \mathbb{R})$ (or $E' = \mathcal{L}(E; \mathbb{C})$) is complete (since $\mathbb{R}$ and $\mathbb{C}$ are complete). The space E' of continuous linear forms on E is called the *dual* of E. It is a subspace of the *algebraic dual* E^* of E which consists of *all* linear forms on E, not necessarily continuous.

It can also be shown that if E, F and G are normed vector spaces, and if G is a Banach space, then $\mathcal{L}_2(E, F; G)$ is a Banach space. The proof is essentially identical.

A.13 Completion of a Normed Vector Space

An easy corollary of Theorem A.62 and Theorem A.61 is that every normed vector space can be embedded in a complete normed vector space, that is, a Banach space.

Theorem A.72. *If $(E, \| \|)$ is a normed vector space, then its completion $(\widehat{E}, \widehat{d})$ as a metric space (where E is given the metric $d(x, y) = \|x - y\|$) can be given a unique vector space structure extending the vector space structure on E, and a norm $\| \|_{\widehat{E}}$, so that $(\widehat{E}, \| \|_{\widehat{E}})$ is a Banach space, and the metric $\widehat{d}$ is associated with the norm $\| \|_{\widehat{E}}$. Furthermore, the isometry $\varphi\colon E \to \widehat{E}$ given by Theorem A.62 is a linear isometry, and $\varphi(E)$ is dense in $\widehat{E}$.*

Proof. The addition operation $+\colon E \times E \to E$ is uniformly continuous because

$$\|(u' + v') - (u'' + v'')\| \leq \|u' - u''\| + \|v' - v''\|.$$

It is not hard to show that $\widehat{E} \times \widehat{E}$ is a complete metric space and that $E \times E$ is dense in $\widehat{E} \times \widehat{E}$. Then, by Theorem A.61, the uniformly continuous function $+$ has a unique continuous extension $+\colon \widehat{E} \times \widehat{E} \to \widehat{E}$.

The map $\cdot\colon \mathbb{R} \times E \to E$ is not uniformly continuous, but for any fixed $\lambda \in \mathbb{R}$, the map $L_\lambda\colon E \to E$ given by $L_\lambda(u) = \lambda \cdot u$ is uniformly continuous, so by Theorem A.61 the function L_λ has a unique continuous extension $L_\lambda\colon \widehat{E} \to \widehat{E}$, which we use to define the scalar multiplication $\cdot\colon \mathbb{R} \times \widehat{E} \to \widehat{E}$. It is easily checked that with the above addition and scalar multiplication, $\widehat{E}$ is a vector space.

Since the norm $\| \|$ on E is uniformly continuous, it has a unique continuous extension $\| \|_{\widehat{E}}\colon \widehat{E} \to \mathbb{R}_+$. The identities $\|u + v\| \leq \|u\| + \|v\|$ and $\|\lambda u\| \leq |\lambda| \, \|u\|$ extend to $\widehat{E}$ by continuity. The equation

$$d(u, v) = \|u - v\|$$

also extends to $\widehat{E}$ by continuity and yields

$$\widehat{d}(\alpha, \beta) = \|\alpha - \beta\|_{\widehat{E}},$$

which shows that $\| \|_{\widehat{E}}$ is indeed a norm, and that the metric $\widehat{d}$ is associated to it. Finally, it is easy to verify that the map φ is linear. The uniqueness of the structure of normed vector space follows from the uniqueness of continuous extensions in Theorem A.61. □

Theorem A.72 and Theorem A.61 will be used to show that every Hermitian space can be embedded in a Hilbert space; see Theorem D.1.

The following version of Theorem A.61 for normed vector spaces will be needed in the theory of integration.

Theorem A.73. *Let E and F be two normed vector spaces, let E_0 be a dense subspace of E, and let $f_0\colon E_0 \to F$ be a continuous function. If f_0 is uniformly continuous and if F is complete, then there is a unique uniformly continuous function $f\colon E \to F$ extending f_0. Furthermore, if f_0 is a continuous linear map, then f is also a linear continuous map, and $\|f\| = \|f_0\|$.*

Proof. We only need to prove the second statement. Given any two vectors $x, y \in E$, since E_0 is dense on E we can pick sequences (x_n) and (y_n) of vectors $x_n, y_n \in E_0$ such that $x = \lim_{n\mapsto\infty} x_n$ and $y = \lim_{n\mapsto\infty} y_n$. Since addition and scalar multiplication are continuous, we get

$$\begin{aligned} x + y &= \lim_{n\mapsto\infty} (x_n + y_n) \\ \lambda x &= \lim_{n\mapsto\infty} (\lambda x_n) \end{aligned}$$

for any $\lambda \in \mathbb{R}$ (or $\lambda \in \mathbb{C}$). Since $f(x)$ is defined by

$$f(x) = \lim_{n\mapsto\infty} f_0(x_n)$$

independently of the sequence (x_n) converging to x, and similarly for $f(y)$ and $f(x+y)$, since f_0 is linear, we have

$$\begin{aligned} f(x+y) &= \lim_{n\mapsto\infty} f_0(x_n + y_n) \\ &= \lim_{n\mapsto\infty} (f_0(x_n) + f_0(y_n)) \\ &= \lim_{n\mapsto\infty} f_0(x_n) + \lim_{n\mapsto\infty} f_0(y_n) \\ &= f(x) + f(y). \end{aligned}$$

Similarly,

$$\begin{aligned} f(\lambda x) &= \lim_{n\mapsto\infty} f_0(\lambda x_n) \\ &= \lim_{n\mapsto\infty} \lambda f_0(x_n) \\ &= \lambda \lim_{n\mapsto\infty} f_0(x_n) \\ &= \lambda f(x). \end{aligned}$$

Therefore, f is linear. Since the norm is continuous, we have

$$\|f(x)\| = \left\| \lim_{n\mapsto\infty} f_0(x_n) \right\| = \lim_{n\mapsto\infty} \|f_0(x_n)\|,$$

and since f_0 is continuous

$$\|f_0(x_n)\| \leq \|f_0\| \, \|x_n\| \quad \text{for all } n \geq 1,$$

so we get

$$\lim_{n \mapsto \infty} \|f_0(x_n)\| \leq \lim_{n \mapsto \infty} \|f_0\| \, \|x_n\| \quad \text{for all } n \geq 1,$$

that is,

$$\|f(x)\| \leq \|f_0\| \, \|x\| \, .$$

Since

$$\|f\| = \sup_{\|x\|=1,\, x \in E} \|f(x)\| \, ,$$

we deduce that $\|f\| \leq \|f_0\|$. But since $E_0 \subseteq E$ and f agrees with f_0 on E_0, we also have

$$\|f_0\| = \sup_{\|x\|=1,\, x \in E_0} \|f_0(x)\| = \sup_{\|x\|=1,\, x \in E_0} \|f(x)\| \leq \sup_{\|x\|=1,\, x \in E} \|f(x)\| = \|f\| \, ,$$

and thus $\|f\| = \|f_0\|$. $\square$

A.14 Futher Readings

A thorough treatment of general topology can be found in Munkres [75, 74], Dixmier [27], Lang [63, 62], Schwartz [84, 83], and Bredon [15].

Appendix B

Vector Norms and Matrix Norms

B.1 Normed Vector Spaces

In order to define how close two vectors or two matrices are, and in order to define the convergence of sequences of vectors or matrices, we can use the notion of a norm. Recall that $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$. Also recall that if $z = a + ib \in \mathbb{C}$ is a complex number, with $a, b \in \mathbb{R}$, then $\overline{z} = a - ib$ and $|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}$ ($|z|$ is the *modulus* of z).

Definition B.1. Let E be a vector space over a field K, where K is either the field $\mathbb{R}$ of reals, or the field $\mathbb{C}$ of complex numbers. A *norm* on E is a function $\| \, \| \colon E \to \mathbb{R}_+$, assigning a nonnegative real number $\|u\|$ to any vector $u \in E$, and satisfying the following conditions for all $x, y \in E$:

(N1) $\|x\| \geq 0$, and $\|x\| = 0$ iff $x = 0$. (positivity)

(N2) $\|\lambda x\| = |\lambda| \, \|x\|$. (homogeneity (or scaling))

(N3) $\|x + y\| \leq \|x\| + \|y\|$. (triangle inequality)

A vector space E together with a norm $\| \, \|$ is called a *normed vector space*.

By (N2), setting $\lambda = -1$, we obtain

$$\|-x\| = \|(-1)x\| = |-1| \, \|x\| = \|x\| \,;$$

that is, $\|-x\| = \|x\|$. From (N3), we have

$$\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\| \,,$$

which implies that

$$\|x\| - \|y\| \leq \|x - y\| \,.$$

By exchanging x and y and using the fact that by (N2),

$$\|y - x\| = \|-(x - y)\| = \|x - y\| \,,$$

we also have

$$\|y\| - \|x\| \leq \|x - y\| .$$

Therefore,

$$|\|x\| - \|y\|| \leq \|x - y\|, \quad \text{for all } x, y \in E. \tag{$*$}$$

Observe that setting $\lambda = 0$ in (N2), we deduce that $\|0\| = 0$ without assuming (N1). Then, by setting $y = 0$ in $(*)$, we obtain

$$|\|x\|| \leq \|x\| , \quad \text{for all } x \in E.$$

Therefore, the condition $\|x\| \geq 0$ in (N1) follows from (N2) and (N3), and (N1) can be replaced by the weaker condition

(N1') For all $x \in E$, if $\|x\| = 0$ then $x = 0$,

A function $\| \, \| : E \to \mathbb{R}$ satisfying axioms (N2) and (N3) is called a *semi-norm*. From the above discussion, a semi-norm also has the properties

$\|x\| \geq 0$ for all $x \in E$, and $\|0\| = 0$.

However, there may be nonzero vectors $x \in E$ such that $\|x\| = 0$. Let us give some examples of normed vector spaces.

Example B.1.

1. Let $E = \mathbb{R}$, and $\|x\| = |x|$, the absolute value of x.

2. Let $E = \mathbb{C}$, and $\|z\| = |z|$, the modulus of z.

3. Let $E = \mathbb{R}^n$ (or $E = \mathbb{C}^n$). There are three standard norms. For every $(x_1, \ldots, x_n) \in E$, we have the norm $\|x\|_1$, defined such that,

$$\|x\|_1 = |x_1| + \cdots + |x_n|,$$

we have the *Euclidean norm* $\|x\|_2$, defined such that,

$$\|x\|_2 = \left(|x_1|^2 + \cdots + |x_n|^2\right)^{\frac{1}{2}} ,$$

and the *sup*-norm $\|x\|_\infty$, defined such that,

$$\|x\|_\infty = \max\{|x_i| \mid 1 \leq i \leq n\}.$$

More generally, we define the *ℓ^p-norm* (for $p \geq 1$) by

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

There are other norms besides the ℓ^p-norms. Here are some examples.

1. For $E = \mathbb{R}^2$,
$$\|(u_1, u_2)\| = |u_1| + 2|u_2|.$$

2. For $E = \mathbb{R}^2$,
$$\|(u_1, u_2)\| = \big((u_1 + u_2)^2 + u_1^2\big)^{1/2}.$$

3. For $E = \mathbb{C}^2$,
$$\|(u_1, u_2)\| = |u_1 + iu_2| + |u_1 - iu_2|.$$

The reader should check that they satisfy all the axioms of a norm.

Some work is required to show the triangle inequality for the ℓ^p-norm.

Proposition B.1. *If E is a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, for every real number $p \geq 1$, the ℓ^p-norm is indeed a norm.*

Proof. The cases $p = 1$ and $p = \infty$ are easy and left to the reader. If $p > 1$, then let $q > 1$ such that
$$\frac{1}{p} + \frac{1}{q} = 1.$$
We will make use of the following fact: for all $\alpha, \beta \in \mathbb{R}$, if $\alpha, \beta \geq 0$, then
$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}. \tag{$*$}$$
To prove the above inequality, we use the fact that the exponential function $t \mapsto e^t$ satisfies the following convexity inequality:
$$e^{\theta x + (1-\theta)y} \leq \theta e^x + (1 - \theta)e^y,$$
for all $x, y \in \mathbb{R}$ and all θ with $0 \leq \theta \leq 1$.

Since the case $\alpha\beta = 0$ is trivial, let us assume that $\alpha > 0$ and $\beta > 0$. If we replace θ by $1/p$, x by $p \log \alpha$ and y by $q \log \beta$, then we get
$$e^{\frac{1}{p} p \log \alpha + \frac{1}{q} q \log \beta} \leq \frac{1}{p} e^{p \log \alpha} + \frac{1}{q} e^{q \log \beta},$$
which simplifies to
$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q},$$
as claimed.

We will now prove that for any two vectors $u, v \in E$, we have
$$\sum_{i=1}^{n} |u_i v_i| \leq \|u\|_p \|v\|_q. \tag{$**$}$$

Since the above is trivial if $u = 0$ or $v = 0$, let us assume that $u \neq 0$ and $v \neq 0$. Then, the inequality $(*)$ with $\alpha = |u_i| / \|u\|_p$ and $\beta = |v_i| / \|v\|_q$ yields

$$\frac{|u_i v_i|}{\|u\|_p \|v\|_q} \leq \frac{|u_i|^p}{p \|u\|_p^p} + \frac{|v_i|^q}{q \|u\|_q^q},$$

for $i = 1, \ldots, n$, and by summing up these inequalities, we get

$$\sum_{i=1}^n |u_i v_i| \leq \|u\|_p \|v\|_q \,,$$

as claimed. To finish the proof, we simply have to prove that property (N3) holds, since (N1) and (N2) are clear. Now, for $i = 1, \ldots, n$, we can write

$$(|u_i| + |v_i|)^p = |u_i|(|u_i| + |v_i|)^{p-1} + |v_i|(|u_i| + |v_i|)^{p-1},$$

so that by summing up these equations we get

$$\sum_{i=1}^n (|u_i| + |v_i|)^p = \sum_{i=1}^n |u_i|(|u_i| + |v_i|)^{p-1} + \sum_{i=1}^n |v_i|(|u_i| + |v_i|)^{p-1},$$

and using the inequality $(**)$, we get

$$\sum_{i=1}^n (|u_i| + |v_i|)^p \leq (\|u\|_p + \|v\|_p) \left(\sum_{i=1}^n (|u_i| + |v_i|)^{(p-1)q} \right)^{1/q}.$$

However, $1/p + 1/q = 1$ implies $pq = p + q$, that is, $(p-1)q = p$, so we have

$$\sum_{i=1}^n (|u_i| + |v_i|)^p \leq (\|u\|_p + \|v\|_p) \left(\sum_{i=1}^n (|u_i| + |v_i|)^p \right)^{1/q},$$

which yields

$$\left(\sum_{i=1}^n (|u_i| + |v_i|)^p \right)^{1/p} \leq \|u\|_p + \|v\|_p \,.$$

Since $|u_i + v_i| \leq |u_i| + |v_i|$, the above implies the triangle inequality $\|u + v\|_p \leq \|u\|_p + \|v\|_p$, as claimed. $\square$

For $p > 1$ and $1/p + 1/q = 1$, the inequality

$$\sum_{i=1}^n |u_i v_i| \leq \left(\sum_{i=1}^n |u_i|^p \right)^{1/p} \left(\sum_{i=1}^n |v_i|^q \right)^{1/q}$$

is known as *Hölder's inequality*. For $p = 2$, it is the *Cauchy–Schwarz inequality*.

Actually, if we define the *Hermitian inner product* $\langle -, - \rangle$ on $\mathbb{C}^n$ by

$$\langle u, v \rangle = \sum_{i=1}^{n} u_i \overline{v}_i,$$

where $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$, then

$$|\langle u, v \rangle| \leq \sum_{i=1}^{n} |u_i \overline{v}_i| = \sum_{i=1}^{n} |u_i v_i|,$$

so Hölder's inequality implies the inequality

$$|\langle u, v \rangle| \leq \|u\|_p \, \|v\|_q$$

also called *Hölder's inequality*, which, for $p = 2$ is the standard Cauchy–Schwarz inequality. The triangle inequality for the ℓ^p-norm,

$$\left(\sum_{i=1}^{n} (|u_i + v_i|)^p \right)^{1/p} \leq \left(\sum_{i=1}^{n} |u_i|^p \right)^{1/p} + \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p},$$

is known as *Minkowski's inequality*.

When we restrict the Hermitian inner product to real vectors, $u, v \in \mathbb{R}^n$, we get the *Euclidean inner product*

$$\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i.$$

It is very useful to observe that if we represent (as usual) $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ (in $\mathbb{R}^n$) by column vectors, then their Euclidean inner product is given by

$$\langle u, v \rangle = u^\top v = v^\top u,$$

and when $u, v \in \mathbb{C}^n$, their Hermitian inner product is given by

$$\langle u, v \rangle = v^* u = \overline{u^* v}.$$

In particular, when $u = v$, in the complex case we get

$$\|u\|_2^2 = u^* u,$$

and in the real case, this becomes

$$\|u\|_2^2 = u^\top u.$$

As convenient as these notations are, we still recommend that you do not abuse them; the notation $\langle u, v \rangle$ is more intrinsic and still "works" when our vector space is infinite dimensional.

The following proposition is easy to show.

Proposition B.2. *The following inequalities hold for all $x \in \mathbb{R}^n$ (or $x \in \mathbb{C}^n$):*

$$\begin{aligned}\|x\|_\infty &\le \|x\|_1 \le n\|x\|_\infty,\\ \|x\|_\infty &\le \|x\|_2 \le \sqrt{n}\|x\|_\infty,\\ \|x\|_2 &\le \|x\|_1 \le \sqrt{n}\|x\|_2.\end{aligned}$$

Proposition B.2 is a special case of a very important result: in a finite-dimensional vector space, any two norms are equivalent.

Definition B.2. Given any (real or complex) vector space E, two norms $\| \|_a$ and $\| \|_b$ are *equivalent* iff there exists some positive reals $C_1, C_2 > 0$, such that

$$\|u\|_a \le C_1 \|u\|_b \quad \text{and} \quad \|u\|_b \le C_2 \|u\|_a, \text{ for all } u \in E.$$

Given any norm $\| \|$ on a vector space of dimension n, for any basis $(e_1, \ldots, e_n)$ of E, observe that for any vector $x = x_1e_1 + \cdots + x_ne_n$, we have

$$\|x\| = \|x_1e_1 + \cdots + x_ne_n\| \le |x_1| \|e_1\| + \cdots + |x_n| \|e_n\| \le C(|x_1| + \cdots + |x_n|) = C \|x\|_1,$$

with $C = \max_{1 \le i \le n} \|e_i\|$ and

$$\|x\|_1 = \|x_1e_1 + \cdots + x_ne_n\| = |x_1| + \cdots + |x_n|.$$

The above implies that

$$| \|u\| - \|v\| | \le \|u - v\| \le C \|u - v\|_1,$$

which means that the map $u \mapsto \|u\|$ is *continuous* with respect to the norm $\| \|_1$.

Let S_1^{n-1} be the unit sphere with respect to the norm $\| \|_1$, namely

$$S_1^{n-1} = \{x \in E \mid \|x\|_1 = 1\}.$$

Now, S_1^{n-1} is a closed and bounded subset of a finite-dimensional vector space, so by Heine–Borel (or equivalently, by Bolzano–Weiertrass), S_1^{n-1} is compact. On the other hand, it is a well known result of analysis that any continuous real-valued function on a nonempty compact set has a minimum and a maximum, and that they are achieved. Using these facts, we can prove the following important theorem:

Theorem B.3. *If E is any real or complex vector space of finite dimension, then any two norms on E are equivalent.*

Proof. It is enough to prove that any norm $\| \|$ is equivalent to the 1-norm. We already proved that the function $x \mapsto \|x\|$ is continuous with respect to the norm $\| \|_1$ and we observed that the unit sphere S_1^{n-1} is compact. Now, we just recalled that because the function $f\colon x \mapsto \|x\|$ is continuous and because S_1^{n-1} is compact, the function f has a minimum m and a maximum

M, and because $\|x\|$ is never zero on S_1^{n-1}, we must have $m > 0$. Consequently, we just proved that if $\|x\|_1 = 1$, then

$$0 < m \leq \|x\| \leq M,$$

so for any $x \in E$ with $x \neq 0$, we get

$$m \leq \|x/\, \|x\|_1\| \leq M,$$

which implies

$$m\, \|x\|_1 \leq \|x\| \leq M\, \|x\|_1 .$$

Since the above inequality holds trivially if $x = 0$, we just proved that $\|\ \|$ and $\|\ \|_1$ are equivalent, as claimed. □

Next, we will consider norms on matrices.

B.2 Matrix Norms

For simplicity of exposition, we will consider the vector spaces $\mathrm{M}_n(\mathbb{R})$ and $\mathrm{M}_n(\mathbb{C})$ of square $n \times n$ matrices. Most results also hold for the spaces $\mathrm{M}_{m,n}(\mathbb{R})$ and $\mathrm{M}_{m,n}(\mathbb{C})$ of rectangular $m \times n$ matrices. Since $n \times n$ matrices can be multiplied, the idea behind matrix norms is that they should behave "well" with respect to matrix multiplication.

Definition B.3. A *matrix norm* $\|\ \|$ on the space of square $n \times n$ matrices in $\mathrm{M}_n(K)$, with $K = \mathbb{R}$ or $K = \mathbb{C}$, is a norm on the vector space $\mathrm{M}_n(K)$, with the additional property called *submultiplicativity* that

$$\|AB\| \leq \|A\|\, \|B\| ,$$

for all $A, B \in \mathrm{M}_n(K)$. A norm on matrices satisfying the above property is often called a *submultiplicative* matrix norm.

Since $I^2 = I$, from $\|I\| = \|I^2\| \leq \|I\|^2$, we get $\|I\| \geq 1$, for every matrix norm.

Before giving examples of matrix norms, we need to review some basic definitions about matrices. Given any matrix $A = (a_{ij}) \in \mathrm{M}_{m,n}(\mathbb{C})$, the *conjugate* $\overline{A}$ of A is the matrix such that

$$\overline{A}_{ij} = \overline{a_{ij}}, \quad 1 \leq i \leq m,\ 1 \leq j \leq n.$$

The *transpose* of A is the $n \times m$ matrix $A^\top$ such that

$$A^\top_{ij} = a_{ji}, \quad 1 \leq i \leq m,\ 1 \leq j \leq n.$$

The *adjoint* of A is the $n \times m$ matrix A^* such that

$$A^* = \overline{(A^\top)} = (\overline{A})^\top.$$

When A is a real matrix, $A^* = A^\top$. A matrix $A \in \mathrm{M}_n(\mathbb{C})$ is *Hermitian* if

$$A^* = A.$$

If A is a real matrix ($A \in \mathrm{M}_n(\mathbb{R})$), we say that A is *symmetric* if

$$A^\top = A.$$

A matrix $A \in \mathrm{M}_n(\mathbb{C})$ is *normal* if

$$AA^* = A^*A,$$

and if A is a real matrix, it is *normal* if

$$AA^\top = A^\top A.$$

A matrix $U \in \mathrm{M}_n(\mathbb{C})$ is *unitary* if

$$UU^* = U^*U = I.$$

A real matrix $Q \in \mathrm{M}_n(\mathbb{R})$ is *orthogonal* if

$$QQ^\top = Q^\top Q = I.$$

Given any matrix $A = (a_{ij}) \in \mathrm{M}_n(\mathbb{C})$, the *trace* $\mathrm{tr}(A)$ of A is the sum of its diagonal elements

$$\mathrm{tr}(A) = a_{11} + \cdots + a_{nn}.$$

It is easy to show that the trace is a linear map, so that

$$\mathrm{tr}(\lambda A) = \lambda \mathrm{tr}(A)$$

and

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B).$$

Moreover, if A is an $m \times n$ matrix and B is an $n \times m$ matrix, it is not hard to show that

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

We also review eigenvalues and eigenvectors. We content ourselves with definition involving matrices.

Definition B.4. Given any square matrix $A \in \mathrm{M}_n(\mathbb{C})$, a complex number $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if there is some *nonzero* vector $u \in \mathbb{C}^n$, such that

$$Au = \lambda u.$$

If λ is an eigenvalue of A, then the *nonzero* vectors $u \in \mathbb{C}^n$ such that $Au = \lambda u$ are called *eigenvectors of A associated with* λ; together with the zero vector, these eigenvectors form a subspace of $\mathbb{C}^n$ denoted by $E_\lambda(A)$, and called the *eigenspace associated with* λ.

Remark: Note that Definition B.4 *requires an eigenvector to be nonzero*. A somewhat unfortunate consequence of this requirement is that the set of eigenvectors is *not* a subspace, since the zero vector is missing! On the positive side, whenever eigenvectors are involved, there is no need to say that they are nonzero. In contrast, even if we allow 0 to be an eigenvector, in order for a scalar λ to be an eigenvalue, there must be a *nonzero vector* u such that $Au = \lambda u$. Without this restriction, since $A0 = \lambda 0 = 0$ for all λ, every scalar would be an eigenvector, which would make the definition of an eigenvalue trivial and useless. The fact that eigenvectors are nonzero is implicitly used in all the arguments involving them, so it seems preferable (but perhaps not as elegant) to stipulate that eigenvectors should be nonzero.

If A is a square real matrix $A \in \mathrm{M}_n(\mathbb{R})$, then we restrict Definition B.4 to real eigenvalues $\lambda \in \mathbb{R}$ and real eigenvectors. However, it should be noted that although every complex matrix always has at least some complex eigenvalue, a real matrix may not have any real eigenvalues. For example, the matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

has the complex eigenvalues i and $-i$, but no real eigenvalues. Thus, typically even for real matrices, we consider complex eigenvalues.

Observe that $\lambda \in \mathbb{C}$ is an eigenvalue of A

- iff $Au = \lambda u$ for some nonzero vector $u \in \mathbb{C}^n$
- iff $(\lambda I - A)u = 0$
- iff the matrix $\lambda I - A$ defines a linear map which has a nonzero kernel, that is,
- iff $\lambda I - A$ not invertible.

However, it is a standard fact of linear algebra that $\lambda I - A$ is not invertible iff

$$\det(\lambda I - A) = 0.$$

Now $\det(\lambda I - A)$ is a polynomial of degree n in the indeterminate λ, in fact, of the form

$$\lambda^n - \mathrm{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A).$$

Thus we see that the eigenvalues of A are the zeros (also called *roots*) of the above polynomial. Since every complex polynomial of degree n has exactly n roots, counted with their multiplicity, we have the following definition:

Definition B.5. Given any square $n \times n$ matrix $A \in \mathrm{M}_n(\mathbb{C})$, the polynomial

$$\det(\lambda I - A) = \lambda^n - \mathrm{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A)$$

is called the *characteristic polynomial* of A. The n (not necessarily distinct) roots $\lambda_1, \ldots, \lambda_n$ of the characteristic polynomial are all the *eigenvalues* of A and constitute the *spectrum* of A. We let

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|$$

be the largest modulus of the eigenvalues of A, called the *spectral radius* of A.

Since the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are the zeros of the polynomial

$$\det(\lambda I - A) = \lambda^n - \mathrm{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A),$$

we deduce that

$$\begin{aligned} \mathrm{tr}(A) &= \lambda_1 + \cdots + \lambda_n \\ \det(A) &= \lambda_1 \cdots \lambda_n. \end{aligned}$$

Proposition B.4. *For any matrix norm $\| \ \|$ on $\mathrm{M}_n(\mathbb{C})$ and for any square $n \times n$ matrix $A \in \mathrm{M}_n(\mathbb{C})$, we have*

$$\rho(A) \leq \|A\|.$$

Proof. Let λ be some eigenvalue of A for which $|\lambda|$ is maximum, that is, such that $|\lambda| = \rho(A)$. If u $(\neq 0)$ is any eigenvector associated with λ and if U is the $n \times n$ matrix whose columns are all u, then $Au = \lambda u$ implies

$$AU = \lambda U,$$

and since

$$|\lambda| \, \|U\| = \|\lambda U\| = \|AU\| \leq \|A\| \, \|U\|$$

and $U \neq 0$, we have $\|U\| \neq 0$, and get

$$\rho(A) = |\lambda| \leq \|A\|,$$

as claimed. ∎

Proposition B.4 also holds for any real matrix norm $\| \ \|$ on $\mathrm{M}_n(\mathbb{R})$ but the proof is more subtle and requires the notion of induced norm. We prove it after giving Definition B.7.

It turns out that if A is a real $n \times n$ symmetric matrix, then the eigenvalues of A are all real and there is some orthogonal matrix Q such that

$$A = Q\mathrm{diag}(\lambda_1, \ldots, \lambda_n)Q^\top,$$

where $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ denotes the matrix whose only nonzero entries (if any) are its diagonal entries, which are the (real) eigenvalues of A. Similarly, if A is a complex $n \times n$ Hermitian matrix, then the eigenvalues of A are all real and there is some unitary matrix U such that

$$A = U\mathrm{diag}(\lambda_1, \ldots, \lambda_n)U^*,$$

where $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ denotes the matrix whose only nonzero entries (if any) are its diagonal entries, which are the (real) eigenvalues of A.

We now return to matrix norms. We begin with the so-called *Frobenius norm*, which is just the norm $\| \, \|_2$ on $\mathbb{C}^{n^2}$, where the $n \times n$ matrix A is viewed as the vector obtained by concatenating together the rows (or the columns) of A. The reader should check that for any $n \times n$ complex matrix $A = (a_{ij})$,

$$\left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2} = \sqrt{\mathrm{tr}(A^*A)} = \sqrt{\mathrm{tr}(AA^*)}.$$

Definition B.6. The *Frobenius norm* $\| \, \|_F$ is defined so that for every square $n \times n$ matrix $A \in \mathrm{M}_n(\mathbb{C})$,

$$\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2} = \sqrt{\mathrm{tr}(AA^*)} = \sqrt{\mathrm{tr}(A^*A)}.$$

The following proposition show that the Frobenius norm is a matrix norm satisfying other nice properties.

Proposition B.5. *The Frobenius norm* $\| \, \|_F$ *on* $\mathrm{M}_n(\mathbb{C})$ *satisfies the following properties:*

(1) It is a matrix norm; that is, $\|AB\|_F \leq \|A\|_F \|B\|_F$*, for all* $A, B \in \mathrm{M}_n(\mathbb{C})$.

(2) It is unitarily invariant, which means that for all unitary matrices U, V*, we have*

$$\|A\|_F = \|UA\|_F = \|AV\|_F = \|UAV\|_F .$$

(3) $\sqrt{\rho(A^*A)} \leq \|A\|_F \leq \sqrt{n}\sqrt{\rho(A^*A)}$*, for all* $A \in \mathrm{M}_n(\mathbb{C})$.

Proof. (1) The only property that requires a proof is the fact $\|AB\|_F \leq \|A\|_F \|B\|_F$. This follows from the Cauchy–Schwarz inequality:

$$\begin{aligned} \|AB\|_F^2 &= \sum_{i,j=1}^{n} \left| \sum_{k=1}^{n} a_{ik} b_{kj} \right|^2 \\ &\leq \sum_{i,j=1}^{n} \left(\sum_{h=1}^{n} |a_{ih}|^2 \right) \left(\sum_{k=1}^{n} |b_{kj}|^2 \right) \\ &= \left(\sum_{i,h=1}^{n} |a_{ih}|^2 \right) \left(\sum_{k,j=1}^{n} |b_{kj}|^2 \right) = \|A\|_F^2 \, \|B\|_F^2 . \end{aligned}$$

(2) We have

$$\|A\|_F^2 = \mathrm{tr}(AA^*) = \mathrm{tr}(AVV^*A^*) = \mathrm{tr}(AV(AV)^*) = \|AV\|_F^2 ,$$

and
$$\|A\|_F^2 = \operatorname{tr}(A^*A) = \operatorname{tr}(A^*U^*UA) = \|UA\|_F^2 .$$
The identity
$$\|A\|_F = \|UAV\|_F$$
follows from the previous two.

(3) It is known by linear algebra that the trace of a matrix is equal to the sum of its eigenvalues. Furthermore, A^*A is symmetric positive semidefinite (which means that its eigenvalues are nonnegative), so $\rho(A^*A)$ is the largest eigenvalue of A^*A and
$$\rho(A^*A) \leq \operatorname{tr}(A^*A) \leq n\rho(A^*A),$$
which yields (3) by taking square roots. $\square$

Remark: The Frobenius norm is also known as the *Hilbert-Schmidt norm* or the *Schur norm*. So many famous names associated with such a simple thing!

B.3 Subordinate Norms

We now give another method for obtaining matrix norms using subordinate norms. First we need a proposition that shows that in a finite-dimensional space, the linear map induced by a matrix is bounded, and thus continuous.

Proposition B.6. *For every norm $\| \;\|$ on $\mathbb{C}^n$ (or $\mathbb{R}^n$), for every matrix $A \in \mathrm{M}_n(\mathbb{C})$ (or $A \in \mathrm{M}_n(\mathbb{R})$), there is a real constant $C_A \geq 0$, such that*
$$\|Au\| \leq C_A \|u\| ,$$
for every vector $u \in \mathbb{C}^n$ (or $u \in \mathbb{R}^n$ if A is real).

Proof. For every basis $(e_1, \ldots, e_n)$ of $\mathbb{C}^n$ (or $\mathbb{R}^n$), for every vector $u = u_1e_1 + \cdots + u_ne_n$, we have
$$\begin{aligned}\|Au\| &= \|u_1A(e_1) + \cdots + u_nA(e_n)\| \\ &\leq |u_1| \|A(e_1)\| + \cdots + |u_n| \|A(e_n)\| \\ &\leq C_1(|u_1| + \cdots + |u_n|) = C_1 \|u\|_1 ,\end{aligned}$$
where $C_1 = \max_{1\leq i\leq n} \|A(e_i)\|$. By Theorem B.3, the norms $\| \;\|$ and $\| \;\|_1$ are equivalent, so there is some constant $C_2 > 0$ so that $\|u\|_1 \leq C_2 \|u\|$ for all u, which implies that
$$\|Au\| \leq C_A \|u\| ,$$
where $C_A = C_1C_2$. $\square$

Proposition B.6 says that every linear map on a finite-dimensional space is *bounded*. This implies that every linear map on a finite-dimensional space is continuous. Actually, it is not hard to show that a linear map on a normed vector space E is bounded iff it is continuous, regardless of the dimension of E.

Proposition B.6 implies that for every matrix $A \in \mathrm{M}_n(\mathbb{C})$ (or $A \in \mathrm{M}_n(\mathbb{R})$),

$$\sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|} \leq C_A.$$

Since $\|\lambda u\| = |\lambda| \, \|u\|$, for every nonzero vector x, we have

$$\frac{\|Ax\|}{\|x\|} = \frac{\|x\| \, \|A(x/\|x\|)\|}{\|x\|} = \|A(x/\|x\|)\|,$$

which implies that

$$\sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|} = \sup_{\substack{x \in \mathbb{C}^n \\ \|x\| = 1}} \|Ax\|.$$

Similarly

$$\sup_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|} = \sup_{\substack{x \in \mathbb{R}^n \\ \|x\| = 1}} \|Ax\|.$$

The above considerations justify the following definition.

Definition B.7. If $\| \ \|$ is any norm on $\mathbb{C}^n$, we define the function $\| \ \|_{\mathrm{op}}$ on $\mathrm{M}_n(\mathbb{C})$ by

$$\|A\|_{\mathrm{op}} = \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|} = \sup_{\substack{x \in \mathbb{C}^n \\ \|x\| = 1}} \|Ax\|.$$

The function $A \mapsto \|A\|_{\mathrm{op}}$ is called the *subordinate matrix norm* or *operator norm* induced by the norm $\| \ \|$.

Another notation for the operator norm of a matrix A (in particular, used by Horn and Johnson [50]), is $|||A|||$.

It is easy to check that the function $A \mapsto \|A\|_{\mathrm{op}}$ is indeed a norm, and by definition, it satisfies the property

$$\|Ax\| \leq \|A\|_{\mathrm{op}} \|x\|, \quad \text{for all } x \in \mathbb{C}^n.$$

A norm $\| \ \|_{\mathrm{op}}$ on $\mathrm{M}_n(\mathbb{C})$ satisfying the above property is said to be *subordinate* to the vector norm $\| \ \|$ on $\mathbb{C}^n$. As a consequence of the above inequality, we have

$$\|ABx\| \leq \|A\|_{\mathrm{op}} \|Bx\| \leq \|A\|_{\mathrm{op}} \|B\|_{\mathrm{op}} \|x\|,$$

for all $x \in \mathbb{C}^n$, which implies that

$$\|AB\|_{\text{op}} \leq \|A\|_{\text{op}} \|B\|_{\text{op}} \quad \text{for all } A, B \in \mathrm{M}_n(\mathbb{C}),$$

showing that $A \mapsto \|A\|_{\text{op}}$ is a matrix norm (it is submultiplicative).

Observe that the operator norm is also defined by

$$\|A\|_{\text{op}} = \inf\{\lambda \in \mathbb{R} \mid \|Ax\| \leq \lambda \|x\|, \text{ for all } x \in \mathbb{C}^n\}.$$

Since the function $x \mapsto \|Ax\|$ is continuous (because $|\, \|Ay\| - \|Ax\|\, | \leq \|Ay - Ax\| \leq C_A \|x - y\|$) and the unit sphere $S^{n-1} = \{x \in \mathbb{C}^n \mid \|x\| = 1\}$ is compact, there is some $x \in \mathbb{C}^n$ such that $\|x\| = 1$ and

$$\|Ax\| = \|A\|_{\text{op}}.$$

Equivalently, there is some $x \in \mathbb{C}^n$ such that $x \neq 0$ and

$$\|Ax\| = \|A\|_{\text{op}} \|x\|.$$

Consequently we can replace sup by max in the definition of $\|A\|_{\text{op}}$ (and inf by min), namely

$$\|A\|_{\text{op}} = \max_{\substack{x \in \mathbb{C}^n \\ \|x\|=1}} \|Ax\|.$$

The definition of an operator norm also implies that

$$\|I\|_{\text{op}} = 1.$$

The above shows that the Frobenius norm is not a subordinate matrix norm for $n \geq 2$ (why?).

If $\|\ \|$ is a vector norm on $\mathbb{C}^n$, the operator norm $\|\ \|_{\text{op}}$ that it induces applies to matrices in $\mathrm{M}_n(\mathbb{C})$. If we are careful to denote vectors and matrices so that no confusion arises, for example, by using lower case letters for vectors and upper case letters for matrices, it should be clear that $\|A\|_{\text{op}}$ is the operator norm of the matrix A and that $\|x\|$ is the vector norm of x. Consequently, following common practice to alleviate notation, we will drop the subscript "op" and simply write $\|A\|$ instead of $\|A\|_{\text{op}}$.

The notion of subordinate norm can be slightly generalized.

Definition B.8. If $K = \mathbb{R}$ or $K = \mathbb{C}$, for any norm $\|\ \|$ on $\mathrm{M}_{m,n}(K)$, and for any two norms $\|\ \|_a$ on K^n and $\|\ \|_b$ on K^m, we say that the norm $\|\ \|$ is *subordinate* to the norms $\|\ \|_a$ and $\|\ \|_b$ if

$$\|Ax\|_b \leq \|A\| \|x\|_a \quad \text{for all } A \in \mathrm{M}_{m,n}(K) \text{ and all } x \in K^n.$$

Remark: For any norm $\| \; \|$ on $\mathbb{C}^n$, we can define the function $\| \; \|_{\mathbb{R}}$ on $\mathrm{M}_n(\mathbb{R})$ by

$$\|A\|_{\mathbb{R}} = \sup_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|} = \sup_{\substack{x \in \mathbb{R}^n \\ \|x\| = 1}} \|Ax\| .$$

The function $A \mapsto \|A\|_{\mathbb{R}}$ is a matrix norm on $\mathrm{M}_n(\mathbb{R})$, and

$$\|A\|_{\mathbb{R}} \leq \|A\| ,$$

for all real matrices $A \in \mathrm{M}_n(\mathbb{R})$. However, it is possible to construct vector norms $\| \; \|$ on $\mathbb{C}^n$ and *real* matrices A such that

$$\|A\|_{\mathbb{R}} < \|A\| .$$

In order to avoid this kind of difficulties, we define subordinate matrix norms over $\mathrm{M}_n(\mathbb{C})$. Luckily, it turns out that $\|A\|_{\mathbb{R}} = \|A\|$ for the vector norms, $\| \; \|_1 , \| \; \|_2$, and $\| \; \|_\infty$.

We now prove Proposition B.4 for real matrix norms.

Proposition B.7. *For any matrix norm $\| \; \|$ on $\mathrm{M}_n(\mathbb{R})$ and for any square $n \times n$ matrix $A \in \mathrm{M}_n(\mathbb{R})$, we have*

$$\rho(A) \leq \|A\| .$$

Proof. We follow the proof in Denis Serre's book [88]. If A is a real matrix, the problem is that the eigenvectors associated with the eigenvalue of maximum modulus may be complex. We use a trick based on the fact that for every matrix A (real or complex),

$$\rho(A^k) = (\rho(A))^k,$$

which is left as an exercise

Pick any complex matrix norm $\| \; \|_c$ on $\mathbb{C}^n$ (for example, the Frobenius norm, or any subordinate matrix norm induced by a norm on $\mathbb{C}^n$). The restriction of $\| \; \|_c$ to real matrices is a real norm that we also denote by $\| \; \|_c$. Now by Theorem B.3, since $\mathrm{M}_n(\mathbb{R})$ has finite dimension n^2, there is some constant $C > 0$ so that

$$\|B\|_c \leq C \|B\| , \quad \text{for all} \quad B \in \mathrm{M}_n(\mathbb{R}).$$

Furthermore, for every $k \geq 1$ and for every real $n \times n$ matrix A, by Proposition B.4, $\rho(A^k) \leq \left\|A^k\right\|_c$, and because $\| \; \|$ is a matrix norm, $\left\|A^k\right\| \leq \|A\|^k$, so we have

$$(\rho(A))^k = \rho(A^k) \leq \left\|A^k\right\|_c \leq C \left\|A^k\right\| \leq C \|A\|^k ,$$

for all $k \geq 1$. It follows that

$$\rho(A) \leq C^{1/k} \|A\| , \quad \text{for all} \quad k \geq 1.$$

However because $C > 0$, we have $\lim_{k \mapsto \infty} C^{1/k} = 1$ (we have $\lim_{k \mapsto \infty} \frac{1}{k} \log(C) = 0$). Therefore, we conclude that

$$\rho(A) \leq \|A\| ,$$

as desired. $\square$

We now determine explicitly what are the subordinate matrix norms associated with the vector norms $\| \|_1$, $\| \|_2$, and $\| \|_\infty$.

Proposition B.8. *For every square matrix $A = (a_{ij}) \in \mathrm{M}_n(\mathbb{C})$, we have*

$$\begin{aligned} \|A\|_1 &= \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_1 = 1}} \|Ax\|_1 = \max_j \sum_{i=1}^n |a_{ij}| \\ \|A\|_\infty &= \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_\infty = 1}} \|Ax\|_\infty = \max_i \sum_{j=1}^n |a_{ij}| \\ \|A\|_2 &= \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_2 = 1}} \|Ax\|_2 = \sqrt{\rho(A^*A)} = \sqrt{\rho(AA^*)}. \end{aligned}$$

Note that $\|A\|_1$ is the maximum of the ℓ^1-norms of the columns of A and $\|A\|_\infty$ is the maximum of the ℓ^1-norms of the rows of A. Furthermore, $\|A^*\|_2 = \|A\|_2$, the norm $\| \|_2$ is unitarily invariant, which means that

$$\|A\|_2 = \|UAV\|_2$$

for all unitary matrices U, V, and if A is a normal matrix, then $\|A\|_2 = \rho(A)$.

Proof. For every vector u, we have

$$\|Au\|_1 = \sum_i \left| \sum_j a_{ij} u_j \right| \leq \sum_j |u_j| \sum_i |a_{ij}| \leq \left(\max_j \sum_i |a_{ij}| \right) \|u\|_1,$$

which implies that

$$\|A\|_1 \leq \max_j \sum_{i=1}^n |a_{ij}|.$$

It remains to show that equality can be achieved. For this let j_0 be some index such that

$$\max_j \sum_i |a_{ij}| = \sum_i |a_{ij_0}|,$$

and let $u_i = 0$ for all $i \neq j_0$ and $u_{j_0} = 1$.

In a similar way, we have

$$\|Au\|_\infty = \max_i \left| \sum_j a_{ij} u_j \right| \leq \left(\max_i \sum_j |a_{ij}| \right) \|u\|_\infty,$$

which implies that

$$\|A\|_\infty \leq \max_i \sum_{j=1}^n |a_{ij}|.$$

To achieve equality, let i_0 be some index such that

$$\max_i \sum_j |a_{ij}| = \sum_j |a_{i_0 j}|.$$

The reader should check that the vector given by

$$u_j = \begin{cases} \frac{\overline{a}_{i_0 j}}{|a_{i_0 j}|} & \text{if } a_{i_0 j} \neq 0 \\ 1 & \text{if } a_{i_0 j} = 0 \end{cases}$$

works.

We have

$$\|A\|_2^2 = \sup_{\substack{x \in \mathbb{C}^n \\ x^* x = 1}} \|Ax\|_2^2 = \sup_{\substack{x \in \mathbb{C}^n \\ x^* x = 1}} x^* A^* A x.$$

Since the matrix A^*A is symmetric, it has real eigenvalues and it can be diagonalized with respect to a unitary matrix. These facts can be used to prove that the function $x \mapsto x^* A^* A x$ has a maximum on the sphere $x^* x = 1$ equal to the largest eigenvalue of A^*A, namely, $\rho(A^*A)$. We postpone the proof until we discuss optimizing quadratic functions. Therefore,

$$\|A\|_2 = \sqrt{\rho(A^*A)}.$$

Let use now prove that $\rho(A^*A) = \rho(AA^*)$. First assume that $\rho(A^*A) > 0$. In this case, there is some eigenvector u ($\neq 0$) such that

$$A^*Au = \rho(A^*A)u,$$

and since $\rho(A^*A) > 0$, we must have $Au \neq 0$. Since $Au \neq 0$,

$$AA^*(Au) = A(A^*Au) = \rho(A^*A)Au$$

which means that $\rho(A^*A)$ is an eigenvalue of AA^*, and thus

$$\rho(A^*A) \leq \rho(AA^*).$$

Because $(A^*)^* = A$, by replacing A by A^*, we get

$$\rho(AA^*) \leq \rho(A^*A),$$

and so $\rho(A^*A) = \rho(AA^*)$.

If $\rho(A^*A) = 0$, then we must have $\rho(AA^*) = 0$, since otherwise by the previous reasoning we would have $\rho(A^*A) = \rho(AA^*) > 0$. Hence, in all case

$$\|A\|_2^2 = \rho(A^*A) = \rho(AA^*) = \|A^*\|_2^2.$$

For any unitary matrices U and V, it is an easy exercise to prove that V^*A^*AV and A^*A have the same eigenvalues, so

$$\|A\|_2^2 = \rho(A^*A) = \rho(V^*A^*AV) = \|AV\|_2^2,$$

and also

$$\|A\|_2^2 = \rho(A^*A) = \rho(A^*U^*UA) = \|UA\|_2^2.$$

Finally, if A is a normal matrix ($AA^* = A^*A$), it can be shown that there is some unitary matrix U so that

$$A = UDU^*,$$

where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix consisting of the eigenvalues of A, and thus

$$A^*A = (UDU^*)^*UDU^* = UD^*U^*UDU^* = UD^*DU^*.$$

However, $D^*D = \mathrm{diag}(|\lambda_1|^2, \ldots, |\lambda_n|^2)$, which proves that

$$\rho(A^*A) = \rho(D^*D) = \max_i |\lambda_i|^2 = (\rho(A))^2,$$

so that $\|A\|_2 = \rho(A)$. $\square$

Definition B.9. For $A = (a_{ij}) \in \mathrm{M}_n(\mathbb{C})$, the norm $\|A\|_2$ is often called the *spectral norm*.

Observe that Property (3) of Proposition B.5 says that

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\, \|A\|_2,$$

which shows that the Frobenius norm is an upper bound on the spectral norm. The Frobenius norm is much easier to compute than the spectral norm.

The reader will check that the above proof still holds if the matrix A is real (change unitary to orthogonal), confirming the fact that $\|A\|_{\mathbb{R}} = \|A\|$ for the vector norms $\| \|_1, \| \|_2$, and $\| \|_\infty$. It is also easy to verify that the proof goes through for *rectangular* $m \times n$ matrices, with the same formulae. Similarly, the Frobenius norm given by

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} = \sqrt{\mathrm{tr}(A^*A)} = \sqrt{\mathrm{tr}(AA^*)}$$

is also a norm on rectangular matrices. For these norms, whenever AB makes sense, we have

$$\|AB\| \leq \|A\|\, \|B\|.$$

Remark: It can be shown that for any two real numbers $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\|A^*\|_q = \|A\|_p = \sup\{\Re(y^*Ax) \mid \|x\|_p = 1, \|y\|_q = 1\} = \sup\{|\langle Ax, y\rangle| \mid \|x\|_p = 1, \|y\|_q = 1\},$$

where $\|A^*\|_q$ and $\|A\|_p$ are the operator norms.

Remark: Let $(E, \| \|)$ and $(F, \| \|)$ be two normed vector spaces (for simplicity of notation, we use the same symbol $\| \|$ for the norms on E and F; this should not cause any confusion). Recall that a function $f\colon E \to F$ is *continuous* if for every $a \in E$, for every $\epsilon > 0$, there is some $\eta > 0$ such that for all $x \in E$,

$$\text{if} \quad \|x - a\| \leq \eta \quad \text{then} \quad \|f(x) - f(a)\| \leq \epsilon.$$

It is not hard to show that a *linear map* $f\colon E \to F$ is continuous iff there is some constant $C \geq 0$ such that

$$\|f(x)\| \leq C \|x\| \text{ for all } x \in E.$$

If so, we say that f is *bounded* (or a *linear bounded operator*). We let $\mathcal{L}(E; F)$ denote the set of all continuous (equivalently, bounded) linear maps from E to F. Then we can define the *operator norm* (or *subordinate norm*) $\| \|$ on $\mathcal{L}(E; F)$ as follows: for every $f \in \mathcal{L}(E; F)$,

$$\|f\| = \sup_{\substack{x \in E \\ x \neq 0}} \frac{\|f(x)\|}{\|x\|} = \sup_{\substack{x \in E \\ \|x\| = 1}} \|f(x)\|,$$

or equivalently by

$$\|f\| = \inf\{\lambda \in \mathbb{R} \mid \|f(x)\| \leq \lambda \|x\|, \text{ for all } x \in E\}.$$

Here because E may be infinite-dimensional, sup can't be replaced by max and inf can't be replaced by min. It is not hard to show that the map $f \mapsto \|f\|$ is a norm on $\mathcal{L}(E; F)$ satisfying the property

$$\|f(x)\| \leq \|f\| \|x\|$$

for all $x \in E$, and that if $f \in \mathcal{L}(E; F)$ and $g \in \mathcal{L}(F; G)$, then

$$\|g \circ f\| \leq \|g\| \|f\|.$$

Operator norms play an important role in functional analysis, especially when the spaces E and F are *complete*.

Appendix C

Basics of Groups and Group Actions

This chapter gathers basics of the theory of groups and group actions.

C.1 Groups, Subgroups, Cosets

Definition C.1. A *group* is a set G equipped with a binary operation $\cdot\colon G \times G \to G$ that associates an element $a \cdot b \in G$ to every pair of elements $a, b \in G$, and having the following properties: $\cdot$ is associative, has an identity element $e \in G$, and every element in G is invertible (w.r.t. $\cdot$). More explicitly, this means that the following equations hold for all $a, b, c \in G$:

(G1) $a \cdot (b \cdot c) = (a \cdot b) \cdot c.$ (associativity);

(G2) $a \cdot e = e \cdot a = a.$ (identity);

(G3) For every $a \in G$, there is some $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e.$ (inverse).

A group G is *abelian* (or *commutative*) if

$$a \cdot b = b \cdot a \quad \text{for all } a, b \in G.$$

A set M together with an operation $\cdot\colon M \times M \to M$ and an element e satisfying only Conditions (G1) and (G2) is called a *monoid*. For example, the set $\mathbb{N} = \{0, 1, \ldots, n, \ldots\}$ of natural numbers is a (commutative) monoid under addition. However, it is not a group.

Some examples of groups are given below.

Example C.1.

1. The set $\mathbb{Z} = \{\ldots, -n, \ldots, -1, 0, 1, \ldots, n, \ldots\}$ of integers is an abelian group under addition, with identity element 0. However, $\mathbb{Z}^* = \mathbb{Z} - \{0\}$ is not a group under multiplication.

2. The set $\mathbb{Q}$ of rational numbers (fractions p/q with $p, q \in \mathbb{Z}$ and $q \neq 0$) is an abelian group under addition, with identity element 0. The set $\mathbb{Q}^* = \mathbb{Q} - \{0\}$ is also an abelian group under multiplication, with identity element 1.

3. Given any nonempty set S, the set of bijections $f\colon S \to S$, also called *permutations of* S, is a group under function composition (i.e., the multiplication of f and g is the composition $g \circ f$), with identity element the identity function id_S. This group is not abelian as soon as S has more than two elements. The permutation group of the set $S = \{1, \ldots, n\}$ is often denoted $\mathfrak{S}_n$ and called the *symmetric group* on n elements.

4. For any positive integer $p \in \mathbb{N}$, define a relation on $\mathbb{Z}$, denoted $m \equiv n \pmod{p}$, as follows:
$$m \equiv n \pmod{p} \quad \text{iff} \quad m - n = kp \quad \text{for some } k \in \mathbb{Z}.$$
The reader will easily check that this is an equivalence relation, and, moreover, it is compatible with respect to addition and multiplication, which means that if $m_1 \equiv n_1 \pmod{p}$ and $m_2 \equiv n_2 \pmod{p}$, then $m_1 + m_2 \equiv n_1 + n_2 \pmod{p}$ and $m_1 m_2 \equiv n_1 n_2 \pmod{p}$. Consequently, we can define an addition operation and a multiplication operation of the set of equivalence classes $\pmod{p}$:
$$[m] + [n] = [m + n]$$
and
$$[m] \cdot [n] = [mn].$$
The reader will easily check that addition of residue classes $\pmod{p}$ induces an abelian group structure with $[0]$ as zero. This group is denoted $\mathbb{Z}/p\mathbb{Z}$.

5. The set of $n \times n$ invertible matrices with real (or complex) coefficients is a group under matrix multiplication, with identity element the identity matrix I_n. This group is called the *general linear group* and is usually denoted by $\mathbf{GL}(n, \mathbb{R})$ (or $\mathbf{GL}(n, \mathbb{C})$).

6. The set of $n \times n$ invertible matrices A with real (or complex) coefficients such that $\det(A) = 1$ is a group under matrix multiplication, with identity element the identity matrix I_n. This group is called the *special linear group* and is usually denoted by $\mathbf{SL}(n, \mathbb{R})$ (or $\mathbf{SL}(n, \mathbb{C})$).

7. The set of $n \times n$ matrices Q with real coefficients such that
$$QQ^\top = Q^\top Q = I_n$$
is a group under matrix multiplication, with identity element the identity matrix I_n; we have $Q^{-1} = Q^\top$. This group is called the *orthogonal group* and is usually denoted by $\mathbf{O}(n)$.

8. The set of $n \times n$ invertible matrices Q with real coefficients such that
$$QQ^\top = Q^\top Q = I_n \quad \text{and} \quad \det(Q) = 1$$
is a group under matrix multiplication, with identity element the identity matrix I_n; as in (6), we have $Q^{-1} = Q^\top$. This group is called the *special orthogonal group* or *rotation group* and is usually denoted by $\mathbf{SO}(n)$.

The groups in (5)–(8) are nonabelian for $n \geq 2$, except for $\mathbf{SO}(2)$ which is abelian (but $\mathbf{O}(2)$ is not abelian).

It is customary to denote the operation of an abelian group G by $+$, in which case the inverse a^{-1} of an element $a \in G$ is denoted by $-a$.

The identity element of a group is *unique*. In fact, we can prove a more general fact:

Proposition C.1. *If a binary operation* $\cdot\colon M \times M \to M$ *is associative and if* $e' \in M$ *is a left identity and* $e'' \in M$ *is a right identity, which means that*
$$e' \cdot a = a \quad \text{for all} \quad a \in M \tag{G2l}$$
and
$$a \cdot e'' = a \quad \text{for all} \quad a \in M, \tag{G2r}$$
then $e' = e''$.

Proof. If we let $a = e''$ in equation (G2l), we get
$$e' \cdot e'' = e'',$$
and if we let $a = e'$ in equation (G2r), we get
$$e' \cdot e'' = e',$$
and thus
$$e' = e' \cdot e'' = e'',$$
as claimed. ☐

Proposition C.1 implies that the identity element of a monoid is unique, and since every group is a monoid, the identity element of a group is unique. Furthermore, every element in a group has a *unique inverse*. This is a consequence of a slightly more general fact:

Proposition C.2. *In a monoid* M *with identity element* e*, if some element* $a \in M$ *has some left inverse* $a' \in M$ *and some right inverse* $a'' \in M$*, which means that*
$$a' \cdot a = e \tag{G3l}$$
and
$$a \cdot a'' = e, \tag{G3r}$$
then $a' = a''$.

Proof. Using (G3l) and the fact that e is an identity element, we have

$$(a' \cdot a) \cdot a'' = e \cdot a'' = a''.$$

Similarly, Using (G3r) and the fact that e is an identity element, we have

$$a' \cdot (a \cdot a'') = a' \cdot e = a'.$$

However, since M is monoid, the operation $\cdot$ is associative, so

$$a' = a' \cdot (a \cdot a'') = (a' \cdot a) \cdot a'' = a'',$$

as claimed. $\square$

Remark: Axioms (G2) and (G3) can be weakened a bit by requiring only (G2r) (the existence of a right identity) and (G3r) (the existence of a right inverse for every element) (or (G2l) and (G3l)). It is a good exercise to prove that the group axioms (G2) and (G3) follow from (G2r) and (G3r).

Definition C.2. If a group G has a finite number n of elements, we say that G is a group of *order* n. If G is infinite, we say that G has *infinite order*. The order of a group is usually denoted by $|G|$ (if G is finite).

Given a group G, for any two subsets $R, S \subseteq G$, we let

$$RS = \{r \cdot s \mid r \in R,\ s \in S\}.$$

In particular, for any $g \in G$, if $R = \{g\}$, we write

$$gS = \{g \cdot s \mid s \in S\},$$

and similarly, if $S = \{g\}$, we write

$$Rg = \{r \cdot g \mid r \in R\}.$$

From now on, we will drop the multiplication sign and write $g_1 g_2$ for $g_1 \cdot g_2$.

Definition C.3. Let G be a group. For any $g \in G$, define L_g, the *left translation by* g, by $L_g(a) = ga$, for all $a \in G$, and R_g, the *right translation by* g, by $R_g(a) = ag$, for all $a \in G$.

The following simple fact is often used.

Proposition C.3. *Given a group G, the translations L_g and R_g are bijections.*

Proof. We show this for L_g, the proof for R_g being similar.

If $L_g(a) = L_g(b)$, then $ga = gb$, and multiplying on the left by g^{-1}, we get $a = b$, so L_g injective. For any $b \in G$, we have $L_g(g^{-1}b) = gg^{-1}b = b$, so L_g is surjective. Therefore, L_g is bijective. $\square$

Definition C.4. Given a group G, a subset H of G is a *subgroup of* G iff

(1) The identity element e of G also belongs to H ($e \in H$);

(2) For all $h_1, h_2 \in H$, we have $h_1 h_2 \in H$;

(3) For all $h \in H$, we have $h^{-1} \in H$.

The proof of the following proposition is left as an exercise.

Proposition C.4. *Given a group G, a subset $H \subseteq G$ is a subgroup of G iff H is nonempty and whenever $h_1, h_2 \in H$, then $h_1 h_2^{-1} \in H$.*

If the group G is finite, then the following criterion can be used.

Proposition C.5. *Given a finite group G, a subset $H \subseteq G$ is a subgroup of G iff*

(1) $e \in H$;

(2) H is closed under multiplication.

Proof. We just have to prove that Condition (3) of Definition C.4 holds. For any $a \in H$, since the left translation L_a is bijective, its restriction to H is injective, and since H is finite, it is also bijective. Since $e \in H$, there is a unique $b \in H$ such that $L_a(b) = ab = e$. However, if a^{-1} is the inverse of a in G, we also have $L_a(a^{-1}) = aa^{-1} = e$, and by injectivity of L_a, we have $a^{-1} = b \in H$. ∎

Example C.2.

1. For any integer $n \in \mathbb{Z}$, the set
$$n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$$
is a subgroup of the group $\mathbb{Z}$.

2. The set of matrices
$$\mathbf{GL}^+(n, \mathbb{R}) = \{A \in \mathbf{GL}(n, \mathbb{R}) \mid \det(A) > 0\}$$
is a subgroup of the group $\mathbf{GL}(n, \mathbb{R})$.

3. The group $\mathbf{SL}(n, \mathbb{R})$ is a subgroup of the group $\mathbf{GL}(n, \mathbb{R})$.

4. The group $\mathbf{O}(n)$ is a subgroup of the group $\mathbf{GL}(n, \mathbb{R})$.

5. The group $\mathbf{SO}(n)$ is a subgroup of the group $\mathbf{O}(n)$, and a subgroup of the group $\mathbf{SL}(n, \mathbb{R})$.

6. It is not hard to show that every 2×2 rotation matrix $R \in \mathbf{SO}(2)$ can be written as

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \text{with} \quad 0 \leq \theta < 2\pi.$$

   Then $\mathbf{SO}(2)$ can be considered as a subgroup of $\mathbf{SO}(3)$ by viewing the matrix

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

   as the matrix

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

7. The set of 2×2 upper-triangular matrices of the form

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \quad a, b, c \in \mathbb{R},\ a, c \neq 0$$

   is a subgroup of the group $\mathbf{GL}(2, \mathbb{R})$.

8. The set V consisting of the four matrices

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$$

   is a subgroup of the group $\mathbf{GL}(2, \mathbb{R})$ called the *Klein four-group*.

Definition C.5. If H is a subgroup of G and $g \in G$ is any element, the sets of the form gH are called *left cosets of H in G* and the sets of the form Hg are called *right cosets of H in G*. The left cosets (resp. right cosets) of H induce an equivalence relation $\sim$ defined as follows: For all $g_1, g_2 \in G$,

$$g_1 \sim g_2 \quad \text{iff} \quad g_1 H = g_2 H$$

(resp. $g_1 \sim g_2$ iff $Hg_1 = Hg_2$). Obviously, $\sim$ is an equivalence relation.

Now, we claim the following fact:

Proposition C.6. *Given a group G and any subgroup H of G, we have $g_1 H = g_2 H$ iff $g_2^{-1} g_1 H = H$ iff $g_2^{-1} g_1 \in H$, for all $g_1, g_2 \in G$.*

Proof. If we apply the bijection $L_{g_2^{-1}}$ to both $g_1 H$ and $g_2 H$ we get $L_{g_2^{-1}}(g_1 H) = g_2^{-1} g_1 H$ and $L_{g_2^{-1}}(g_2 H) = H$, so $g_1 H = g_2 H$ iff $g_2^{-1} g_1 H = H$. If $g_2^{-1} g_1 H = H$, since $1 \in H$, we get $g_2^{-1} g_1 \in H$. Conversely, if $g_2^{-1} g_1 \in H$, since H is a group, the left translation $L_{g_2^{-1} g_1}$ is a bijection of H, so $g_2^{-1} g_1 H = H$. Thus, $g_2^{-1} g_1 H = H$ iff $g_2^{-1} g_1 \in H$. $\square$

It follows that the equivalence class of an element $g \in G$ is the coset gH (resp. Hg). Since L_g is a bijection between H and gH, the cosets gH all have the same cardinality. The map $L_{g^{-1}} \circ R_g$ is a bijection between the left coset gH and the right coset Hg, so they also have the same cardinality. Since the distinct cosets gH form a partition of G, we obtain the following fact:

Proposition C.7. *(Lagrange) For any finite group G and any subgroup H of G, the order h of H divides the order n of G.*

Definition C.6. Given a finite group G and a subgroup H of G, if $n = |G|$ and $h = |H|$, then the ratio n/h is denoted by $(G : H)$ and is called the *index of H in G*.

The index $(G : H)$ is the number of left (and right) cosets of H in G. Proposition C.7 can be stated as

$$|G| = (G : H)|H|.$$

The set of left cosets of H in G (which, in general, is **not** a group) is denoted G/H. The "points" of G/H are obtained by "collapsing" all the elements in a coset into a single element.

Example C.3.

1. Let n be any positive integer, and consider the subgroup $n\mathbb{Z}$ of $\mathbb{Z}$ (under addition). The coset of 0 is the set $\{0\}$, and the coset of any nonzero integer $m \in \mathbb{Z}$ is

$$m + n\mathbb{Z} = \{m + nk \mid k \in \mathbb{Z}\}.$$

By dividing m by n, we have $m = nq + r$ for some unique r such that $0 \leq r \leq n - 1$. But then we see that r is the smallest positive element of the coset $m + n\mathbb{Z}$. This implies that there is a bijection betwen the cosets of the subgroup $n\mathbb{Z}$ of $\mathbb{Z}$ and the set of residues $\{0, 1, \ldots, n-1\}$ modulo n, or equivalently a bijection with $\mathbb{Z}/n\mathbb{Z}$.

2. The cosets of $\mathbf{SL}(n, \mathbb{R})$ in $\mathbf{GL}(n, \mathbb{R})$ are the sets of matrices

$$A\,\mathbf{SL}(n, \mathbb{R}) = \{AB \mid B \in \mathbf{SL}(n, \mathbb{R})\},\ A \in \mathbf{GL}(n, \mathbb{R}).$$

Since A is invertible, $\det(A) \neq 0$, and we can write $A = (\det(A))^{1/n}((\det(A))^{-1/n}A)$ if $\det(A) > 0$ and $A = (-\det(A))^{1/n}((-\det(A))^{-1/n}A)$ if $\det(A) < 0$. But we have $(\det(A))^{-1/n}A \in \mathbf{SL}(n, \mathbb{R})$ if $\det(A) > 0$ and $-(-\det(A))^{-1/n}A \in \mathbf{SL}(n, \mathbb{R})$ if $\det(A) < 0$, so the coset $A\,\mathbf{SL}(n, \mathbb{R})$ contains the matrix

$$(\det(A))^{1/n}I_n \quad \text{if} \quad \det(A) > 0, \quad -(-\det(A))^{1/n}I_n \quad \text{if} \quad \det(A) < 0.$$

It follows that there is a bijection between the cosets of $\mathbf{SL}(n, \mathbb{R})$ in $\mathbf{GL}(n, \mathbb{R})$ and $\mathbb{R}$.

3. The cosets of $\mathbf{SO}(n)$ in $\mathbf{GL}^+(n, \mathbb{R})$ are the sets of matrices

$$A\,\mathbf{SO}(n) = \{AQ \mid Q \in \mathbf{SO}(n)\},\ A \in \mathbf{GL}^+(n, \mathbb{R}).$$

It can be shown (using the polar form for matrices) that there is a bijection between the cosets of $\mathbf{SO}(n)$ in $\mathbf{GL}^+(n, \mathbb{R})$ and the set of $n \times n$ symmetric, positive, definite matrices; these are the symmetric matrices whose eigenvalues are strictly positive.

4. The cosets of $\mathbf{SO}(2)$ in $\mathbf{SO}(3)$ are the sets of matrices

$$Q\,\mathbf{SO}(2) = \{QR \mid R \in \mathbf{SO}(2)\},\ Q \in \mathbf{SO}(3).$$

The group $\mathbf{SO}(3)$ moves the points on the sphere S^2 in $\mathbb{R}^3$, namely for any $x \in S^2$,

$$x \mapsto Qx \quad \text{for any rotation } Q \in \mathbf{SO}(3).$$

Here,

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}.$$

Let $N = (0, 0, 1)$ be the north pole on the sphere S^2. Then it is not hard to show that $\mathbf{SO}(2)$ is precisely the subgroup of $\mathbf{SO}(3)$ that leaves N fixed. As a consequence, all rotations QR in the coset $Q\,\mathbf{SO}(2)$ map N to the same point $QN \in S^2$, and it can be shown that there is a bijection between the cosets of $\mathbf{SO}(2)$ in $\mathbf{SO}(3)$ and the points on S^2. The surjectivity of this map has to do with the fact that the action of $\mathbf{SO}(3)$ on S^2 is transitive, which means that for any point $x \in S^2$, there is some rotation $Q \in \mathbf{SO}(3)$ such that $QN = x$.

It is tempting to define a multiplication operation on left cosets (or right cosets) by setting

$$(g_1H)(g_2H) = (g_1g_2)H,$$

but this operation is not well defined in general, unless the subgroup H possesses a special property. In Example C.3, it is possible to define multiplication of cosets in (1), but it is not possible in (2) and (3).

The property of the subgroup H that allows defining a multiplication operation on left cosets is typical of the kernels of group homomorphisms, so we are led to the following definition.

Definition C.7. Given any two groups G and G', a function $\varphi\colon G \to G'$ is a *homomorphism* iff

$$\varphi(g_1g_2) = \varphi(g_1)\varphi(g_2), \quad \text{for all } g_1, g_2 \in G.$$

Taking $g_1 = g_2 = e$ (in G), we see that

$$\varphi(e) = e',$$

and taking $g_1 = g$ and $g_2 = g^{-1}$, we see that

$$\varphi(g^{-1}) = (\varphi(g))^{-1}.$$

Example C.4.

1. The map $\varphi\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ given by $\varphi(m) = m \bmod n$ for all $m \in \mathbb{Z}$ is a homomorphism.

2. The map $\det\colon \mathbf{GL}(n, \mathbb{R}) \to \mathbb{R}$ is a homomorphism because $\det(AB) = \det(A)\det(B)$ for any two matrices A, B. Similarly, the map $\det\colon \mathbf{O}(n) \to \mathbb{R}$ is a homomorphism.

If $\varphi\colon G \to G'$ and $\psi\colon G' \to G''$ are group homomorphisms, then $\psi \circ \varphi\colon G \to G''$ is also a homomorphism. If $\varphi\colon G \to G'$ is a homomorphism of groups, and if $H \subseteq G$, $H' \subseteq G'$ are two subgroups, then it is easily checked that

$$\mathrm{Im}\ H = \varphi(H) = \{\varphi(g) \mid g \in H\}$$

is a subgroup of G' and

$$\varphi^{-1}(H') = \{g \in G \mid \varphi(g) \in H'\}$$

is a subgroup of G. In particular, when $H' = \{e'\}$, we obtain the *kernel*, $\mathrm{Ker}\ \varphi$, of φ.

Definition C.8. If $\varphi\colon G \to G'$ is a homomorphism of groups, and if $H \subseteq G$ is a subgroup of G, then the subgroup of G',

$$\mathrm{Im}\ H = \varphi(H) = \{\varphi(g) \mid g \in H\},$$

is called the *image of H by φ*, and the subgroup of G,

$$\mathrm{Ker}\ \varphi = \{g \in G \mid \varphi(g) = e'\},$$

is called the *kernel* of φ.

Example C.5.

1. The kernel of the homomorphism $\varphi\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ is $n\mathbb{Z}$.

2. The kernel of the homomorphism $\det\colon \mathbf{GL}(n, \mathbb{R}) \to \mathbb{R}$ is $\mathbf{SL}(n, \mathbb{R})$. Similarly, the kernel of the homomorphism $\det\colon \mathbf{O}(n) \to \mathbb{R}$ is $\mathbf{SO}(n)$.

The following characterization of the injectivity of a group homomorphism is used all the time.

Proposition C.8. *If $\varphi\colon G \to G'$ is a homomorphism of groups, then $\varphi\colon G \to G'$ is injective iff* $\mathrm{Ker}\ \varphi = \{e\}$. *(We also write* $\mathrm{Ker}\ \varphi = (0)$.*)*

Proof. Assume φ is injective. Since $\varphi(e) = e'$, if $\varphi(g) = e'$, then $\varphi(g) = \varphi(e)$, and by injectivity of φ we must have $g = e$, so $\mathrm{Ker}\ \varphi = \{e\}$.

Conversely, assume that $\mathrm{Ker}\ \varphi = \{e\}$. If $\varphi(g_1) = \varphi(g_2)$, then by multiplication on the left by $(\varphi(g_1))^{-1}$ we get

$$e' = (\varphi(g_1))^{-1}\varphi(g_1) = (\varphi(g_1))^{-1}\varphi(g_2),$$

and since φ is a homomorphism $(\varphi(g_1))^{-1} = \varphi(g_1^{-1})$, so

$$e' = (\varphi(g_1))^{-1}\varphi(g_2) = \varphi(g_1^{-1})\varphi(g_2) = \varphi(g_1^{-1}g_2).$$

This shows that $g_1^{-1}g_2 \in \mathrm{Ker}\ \varphi$, but since $\mathrm{Ker}\ \varphi = \{e\}$ we have $g_1^{-1}g_2 = e$, and thus $g_2 = g_1$, proving that φ is injective. $\square$

Definition C.9. We say that a group homomorphism $\varphi\colon G \to G'$ is an *isomorphism* if there is a homomorphism $\psi\colon G' \to G$, so that

$$\psi \circ \varphi = \mathrm{id}_G \quad \text{and} \quad \varphi \circ \psi = \mathrm{id}_{G'}. \tag{†}$$

If φ is an isomorphism we say that the groups G and G' are *isomorphic*. When $G' = G$, a group isomorphism is called an *automorphism*.

The reasoning used in the proof of Proposition C.2 shows that if a a group homomorphism $\varphi\colon G \to G'$ is an isomorphism, then the homomorphism $\psi\colon G' \to G$ satisfying Condition (†) is unique. This homomorphism is denoted φ^{-1}.

The left translations L_g and the right translations R_g are automorphisms of G.

Suppose $\varphi\colon G \to G'$ is a bijective homomorphism, and let φ^{-1} be the inverse of φ (as a function). Then for all $a, b \in G$, we have

$$\varphi(\varphi^{-1}(a)\varphi^{-1}(b)) = \varphi(\varphi^{-1}(a))\varphi(\varphi^{-1}(b)) = ab,$$

and so

$$\varphi^{-1}(ab) = \varphi^{-1}(a)\varphi^{-1}(b),$$

which proves that φ^{-1} is a homomorphism. Therefore, we proved the following fact.

Proposition C.9. *A bijective group homomorphism $\varphi\colon G \to G'$ is an isomorphism.*

Observe that the property

$$gH = Hg, \quad \text{for all } g \in G. \tag{$*$}$$

is equivalent by multiplication on the right by g^{-1} to

$$gHg^{-1} = H, \quad \text{for all } g \in G,$$

and the above is equivalent to

$$gHg^{-1} \subseteq H, \quad \text{for all } g \in G. \tag{$**$}$$

This is because $gHg^{-1} \subseteq H$ implies $H \subseteq g^{-1}Hg$, and this for all $g \in G$.

Proposition C.10. *Let $\varphi\colon G \to G'$ be a group homomorphism. Then $H = \mathrm{Ker}\ \varphi$ satisfies Property ($**$), and thus Property ($*$).*

Proof. We have

$$\varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g^{-1}) = \varphi(g)e'\varphi(g)^{-1} = \varphi(g)\varphi(g)^{-1} = e',$$

for all $h \in H = \mathrm{Ker}\ \varphi$ and all $g \in G$. Thus, by definition of $H = \mathrm{Ker}\ \varphi$, we have $gHg^{-1} \subseteq H$. $\square$

Definition C.10. For any group G, a subgroup N of G is a *normal subgroup* of G iff

$$gNg^{-1} = N, \quad \text{for all } g \in G.$$

This is denoted by $N \lhd G$.

Proposition C.10 shows that the kernel $\mathrm{Ker}\ \varphi$ of a homomorphism $\varphi\colon G \to G'$ is a normal subgroup of G.

Observe that if G is abelian, then *every* subgroup of G is normal.

Consider Example C.2. Let $R \in \mathbf{SO}(2)$ and $A \in \mathbf{SL}(2, \mathbb{R})$ be the matrices

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Then

$$A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

and we have

$$ARA^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix},$$

and clearly $ARA^{-1} \notin \mathbf{SO}(2)$. Therefore $\mathbf{SO}(2)$ is not a normal subgroup of $\mathbf{SL}(2, \mathbb{R})$. The same counter-example shows that $\mathbf{O}(2)$ is not a normal subgroup of $\mathbf{GL}(2, \mathbb{R})$.

Let $R \in \mathbf{SO}(2)$ and $Q \in \mathbf{SO}(3)$ be the matrices

$$R = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then

$$Q^{-1} = Q^{\top} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

and we have

$$QRQ^{-1} = \begin{pmatrix}1 & 0 & 0\\ 0 & 0 & -1\\ 0 & 1 & 0\end{pmatrix}\begin{pmatrix}0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & -1 & 0\end{pmatrix} = \begin{pmatrix}0 & -1 & 0\\ 0 & 0 & -1\\ 1 & 0 & 0\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & -1 & 0\end{pmatrix}$$
$$= \begin{pmatrix}0 & 0 & -1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{pmatrix}.$$

Observe that $QRQ^{-1} \notin \mathbf{SO}(2)$, so $\mathbf{SO}(2)$ is not a normal subgroup of $\mathbf{SO}(3)$.

Let T and $A \in \mathbf{GL}(2, \mathbb{R})$ be the following matrices

$$T = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \quad A = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}.$$

We have

$$A^{-1} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} = A,$$

and

$$ATA^{-1} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}.$$

The matrix T is upper triangular, but ATA^{-1} is not, so the group of 2×2 upper triangular matrices is not a normal subgroup of $\mathbf{GL}(2, \mathbb{R})$.

Let $Q \in V$ and $A \in \mathbb{GL}(2, \mathbb{R})$ be the following matrices

$$Q = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad A = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}.$$

We have

$$A^{-1} = \begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix}$$

and

$$AQA^{-1} = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}1 & -1\\ 0 & -1\end{pmatrix}\begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}1 & -2\\ 0 & -1\end{pmatrix}.$$

Clearly $AQA^{-1} \notin V$, which shows that the Klein four group is not a normal subgroup of $\mathbf{GL}(2, \mathbb{R})$.

The reader should check that the subgroups $n\mathbb{Z}$, $\mathbf{GL}^+(n, \mathbb{R})$, $\mathbf{SL}(n, \mathbb{R})$, and $\mathbf{SO}(n, \mathbb{R})$ as a subgroup of $\mathbf{O}(n, \mathbb{R})$, are normal subgroups.

If N is a normal subgroup of G, the equivalence relation $\sim$ induced by left cosets (see Definition C.5) is the same as the equivalence induced by right cosets. Furthermore, this equivalence relation is a *congruence*, which means that: For all $g_1, g_2, g'_1, g'_2 \in G$,

(1) If $g_1N = g_1'N$ and $g_2N = g_2'N$, then $g_1g_2N = g_1'g_2'N$, and

(2) If $g_1N = g_2N$, then $g_1^{-1}N = g_2^{-1}N$.

As a consequence, we can define a group structure on the set $G/\sim$ of equivalence classes modulo $\sim$, by setting

$$(g_1N)(g_2N) = (g_1g_2)N.$$

Definition C.11. Let G be a group and N be a normal subgroup of G. The group obtained by defining the multiplication of (left) cosets by

$$(g_1N)(g_2N) = (g_1g_2)N, \quad g_1, g_2 \in G$$

is denoted G/N, and called the *quotient of G by N*. The equivalence class gN of an element $g \in G$ is also denoted $\overline{g}$ (or $[g]$). The map $\pi\colon G \to G/N$ given by

$$\pi(g) = \overline{g} = gN$$

is a group homomorphism called the *canonical projection*.

Since the kernel of a homomorphism is a normal subgroup, we obtain the following very useful result.

Proposition C.11. *Given a homomorphism of groups $\varphi\colon G \to G'$, the groups $G/\mathrm{Ker}\ \varphi$ and $\mathrm{Im}\ \varphi = \varphi(G)$ are isomorphic.*

Proof. Since φ is surjective onto its image, we may assume that φ is surjective, so that $G' = \mathrm{Im}\ \varphi$. We define a map $\overline{\varphi}\colon G/\mathrm{Ker}\ \varphi \to G'$ as follows:

$$\overline{\varphi}(\overline{g}) = \varphi(g), \quad g \in G.$$

We need to check that the definition of this map does not depend on the representative chosen in the coset $\overline{g} = g\,\mathrm{Ker}\ \varphi$, and that it is a homomorphism. If g' is another element in the coset $g\,\mathrm{Ker}\ \varphi$, which means that $g' = gh$ for some $h \in \mathrm{Ker}\,\varphi$, then

$$\varphi(g') = \varphi(gh) = \varphi(g)\varphi(h) = \varphi(g)e' = \varphi(g),$$

since $\varphi(h) = e'$ as $h \in \mathrm{Ker}\ \varphi$. This shows that

$$\overline{\varphi}(\overline{g'}) = \varphi(g') = \varphi(g) = \overline{\varphi}(\overline{g}),$$

so the map $\overline{\varphi}$ is well defined. It is a homomorphism because

$$\begin{aligned}
\overline{\varphi}(\overline{g}\overline{g'}) &= \overline{\varphi}(\overline{gg'}) \\
&= \varphi(gg') \\
&= \varphi(g)\varphi(g') \\
&= \overline{\varphi}(\overline{g})\overline{\varphi}(\overline{g'}).
\end{aligned}$$

The map $\overline{\varphi}$ is injective because $\overline{\varphi}(\overline{g}) = e'$ iff $\varphi(g) = e'$ iff $g \in \mathrm{Ker}\ \varphi$, iff $\overline{g} = \overline{e}$. The map $\overline{\varphi}$ is surjective because φ is surjective. Therefore $\overline{\varphi}$ is a bijective homomorphism, and thus an isomorphism, as claimed. $\square$

Proposition C.11 is called the *first isomorphism theorem*.

A useful way to construct groups is the *direct product* construction.

Definition C.12. Given two groups G an H, we let $G \times H$ be the Cartesian product of the sets G and H with the multiplication operation $\cdot$ given by

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2).$$

It is immediately verified that $G \times H$ is a group called the *direct product* of G and H.

Similarly, given any n groups $G_1, \ldots, G_n$, we can define the direct product $G_1 \times \cdots \times G_n$ is a similar way.

If G is an abelian group and $H_1, \ldots, H_n$ are subgroups of G, the situation is simpler. Consider the map

$$a \colon H_1 \times \cdots \times H_n \to G$$

given by

$$a(h_1, \ldots, h_n) = h_1 + \cdots + h_n,$$

using $+$ for the operation of the group G. It is easy to verify that a is a group homomorphism, so its image is a subgroup of G denoted by $H_1 + \cdots + H_n$, and called the *sum* of the groups H_i. The following proposition will be needed.

Proposition C.12. *Given an abelian group G, if H_1 and H_2 are any subgroups of G such that $H_1 \cap H_2 = \{0\}$, then the map a is an isomorphism*

$$a \colon H_1 \times H_2 \to H_1 + H_2.$$

Proof. The map is surjective by definition, so we just have to check that it is injective. For this, we show that $\mathrm{Ker}\, a = \{(0, 0)\}$. We have $a(a_1, a_2) = 0$ iff $a_1 + a_2 = 0$ iff $a_1 = -a_2$. Since $a_1 \in H_1$ and $a_2 \in H_2$, we see that $a_1, a_2 \in H_1 \cap H_2 = \{0\}$, so $a_1 = a_2 = 0$, which proves that $\mathrm{Ker}\, a = \{(0, 0)\}$. $\square$

Under the conditions of Proposition C.12, namely $H_1 \cap H_2 = \{0\}$, the group $H_1 + H_2$ is called the *direct sum* of H_1 and H_2; it is denoted by $H_1 \oplus H_2$, and we have an isomorphism $H_1 \times H_2 \cong H_1 \oplus H_2$.

C.2 Group Actions: Part I, Definition and Examples

If X is a set (usually some kind of geometric space, for example, the sphere in $\mathbb{R}^3$, the upper half-plane, etc.), the "symmetries" of X are often captured by the action of a group G on X. In fact, if G is a Lie group and the action satisfies some simple properties, the set X can be given a manifold structure which makes it a projection (quotient) of G, a so-called "homogeneous space."

Definition C.13. Given a set X and a group G, a *left action of G on X* (for short, an *action of G on X*) is a function $\varphi\colon G \times X \to X$, such that:

(1) For all $g, h \in G$ and all $x \in X$,
$$\varphi(g, \varphi(h, x)) = \varphi(gh, x),$$

(2) For all $x \in X$,
$$\varphi(1, x) = x,$$
where $1 \in G$ is the identity element of G.

To alleviate the notation, we usually write $g \cdot x$ or even gx for $\varphi(g, x)$, in which case the above axioms read:

(1) For all $g, h \in G$ and all $x \in X$,
$$g \cdot (h \cdot x) = gh \cdot x,$$

(2) For all $x \in X$,
$$1 \cdot x = x.$$

The set X is called a *(left) G-set*. The action φ is *faithful* or *effective* iff for every g, if $g \cdot x = x$ for all $x \in X$, then $g = 1$. Faithful means that if the action of some element g behaves like the identity, then g must be the identity element. The action φ is *transitive* iff for any two elements $x, y \in X$, there is some $g \in G$ so that $g \cdot x = y$.

Given an action $\varphi\colon G \times X \to X$, for every $g \in G$, we have a function $\varphi_g\colon X \to X$ defined by
$$\varphi_g(x) = g \cdot x, \quad \text{for all } x \in X.$$
Observe that φ_g has $\varphi_{g^{-1}}$ as inverse, since
$$\varphi_{g^{-1}}(\varphi_g(x)) = \varphi_{g^{-1}}(g \cdot x) = g^{-1} \cdot (g \cdot x) = (g^{-1}g) \cdot x = 1 \cdot x = x,$$
and similarly, $\varphi_g \circ \varphi_{g^{-1}} = \mathrm{id}$. Therefore, φ_g is a bijection of X; that is, φ_g is a permutation of X. Moreover, we check immediately that
$$\varphi_g \circ \varphi_h = \varphi_{gh},$$
so the map $g \mapsto \varphi_g$ is a group homomorphism from G to $\mathfrak{S}_X$, the group of permutations of X. With a slight abuse of notation, this group homomorphism $G \longrightarrow \mathfrak{S}_X$ is also denoted φ.

Conversely, it is easy to see that any group homomorphism $\varphi\colon G \to \mathfrak{S}_X$ yields a group action $\cdot\colon G \times X \longrightarrow X$, by setting
$$g \cdot x = \varphi(g)(x).$$
Observe that an action φ is faithful iff the group homomorphism $\varphi\colon G \to \mathfrak{S}_X$ is injective, i.e. iff φ has a trivial kernel. Also, we have $g \cdot x = y$ iff $g^{-1} \cdot y = x$, since $(gh) \cdot x = g \cdot (h \cdot x)$ and $1 \cdot x = x$, for all $g, h \in G$ and all $x \in X$.

Definition C.14. Given two G-sets X and Y, a function $f\colon X \to Y$ is said to be *equivariant*, or a *G-map*, iff for all $x \in X$ and all $g \in G$, we have

$$f(g \cdot x) = g \cdot f(x).$$

Equivalently, if the G-actions are denoted by $\varphi\colon G \times X \to X$ and $\psi\colon G \times Y \to Y$, we have the following commutative diagram for all $g \in G$:

$$\begin{tikzcd} X \arrow[r, "\varphi_g"] \arrow[d, "f"'] & X \arrow[d, "f"] \\ Y \arrow[r, "\psi_g"'] & Y. \end{tikzcd}$$

Remark: We can also define a *right action* $\cdot\colon X \times G \to X$ of a group G on a set X as a map satisfying the conditions

(1) For all $g, h \in G$ and all $x \in X$,

$$(x \cdot g) \cdot h = x \cdot gh,$$

(2) For all $x \in X$,

$$x \cdot 1 = x.$$

Every notion defined for left actions is also defined for right actions in the obvious way.

However, one change is necessary. For every $g \in G$, the map $\varphi_g\colon X \to X$ must be defined as

$$\varphi_g(x) = x \cdot g^{-1},$$

in order for the map $g \mapsto \varphi_g$ from G to $\mathfrak{S}_X$ to be a homomorphism ($\varphi_g \circ \varphi_h = \varphi_{gh}$). Conversely, given a homomorphism $\varphi\colon G \to \mathfrak{S}_X$, we get a right action $\cdot\colon X \times G \longrightarrow X$ by setting

$$x \cdot g = \varphi(g^{-1})(x).$$

Here are some examples of (left) group actions.

Example C.6. The unit sphere S^2 (more generally, S^{n-1}).

Recall that for any $n \geq 1$, the *(real) unit sphere* S^{n-1} is the set of points in $\mathbb{R}^n$ given by

$$S^{n-1} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 = 1\}.$$

In particular, S^2 is the usual sphere in $\mathbb{R}^3$. Since the group $\mathbf{SO}(3) = \mathbf{SO}(3, \mathbb{R})$ consists of (orientation preserving) linear isometries, i.e., *linear* maps that are distance preserving (and of determinant $+1$), and every linear map leaves the origin fixed, we see that any rotation maps S^2 into itself.

Beware that this would be false if we considered the group of *affine* isometries $\mathbf{SE}(3)$ of $\mathbb{E}^3$. For example, a screw motion does *not* map S^2 into itself, even though it is distance preserving, because the origin is translated.

Thus, for $X = S^2$ and $G = \mathbf{SO}(3)$, we have an action $\cdot\colon \mathbf{SO}(3) \times S^2 \to S^2$, given by the matrix multiplication

$$R \cdot x = Rx.$$

The verification that the above is indeed an action is trivial. This action is transitive. This is because, for any two points x, y on the sphere S^2, there is a rotation whose axis is perpendicular to the plane containing x, y and the center O of the sphere (this plane is not unique when x and y are antipodal, i.e., on a diameter) mapping x to y. See Figure C.1.

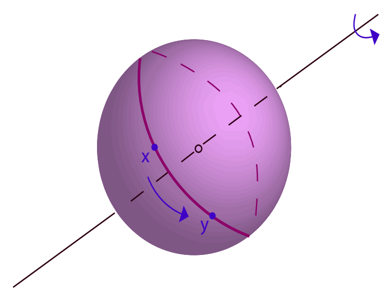


Figure C.1: The rotation which maps x to y.

Similarly, for any $n \geq 1$, let $X = S^{n-1}$ and $G = \mathbf{SO}(n)$ and define the action $\cdot\colon \mathbf{SO}(n) \times S^{n-1} \to S^{n-1}$ as $R \cdot x = Rx$. It is easy to show that this action is transitive.

Analogously, we can define the *(complex) unit sphere* Σ^{n-1}, as the set of points in $\mathbb{C}^n$ given by

$$\Sigma^{n-1} = \{(z_1, \ldots, z_n) \in \mathbb{C}^n \mid z_1\overline{z}_1 + \cdots + z_n\overline{z}_n = 1\}.$$

If we write $z_j = x_j + iy_j$, with $x_j, y_j \in \mathbb{R}$, then

$$\Sigma^{n-1} = \{(x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{R}^{2n} \mid x_1^2 + \cdots + x_n^2 + y_1^2 + \cdots + y_n^2 = 1\}.$$

Therefore, we can view the complex sphere Σ^{n-1} (in $\mathbb{C}^n$) as the real sphere S^{2n-1} (in $\mathbb{R}^{2n}$). By analogy with the real case, we can define for $X = \Sigma^{n-1}$ and $G = \mathbf{SU}(n)$ an action $\cdot\colon \mathbf{SU}(n) \times \Sigma^{n-1} \to \Sigma^{n-1}$ of the group $\mathbf{SU}(n)$ of *linear* maps of $\mathbb{C}^n$ preserving the Hermitian inner product (and the origin, as all linear maps do), and this action is transitive.

One should not confuse the unit sphere Σ^{n-1} with the hypersurface $S^{n-1}_{\mathbb{C}}$, given by

$$S^{n-1}_{\mathbb{C}} = \{(z_1, \ldots, z_n) \in \mathbb{C}^n \mid z_1^2 + \cdots + z_n^2 = 1\}.$$

For instance, one should check that a line L through the origin intersects Σ^{n-1} in a circle, whereas it intersects $S^{n-1}_{\mathbb{C}}$ in exactly two points! Recall for a fixed $u = (x_1, \ldots x_n, y_1, \ldots y_n) \in \mathbb{C}^n$, that $L = \{\gamma u \mid \gamma \in \mathbb{C}\}$. Since $\gamma = \rho(\cos\theta + i\sin\theta)$, we deduce that L is actually the two dimensional subspace through the origin spanned by the orthogonal vectors $(x_1, \ldots x_n, y_1, \ldots y_n)$ and $(-y_1, \cdots - y_n, x_1, \ldots x_n)$.

Example C.7. The upper half-plane.

The *upper half-plane* H is the open subset of $\mathbb{R}^2$ consisting of all points $(x, y) \in \mathbb{R}^2$, with $y > 0$. It is convenient to identify H with the set of complex numbers $z \in \mathbb{C}$ such that $\Im z > 0$. Then we can let $X = H$ and $G = \mathbf{SL}(2, \mathbb{R})$ and define an action $\cdot\colon \mathbf{SL}(2, \mathbb{R}) \times H \to H$ of the group $\mathbf{SL}(2, \mathbb{R})$ on H, as follows: For any $z \in H$, for any $A \in \mathbf{SL}(2, \mathbb{R})$,

$$A \cdot z = \frac{az + b}{cz + d},$$

where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $ad - bc = 1$.

It is easily verified that $A \cdot z$ is indeed always well defined and in H when $z \in H$ (check this). To see why this action is transitive, let z and w be two arbitrary points of H where $z = x + iy$ and $w = u + iv$ with $x, u \in \mathbb{R}$ and $y, v \in \mathbb{R}^+$ (i.e. y and v are positive real numbers). Define $A = \begin{pmatrix} \sqrt{\frac{v}{y}} & \frac{uy - vx}{\sqrt{yv}} \\ 0 & \sqrt{\frac{y}{v}} \end{pmatrix}$. Note that $A \in \mathbf{SL}(2, \mathbb{R})$. A routine calculation shows that $A \cdot z = w$.

Before introducing Example C.8, we need to define the groups of Möbius transformations and the Riemann sphere. Maps of the form

$$z \mapsto \frac{az + b}{cz + d},$$

where $z \in \mathbb{C}$ and $ad - bc = 1$, are called *Möbius transformations*. Here, $a, b, c, d \in \mathbb{R}$, but in general, we allow $a, b, c, d \in \mathbb{C}$. Actually, these transformations are not necessarily defined everywhere on $\mathbb{C}$, for example, for $z = -d/c$ if $c \neq 0$. To fix this problem, we add a "point at infinity" ∞ to $\mathbb{C}$, and define Möbius transformations as functions $\mathbb{C} \cup \{\infty\} \longrightarrow \mathbb{C} \cup \{\infty\}$. If $c = 0$, the Möbius transformation sends ∞ to itself, otherwise, $-d/c \mapsto \infty$ and $\infty \mapsto a/c$.

The space $\mathbb{C} \cup \{\infty\}$ can be viewed as the plane $\mathbb{R}^2$ extended with a point at infinity. Using a stereographic projection from the sphere S^2 to the plane (say from the north pole to the

equatorial plane), we see that there is a bijection between the sphere S^2 and $\mathbb{C} \cup \{\infty\}$. More precisely, the *stereographic projection* σ_N of the sphere S^2 from the north pole $N = (0,0,1)$ to the plane $z = 0$ (extended with the point at infinity ∞) is given by

$$(x,y,z) \in S^2 - \{(0,0,1)\} \mapsto \left(\frac{x}{1-z}, \frac{y}{1-z}\right) = \frac{x+iy}{1-z} \in \mathbb{C}, \quad \text{with} \quad (0,0,1) \mapsto \infty.$$

The inverse stereographic projection σ_N^{-1} is given by

$$(u,v) \mapsto \left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, \frac{u^2+v^2-1}{u^2+v^2+1}\right), \quad \text{with} \quad \infty \mapsto (0,0,1).$$

Intuitively, the inverse stereographic projection "wraps" the equatorial plane around the sphere. See Figure C.2.

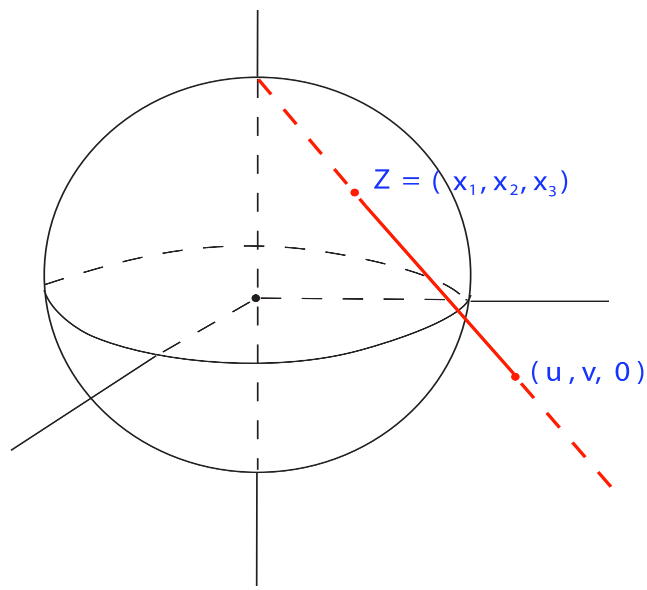


Figure C.2: Inverse stereographic projection.

The space $\mathbb{C} \cup \{\infty\}$ is known as the *Riemann sphere*. We will see shortly that $\mathbb{C} \cup \{\infty\} \cong S^2$ is also the complex projective line $\mathbb{CP}^1$. In summary, Möbius transformations are bijections of the Riemann sphere. It is easy to check that these transformations form a group under composition for all $a,b,c,d \in \mathbb{C}$, with $ad - bc = 1$. This is the *Möbius group*, denoted $\mathbf{M\ddot{o}b}^+$. The Möbius transformations corresponding to the case $a,b,c,d \in \mathbb{R}$, with $ad - bc = 1$ form a subgroup of $\mathbf{M\ddot{o}b}^+$ denoted $\mathbf{M\ddot{o}b}^+_{\mathbb{R}}$.

The map from $\mathbf{SL}(2,\mathbb{C})$ to $\mathbf{M\ddot{o}b}^+$ that sends $A \in \mathbf{SL}(2,\mathbb{C})$ to the corresponding Möbius transformation is a surjective group homomorphism, and one checks easily that its kernel is $\{-I, I\}$ (where I is the 2×2 identity matrix). Therefore, the Möbius group $\mathbf{M\ddot{o}b}^+$ is

isomorphic to the quotient group $\mathbf{SL}(2,\mathbb{C})/\{-I, I\}$, denoted $\mathbf{PSL}(2,\mathbb{C})$. This latter group turns out to be the group of projective transformations of the projective space $\mathbb{CP}^1$. The same reasoning shows that the subgroup $\mathbf{M\ddot{o}b}^+_{\mathbb{R}}$ is isomorphic to $\mathbf{SL}(2,\mathbb{R})/\{-I, I\}$, denoted $\mathbf{PSL}(2,\mathbb{R})$.

Example C.8. The Riemann sphere $\mathbb{C} \cup \{\infty\}$.

Let $X = \mathbb{C} \cup \{\infty\}$ and $G = \mathbf{SL}(2,\mathbb{C})$. The group $\mathbf{SL}(2,\mathbb{C})$ acts on $\mathbb{C} \cup \{\infty\} \cong S^2$ the same way that $\mathbf{SL}(2,\mathbb{R})$ acts on H, namely: For any $A \in \mathbf{SL}(2,\mathbb{C})$, for any $z \in \mathbb{C} \cup \{\infty\}$,

$$A \cdot z = \frac{az + b}{cz + d},$$

where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{with} \quad ad - bc = 1.$$

This action is transitive, an exercise we leave for the reader.

Example C.9. The unit disk.

One may recall from complex analysis that the scaled (complex) Möbius transformation

$$z \mapsto \frac{z - i}{z + i}$$

is a biholomorphic or analytic isomorphism between the upper half plane H and the open unit disk

$$D = \{z \in \mathbb{C} \mid |z| < 1\}.$$

As a consequence, it is possible to define a transitive action of $\mathbf{SL}(2,\mathbb{R})$ on D. This can be done in a more direct fashion, using a group isomorphic to $\mathbf{SL}(2,\mathbb{R})$, namely, $\mathbf{SU}(1,1)$ (a group of complex matrices), but we don't want to do this right now.

Example C.10. The unit Riemann sphere revisited.

Another interesting action is the action of $\mathbf{SU}(2)$ on the extended plane $\mathbb{C} \cup \{\infty\}$. Recall that the group $\mathbf{SU}(2)$ consists of all complex matrices of the form

$$A = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \qquad \alpha, \beta \in \mathbb{C}, \qquad \alpha\overline{\alpha} + \beta\overline{\beta} = 1,$$

Let $X = \mathbb{C} \cup \{\infty\}$ and $G = \mathbf{SU}(2)$. The action $\cdot\colon \mathbf{SU}(2) \times (\mathbb{C} \cup \{\infty\}) \to \mathbb{C} \cup \{\infty\}$ is given by

$$A \cdot w = \frac{\alpha w + \beta}{-\overline{\beta} w + \overline{\alpha}}, \quad w \in \mathbb{C} \cup \{\infty\}.$$

This action is transitive, but the proof of this fact relies on the surjectivity of the group homomorphism

$$\rho\colon \mathbf{SU}(2) \to \mathbf{SO}(3)$$

defined below, and the stereographic projection σ_N from S^2 onto $\mathbb{C} \cup \{\infty\}$. In particular, take $z, w \in \mathbb{C} \cup \{\infty\}$, use the inverse stereographic projection to obtain two points on S^2, namely $\sigma_N^{-1}(z)$ and $\sigma_N^{-1}(w)$. Then apply the appropriate rotation $R \in \mathbf{SO}(3)$ to map $\sigma_N^{-1}(z)$ onto $\sigma_N^{-1}(w)$. Such a rotation exists by the argument presented in Example C.6. Since $\rho\colon \mathbf{SU}(2) \to \mathbf{SO}(3)$ is surjective (see below), we know there must exist $A \in \mathbf{SU}(2)$ such that $\rho(A) = R$ and $A \cdot z = w$.

Using the stereographic projection σ_N from S^2 onto $\mathbb{C} \cup \{\infty\}$ and its inverse σ_N^{-1}, we can define an action of $\mathbf{SU}(2)$ on S^2 by

$$A \cdot (x, y, z) = \sigma_N^{-1}(A \cdot \sigma_N(x, y, z)), \quad (x, y, z) \in S^2.$$

Although this is not immediately obvious, it turns out that $\mathbf{SU}(2)$ acts on S^2 by maps that are restrictions of linear maps to S^2, and since these linear maps preserve S^2, they are orthogonal transformations. Thus, we obtain a continuous (in fact, smooth) group homomorphism

$$\rho\colon \mathbf{SU}(2) \to \mathbf{O}(3).$$

Since $\mathbf{SU}(2)$ is connected and ρ is continuous, the image of $\mathbf{SU}(2)$ is contained in the connected component of I in $\mathbf{O}(3)$, namely $\mathbf{SO}(3)$, so ρ is a homomorphism

$$\rho\colon \mathbf{SU}(2) \to \mathbf{SO}(3).$$

We will see that this homomorphism is surjective and that its kernel is $\{I, -I\}$. The upshot is that we have an isomorphism

$$\mathbf{SO}(3) \cong \mathbf{SU}(2)/\{I, -I\}.$$

The homomorphism ρ is a way of describing how a unit quaternion (any element of $\mathbf{SU}(2)$) induces a rotation, *via* the stereographic projection and its inverse. If we write $\alpha = a + ib$ and $\beta = c + id$, a rather tedious computation yields

$$\rho(A) = \begin{pmatrix} a^2 - b^2 - c^2 + d^2 & -2ab - 2cd & -2ac + 2bd \\ 2ab - 2cd & a^2 - b^2 + c^2 - d^2 & -2ad - 2bc \\ 2ac + 2bd & 2ad - 2bc & a^2 + b^2 - c^2 - d^2 \end{pmatrix}.$$

One can check that $\rho(A)$ is indeed a rotation matrix which represents the rotation whose axis is the line determined by the vector $(d, -c, b)$ and whose angle $\theta \in [-\pi, \pi]$ is determined by

$$\cos\frac{\theta}{2} = |a|.$$

We can also compute the derivative $d\rho_I\colon \mathfrak{su}(2) \to \mathfrak{so}(3)$ of ρ at I as follows. Recall that $\mathfrak{su}(2)$ consists of all complex matrices of the form

$$\begin{pmatrix} ib & c + id \\ -c + id & -ib \end{pmatrix}, \quad b, c, d \in \mathbb{R},$$

so pick the following basis for $\mathfrak{su}(2)$,

$$X_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

and define the curves in $\mathbf{SU}(2)$ through I given by

$$c_1(t) = \begin{pmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{pmatrix}, \quad c_2(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}, \quad c_3(t) = \begin{pmatrix} \cos t & i\sin t \\ i\sin t & \cos t \end{pmatrix}.$$

It is easy to check that $c_i'(0) = X_i$ for $i = 1, 2, 3$, and that

$$d\rho_I(X_1) = 2\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad d\rho_I(X_2) = 2\begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad d\rho_I(X_3) = 2\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Thus we have

$$d\rho_I(X_1) = 2E_3, \quad d\rho_I(X_2) = -2E_2, \quad d\rho_I(X_3) = 2E_1,$$

where (E_1, E_2, E_3) is the basis of $\mathfrak{so}(3)$ given by

$$\left(E_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \; E_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \; E_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right),$$

which means that $d\rho_I$ is an isomorphism between the Lie algebras $\mathfrak{su}(2)$ and $\mathfrak{so}(3)$.

Recall that we have the commutative diagram

$$\begin{tikzcd} \mathbf{SU}(2) \arrow[r, "\rho"] & \mathbf{SO}(3) \\ \mathfrak{su}(2) \arrow[u, "\exp"] \arrow[r, "d\rho_I"'] & \mathfrak{so}(3) \arrow[u, "\exp"'] \end{tikzcd}.$$

Since $d\rho_I$ is surjective and the exponential map $\exp\colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ is surjective, we conclude that ρ is surjective. (We also know that $\exp\colon \mathfrak{su}(2) \to \mathbf{SU}(2)$ is surjective.) Observe that $\rho(-A) = \rho(A)$, and it is easy to check that $\operatorname{Ker}\rho = \{I, -I\}$.

Example C.11. The set of $n \times n$ symmetric, positive, definite matrices, $\mathbf{SPD}(n)$.

Let $X = \mathbf{SPD}(n)$ and $G = \mathbf{GL}(n)$. The group $\mathbf{GL}(n) = \mathbf{GL}(n, \mathbb{R})$ acts on $\mathbf{SPD}(n)$ as follows: for all $A \in \mathbf{GL}(n)$ and all $S \in \mathbf{SPD}(n)$,

$$A \cdot S = ASA^\top.$$

It is easily checked that $ASA^\top$ is in $\mathbf{SPD}(n)$ if S is in $\mathbf{SPD}(n)$. First observe that $ASA^\top$ is symmetric since

$$(ASA^\top)^\top = AS^\top A^\top = ASA^\top.$$

Next recall the following characterization of positive definite matrix, namely

$$y^\top S y > 0, \qquad \text{whenever } y \neq 0.$$

We want to show $x^\top (A^\top S A) x > 0$ for all $x \neq 0$. Since A is invertible, we have $x = A^{-1} y$ for some nonzero y, and hence

$$\begin{aligned} x^\top (A^\top S A) x &= y^\top (A^{-1})^\top A^\top S A A^{-1} y \\ &= y^\top S y > 0. \end{aligned}$$

Hence $A^\top S A$ is positive definite. This action is transitive because every SPD matrix S can be written as $S = AA^\top$, for some invertible matrix A (prove this as an exercise). Given any two SPD matrices $S_1 = A_1 A_1^\top$ and $S_2 = A_2 A_2^\top$ with A_1 and A_2 invertible, if $A = A_2 A_1^{-1}$, we have

$$\begin{aligned} A \cdot S_1 &= A_2 A_1^{-1} S_1 (A_2 A_1^{-1})^\top = A_2 A_1^{-1} S_1 (A_1^\top)^{-1} A_2^\top \\ &= A_2 A_1^{-1} A_1 A_1^\top (A_1^\top)^{-1} A_2^\top = A_2 A_2^\top = S_2. \end{aligned}$$

Example C.12. The projective spaces $\mathbb{RP}^n$ and $\mathbb{CP}^n$.

The *(real) projective space* $\mathbb{RP}^n$ is the set of all lines through the origin in $\mathbb{R}^{n+1}$; that is, the set of one-dimensional subspaces of $\mathbb{R}^{n+1}$ (where $n \geq 0$). Since a one-dimensional subspace $L \subseteq \mathbb{R}^{n+1}$ is spanned by any nonzero vector $u \in L$, we can view $\mathbb{RP}^n$ as the set of equivalence classes of nonzero vectors in $\mathbb{R}^{n+1} - \{0\}$ modulo the equivalence relation

$$u \sim v \quad \text{iff} \quad v = \lambda u, \quad \text{for some} \quad \lambda \in \mathbb{R},\ \lambda \neq 0.$$

In terms of this definition, there is a projection $pr\colon (\mathbb{R}^{n+1} - \{0\}) \to \mathbb{RP}^n$, given by $pr(u) = [u]_\sim$, the equivalence class of u modulo $\sim$. Write $[u]$ for the line defined by the nonzero vector u. Since every line L in $\mathbb{R}^{n+1}$ intersects the sphere S^n in two antipodal points, we can view $\mathbb{RP}^n$ as the quotient of the sphere S^n by identification of antipodal points. See Figures C.3 and C.4.

Let $X = \mathbb{RP}^n$ and $G = \mathbf{SO}(n+1)$. We define an action of $\mathbf{SO}(n+1)$ on $\mathbb{RP}^n$ as follows: For any line $L = [u]$, for any $R \in \mathbf{SO}(n+1)$,

$$R \cdot L = [Ru].$$

Since R is linear, the line $[Ru]$ is well defined; that is, does not depend on the choice of $u \in L$. The reader can show that this action is transitive.

The *(complex) projective space* $\mathbb{CP}^n$ is defined analogously as the set of all lines through the origin in $\mathbb{C}^{n+1}$; that is, the set of one-dimensional subspaces of $\mathbb{C}^{n+1}$ (where $n \geq 0$). This time, we can view $\mathbb{CP}^n$ as the set of equivalence classes of vectors in $\mathbb{C}^{n+1} - \{0\}$ modulo the equivalence relation

$$u \sim v \quad \text{iff} \quad v = \lambda u, \quad \text{for some} \quad \lambda \neq 0 \in \mathbb{C}.$$

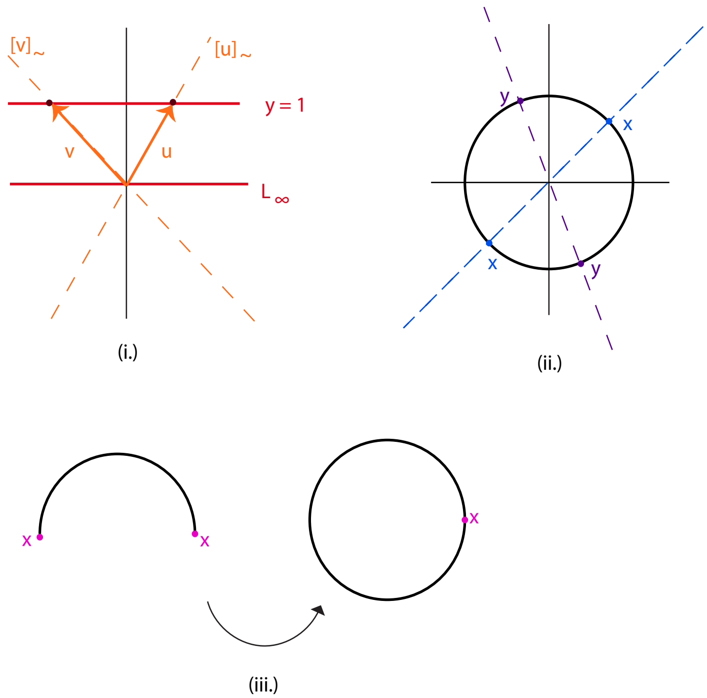


Figure C.3: Three constructions for $\mathbb{RP}^1 \cong S^1$. Illustration (i.) applies the equivalence relation. Since any line through the origin, excluding the x-axis, intersects the line $y = 1$, its equivalence class is represented by its point of intersection on $y = 1$. Hence, $\mathbb{RP}^n$ is the disjoint union of the line $y = 1$ and the point of infinity given by the x-axis. Illustration (ii.) represents $\mathbb{RP}^1$ as the quotient of the circle S^1 by identification of antipodal points. Illustration (iii.) is a variation which glues the equatorial points of the upper semicircle.

We have the projection $pr\colon \mathbb{C}^{n+1} - \{0\} \to \mathbb{CP}^n$, given by $pr(u) = [u]_\sim$, the equivalence class of u modulo $\sim$. Again, write $[u]$ for the line defined by the nonzero vector u. Let $X = \mathbb{CP}^n$ and $G = \mathbf{SU}(n+1)$. We define an action of $\mathbf{SU}(n+1)$ on $\mathbb{CP}^n$ as follows: For any line $L = [u]$, for any $R \in \mathbf{SU}(n+1)$,

$$R \cdot L = [Ru].$$

Again, this action is well defined and it is transitive. (Check this.)

Before progressing to our final example of group actions, we take a moment to construct $\mathbb{CP}^n$ as a quotient space of S^{2n+1}. Recall that $\Sigma^n \subseteq \mathbb{C}^{n+1}$, the unit sphere in $\mathbb{C}^{n+1}$, is defined by

$$\Sigma^n = \{(z_1, \ldots, z_{n+1}) \in \mathbb{C}^{n+1} \mid z_1\overline{z}_1 + \cdots + z_{n+1}\overline{z}_{n+1} = 1\}.$$

For any line $L = [u]$, where $u \in \mathbb{C}^{n+1}$ is a nonzero vector, writing $u = (u_1, \ldots, u_{n+1})$, a point $z \in \mathbb{C}^{n+1}$ belongs to L iff $z = \lambda(u_1, \ldots, u_{n+1})$, for some $\lambda \in \mathbb{C}$. Therefore, the intersection

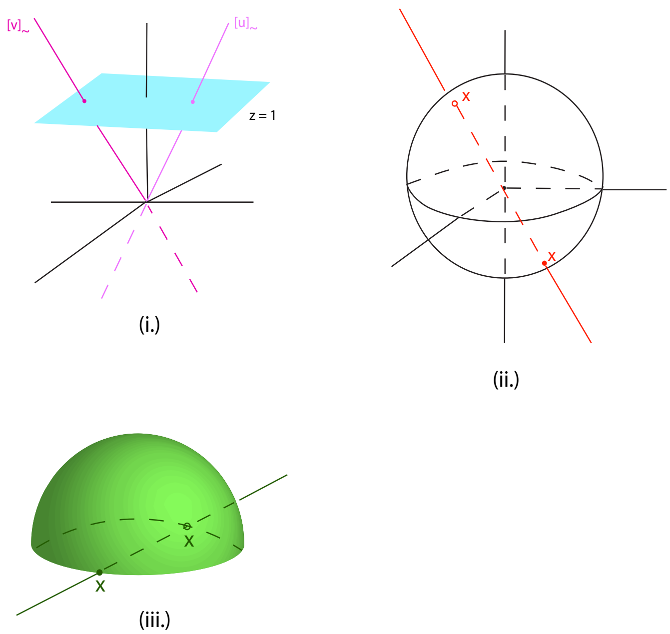


Figure C.4: Three constructions for $\mathbb{RP}^2$. Illustration (i.) applies the equivalence relation. Since any line through the origin which is not contained in the xy-plane intersects the plane $z = 1$, its equivalence class is represented by its point of intersection on $z = 1$. Hence, $\mathbb{RP}^2$ is the disjoint union of the plane $z = 1$ and the copy of $\mathbb{RP}^1$ provided by the xy-plane. Illustration (ii.) represents $\mathbb{RP}^2$ as the quotient of the sphere S^2 by identification of antipodal points. Illustration (iii.) is a variation which glues the antipodal points on boundary of the unit disk, which is represented as as the upper hemisphere.

$L \cap \Sigma^n$ of the line L and the sphere Σ^n is given by

$$L \cap \Sigma^n = \{\lambda(u_1, \ldots, u_{n+1}) \in \mathbb{C}^{n+1} \mid \lambda \in \mathbb{C},\ \lambda\overline{\lambda}(u_1\overline{u}_1 + \cdots + u_{n+1}\overline{u}_{n+1}) = 1\},$$

i.e.,

$$L \cap \Sigma^n = \left\{ \lambda(u_1, \ldots, u_{n+1}) \in \mathbb{C}^{n+1} \;\middle|\; \lambda \in \mathbb{C},\ |\lambda| = \frac{1}{\sqrt{|u_1|^2 + \cdots + |u_{n+1}|^2}} \right\}.$$

Thus, we see that there is a bijection between $L \cap \Sigma^n$ and the circle S^1; that is, geometrically $L \cap \Sigma^n$ is a circle. Moreover, since any line L through the origin is determined by just one other point, we see that for any two lines L_1 and L_2 through the origin,

$$L_1 \neq L_2 \quad \text{iff} \quad (L_1 \cap \Sigma^n) \cap (L_2 \cap \Sigma^n) = \emptyset.$$

However, Σ^n is the sphere S^{2n+1} in $\mathbb{R}^{2n+2}$. It follows that $\mathbb{CP}^n$ is the quotient of S^{2n+1} by the equivalence relation $\sim$ defined such that

$$y \sim z \quad \text{iff} \quad y, z \in L \cap \Sigma^n, \quad \text{for some line, } L \text{, through the origin.}$$

Therefore, we can write

$$S^{2n+1}/S^1 \cong \mathbb{CP}^n.$$

The case $n = 1$ is particularly interesting, as it turns out that

$$S^3/S^1 \cong S^2.$$

This is the famous *Hopf fibration*. To show this, proceed as follows: As

$$S^3 \cong \Sigma^1 = \{(z, z') \in \mathbb{C}^2 \mid |z|^2 + |z'|^2 = 1\},$$

define a map, $\mathrm{HF}\colon S^3 \to S^2$, by

$$\mathrm{HF}((z, z')) = (2z\overline{z'}, |z|^2 - |z'|^2).$$

We leave as a homework exercise to prove that this map has range S^2 and that

$$\mathrm{HF}((z_1, z_1')) = \mathrm{HF}((z_2, z_2')) \quad \text{iff} \quad (z_1, z_1') = \lambda(z_2, z_2'), \quad \text{for some } \lambda \text{ with } |\lambda| = 1.$$

In other words, for any point, $p \in S^2$, the inverse image $\mathrm{HF}^{-1}(p)$ (also called *fibre* over p) is a circle on S^3. Consequently, S^3 can be viewed as the union of a family of disjoint circles. This is the *Hopf fibration*. It is possible to visualize the Hopf fibration using the stereographic projection from S^3 onto $\mathbb{R}^3$. This is a beautiful and puzzling picture. For example, see Berger [4]. Therefore, HF induces a bijection from $\mathbb{CP}^1$ to S^2, and it is a homeomorphism.

Example C.13. Affine spaces.

Let X be a set and E a real vector space. A transitive and faithful action $\cdot\colon E \times X \to X$ of the additive group of E on X makes X into an *affine space*. The intuition is that the members of E are translations.

Those familiar with affine spaces as in Gallier [37] (Chapter 2) or Berger [4] will point out that if X is an affine space, then not only is the action of E on X transitive, but more is true: For any two points $a, b \in E$, there is a *unique* vector $u \in E$, such that $u \cdot a = b$. By the way, the action of E on X is usually considered to be a right action and is written additively, so $u \cdot a$ is written $a + u$ (the result of translating a by u). Thus, it would seem that we have to require more of our action. However, this is not necessary because E (under addition) is *abelian*. More precisely, we have the proposition

Proposition C.13. *If G is an abelian group acting on a set X and the action $\cdot\colon G \times X \to X$ is transitive and faithful, then for any two elements $x, y \in X$, there is a unique $g \in G$ so that $g \cdot x = y$ (the action is simply transitive).*

Proof. Since our action is transitive, there is at least some $g \in G$ so that $g \cdot x = y$. Assume that we have $g_1, g_2 \in G$ with

$$g_1 \cdot x = g_2 \cdot x = y.$$

We shall prove that

$$g_1 \cdot z = g_2 \cdot z, \quad \text{for all } z \in X.$$

This implies that

$$g_1 g_2^{-1} \cdot z = z, \quad \text{for all } z \in X.$$

As our action is faithful, $g_1 g_2^{-1} = 1$, and we must have $g_1 = g_2$, which proves our proposition.

Pick any $z \in X$. As our action is transitive, there is some $h \in G$ so that $z = h \cdot x$. Then, we have

$$\begin{aligned}
g_1 \cdot z &= g_1 \cdot (h \cdot x) \\
&= (g_1 h) \cdot x \\
&= (h g_1) \cdot x \qquad \text{(since } G \text{ is abelian)} \\
&= h \cdot (g_1 \cdot x) \\
&= h \cdot (g_2 \cdot x) \qquad \text{(since} \quad g_1 \cdot x = g_2 \cdot x) \\
&= (h g_2) \cdot x \\
&= (g_2 h) \cdot x \qquad \text{(since } G \text{ is abelian)} \\
&= g_2 \cdot (h \cdot x) \\
&= g_2 \cdot z.
\end{aligned}$$

Therefore, $g_1 \cdot z = g_2 \cdot z$ for all $z \in X$, as claimed. ☐

C.3 Group Actions: Part II, Stabilizers and Homogeneous Spaces

Now that we have an understanding of how a group G acts on a set X, we may use this action to form new topological spaces, namely homogeneous spaces. In the construction of homogeneous spaces, the subset of group elements that leaves some given element $x \in X$ fixed plays an important role.

Definition C.15. Given an action $\cdot\colon G \times X \to X$ of a group G on a set X, for any $x \in X$, the group G_x (also denoted $\mathrm{Stab}_G(x)$), called the *stabilizer* of x or *isotropy group at* x, is given by

$$G_x = \{g \in G \mid g \cdot x = x\}.$$

We have to verify that G_x is indeed a subgroup of G, but this is easy. Indeed, if $g \cdot x = x$ and $h \cdot x = x$, then we also have $h^{-1} \cdot x = x$ and so, we get $gh^{-1} \cdot x = x$, proving that G_x is a subgroup of G. In general, G_x is **not** a normal subgroup.

Observe that

$$G_{g\cdot x} = gG_xg^{-1},$$

for all $g \in G$ and all $x \in X$. Indeed,

$$\begin{aligned} G_{g\cdot x} &= \{h \in G \mid h \cdot (g \cdot x) = g \cdot x\} \\ &= \{h \in G \mid hg \cdot x = g \cdot x\} \\ &= \{h \in G \mid g^{-1}hg \cdot x = x\}, \end{aligned}$$

which shows $g^{-1}G_{g\cdot x}g \subseteq G_x$, or equivalently that $G_{g\cdot x} \subseteq gG_xg^{-1}$. It remains to show that $gG_xg^{-1} \subseteq G_{g\cdot x}$. Take an element of gG_xg^{-1}, which has the form ghg^{-1} with $h \cdot x = x$. Since $h \cdot x = x$, we have $(ghg^{-1}) \cdot gx = gx$, which shows that $ghg^{-1} \in G_{g\cdot x}$.

Because $G_{g\cdot x} = gG_xg^{-1}$, the stabilizers of x and $g \cdot x$ are conjugate of each other.

When the action of G on X is transitive, for any fixed $x \in G$, the set X is a quotient (as a set, not as group) of G by G_x. Indeed, we can define the map, $\pi_x \colon G \to X$, by

$$\pi_x(g) = g \cdot x, \quad \text{for all } g \in G.$$

Observe that

$$\pi_x(gG_x) = (gG_x) \cdot x = g \cdot (G_x \cdot x) = g \cdot x = \pi_x(g).$$

This shows that $\pi_x \colon G \to X$ induces a quotient map $\overline{\pi}_x \colon G/G_x \to X$, from the set G/G_x of (left) cosets of G_x to X, defined by

$$\overline{\pi}_x(gG_x) = g \cdot x.$$

Since

$$\pi_x(g) = \pi_x(h) \quad \text{iff} \quad g \cdot x = h \cdot x \quad \text{iff} \quad g^{-1}h \cdot x = x \quad \text{iff} \quad g^{-1}h \in G_x \quad \text{iff} \quad gG_x = hG_x,$$

we deduce that $\overline{\pi}_x \colon G/G_x \to X$ is injective. However, since our action is transitive, for every $y \in X$, there is some $g \in G$ so that $g \cdot x = y$, and so $\overline{\pi}_x(gG_x) = g \cdot x = y$; that is, the map $\overline{\pi}_x$ is also surjective. Therefore, the map $\overline{\pi}_x \colon G/G_x \to X$ is a bijection (of sets, not groups). The map $\pi_x \colon G \to X$ is also surjective. Let us record this important fact as

Proposition C.14. *If $\cdot \colon G \times X \to X$ is a transitive action of a group G on a set X, for every fixed $x \in X$, the surjection $\pi_x \colon G \to X$ given by*

$$\pi_x(g) = g \cdot x$$

induces a bijection

$$\overline{\pi}_x \colon G/G_x \to X,$$

where G_x is the stabilizer of x. See Figure C.5.

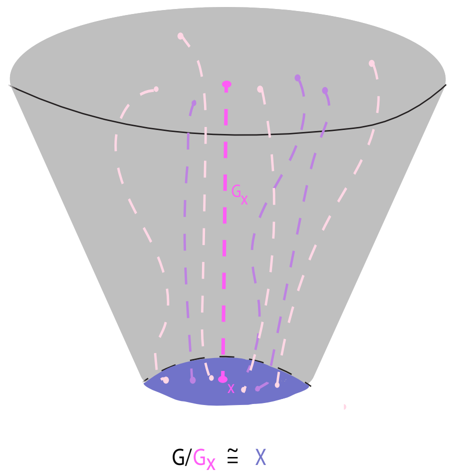


Figure C.5: A schematic representation of $G/G_x \cong X$, where G is the gray solid, X is its purple circular base, and G_x is the pink vertical strand. The dotted strands are the fibres gG_x.

The map $\pi_x \colon G \to X$ (corresponding to a fixed $x \in X$) is sometimes called a *projection* of G onto X. Proposition C.14 shows that for every $y \in X$, the subset $\pi_x^{-1}(y)$ of G (called the *fibre above* y) is equal to some coset gG_x of G, and thus is in bijection with the group G_x itself. We can think of G as a moving family of fibres G_x parametrized by X. This point of view of viewing a space as a moving family of simpler spaces is typical in (algebraic) geometry, and underlies the notion of (principal) fibre bundle.

Note that if the action $\cdot \colon G \times X \to X$ is transitive, then the stabilizers G_x and G_y of any two elements $x, y \in X$ are isomorphic, as they as conjugates. Thus, in this case, it is enough to compute one of these stabilizers for a "convenient" x.

As the situation of Proposition C.14 is of particular interest, we make the following definition:

Definition C.16. A set X is said to be a *homogeneous space* if there is a transitive action $\cdot \colon G \times X \to X$ of some group G on X.

We see that all the spaces of Examples C.6–C.13, are homogeneous spaces. Another example that will play an important role when we deal with Lie groups is the situation where we have a group G, a subgroup H of G (not necessarily normal), and where $X = G/H$, the set of left cosets of G modulo H. The group G acts on G/H by left multiplication:

$$a \cdot (gH) = (ag)H,$$

where $a, g \in G$. This action is clearly transitive and one checks that the stabilizer of gH is gHg^{-1}. If G is a topological group and H is a closed subgroup of G (see later for an explanation), it turns out that G/H is Hausdorff. If G is a Lie group, we obtain a manifold.

Even if G and X are topological spaces and the action $\cdot\colon G \times X \to X$ is continuous, in general, the space G/G_x under the quotient topology is **not** homeomorphic to X.

We will give later sufficient conditions that insure that X is indeed a topological space or even a manifold. In particular, X will be a manifold when G is a Lie group.

In general, an action $\cdot\colon G \times X \to X$ is not transitive on X, but for every $x \in X$, it is transitive on the set

$$O(x) = G \cdot x = \{g \cdot x \mid g \in G\}.$$

Such a set is called the *orbit* of x. The orbits are the equivalence classes of the following equivalence relation:

Definition C.17. Given an action $\cdot\colon G \times X \to X$ of some group G on X, the equivalence relation $\sim$ on X is defined so that, for all $x, y \in X$,

$$x \sim y \quad \text{iff} \quad y = g \cdot x, \quad \text{for some } g \in G.$$

For every $x \in X$, the equivalence class of x is the *orbit of* x, denoted $O(x)$ or $G \cdot x$, with

$$G \cdot x = O(x) = \{g \cdot x \mid g \in G\}.$$

The set of orbits is denoted X/G.

We warn the reader that some authors use the notation $G\backslash X$ for the the set of orbits $G \cdot x$, by analogy with right cosets Hg of a subgroup H of G.

The orbit space X/G is obtained from X by an identification (or merging) process: For every orbit, all points in that orbit are merged into a single point. This is akin to the process of forming the identification topology. For example, if $X = S^2$ and G is the group consisting of the restrictions of the two linear maps I and $-I$ of $\mathbb{R}^3$ to S^2 (where $(-I)(x) = -x$ for all $x \in \mathbb{R}^3$), then

$$X/G = S^2/\{I, -I\} \cong \mathbb{RP}^2.$$

See Figure C.4. More generally, if S^n is the n-sphere in $\mathbb{R}^{n+1}$, then we have a bijection between the orbit space $S^n/\{I, -I\}$ and $\mathbb{RP}^n$:

$$S^n/\{I, -I\} \cong \mathbb{RP}^n.$$

Many manifolds can be obtained in this fashion, including the torus, the Klein bottle, the Möbius band, *etc.*

Since the action of G is transitive on $O(x)$, by Proposition C.14, we see that for every $x \in X$, we have a bijection

$$O(x) \cong G/G_x.$$

As a corollary, if both X and G are finite, for any set $A \subseteq X$ of representatives from every orbit, we have the *orbit formula*:

$$|X| = \sum_{a \in A} [G \colon G_a] = \sum_{a \in A} |G|/|G_a|.$$

Even if a group action $\cdot\colon G \times X \to X$ is not transitive, when X is a manifold, we can consider the set of orbits X/G, and if the action of G on X satisfies certain conditions, X/G is actually a manifold. Manifolds arising in this fashion are often called *orbifolds*. In summary, we see that manifolds arise in at least two ways from a group action:

(1) As homogeneous spaces G/G_x, if the action is transitive.

(2) As orbifolds X/G (under certain conditions on the action).

Of course, in both cases, the action must satisfy some additional properties.

For the rest of this section, we reconsider Examples C.6–C.13 in the context of homogeneous space by determining some stabilizers for those actions.

(a) Consider the action $\cdot\colon \mathbf{SO}(n) \times S^{n-1} \to S^{n-1}$ of $\mathbf{SO}(n)$ on the sphere S^{n-1} $(n \geq 1)$ defined in Example C.6. Since this action is transitive, we can determine the stabilizer of any convenient element of S^{n-1}, say $e_1 = (1, 0, \ldots, 0)$. In order for any $R \in \mathbf{SO}(n)$ to leave e_1 fixed, the first column of R must be e_1, so R is an orthogonal matrix of the form

$$R = \begin{pmatrix} 1 & U \\ 0 & S \end{pmatrix}, \quad \text{with} \quad \det(S) = 1,$$

where U is a $1 \times (n-1)$ row vector. As the rows of R must be unit vectors, we see that $U = 0$ and $S \in \mathbf{SO}(n-1)$. Therefore, the stabilizer of e_1 is isomorphic to $\mathbf{SO}(n-1)$, and we deduce the bijection

$$\mathbf{SO}(n)/\mathbf{SO}(n-1) \cong S^{n-1}.$$

Strictly speaking, $\mathbf{SO}(n-1)$ is not a subgroup of $\mathbf{SO}(n)$, and in all rigor, we should consider the subgroup $\widetilde{\mathbf{SO}}(n-1)$ of $\mathbf{SO}(n)$ consisting of all matrices of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & S \end{pmatrix}, \quad \text{with} \quad \det(S) = 1,$$

and write

$$\mathbf{SO}(n)/\widetilde{\mathbf{SO}}(n-1) \cong S^{n-1}.$$

However, it is common practice to identify $\mathbf{SO}(n-1)$ with $\widetilde{\mathbf{SO}}(n-1)$.

When $n = 2$, as $\mathbf{SO}(1) = \{1\}$, we find that $\mathbf{SO}(2) \cong S^1$, a circle, a fact that we already knew. When $n = 3$, we find that $\mathbf{SO}(3)/\mathbf{SO}(2) \cong S^2$. This says that $\mathbf{SO}(3)$ is somehow the result of glueing circles to the surface of a sphere (in $\mathbb{R}^3$), in such a way that these circles do not intersect. This is hard to visualize!

A similar argument for the complex unit sphere Σ^{n-1} shows that

$$\mathbf{SU}(n)/\mathbf{SU}(n-1) \cong \Sigma^{n-1} \cong S^{2n-1}.$$

Again, we identify $\mathbf{SU}(n-1)$ with a subgroup of $\mathbf{SU}(n)$, as in the real case. In particular, when $n = 2$, as $\mathbf{SU}(1) = \{1\}$, we find that

$$\mathbf{SU}(2) \cong S^3;$$

that is, the group $\mathbf{SU}(2)$ is topologically the sphere S^3! Actually, this is not surprising if we remember that $\mathbf{SU}(2)$ is in fact the group of unit quaternions.

(b) We saw in Example C.7 that the action $\cdot\colon \mathbf{SL}(2,\mathbb{R}) \times H \to H$ of the group $\mathbf{SL}(2,\mathbb{R})$ on the upper half plane is transitive. Let us find out what the stabilizer of $z = i$ is. We should have

$$\frac{ai+b}{ci+d} = i,$$

that is, $ai + b = -c + di$, i.e.,

$$(d-a)i = b + c.$$

Since a, b, c, d are real, we must have $d = a$ and $b = -c$. Moreover, $ad - bc = 1$, so we get $a^2 + b^2 = 1$. We conclude that a matrix in $\mathbf{SL}(2,\mathbb{R})$ fixes i iff it is of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad \text{with} \quad a^2 + b^2 = 1.$$

Clearly, these are the rotation matrices in $\mathbf{SO}(2)$, and so the stabilizer of i is $\mathbf{SO}(2)$. We conclude that

$$\mathbf{SL}(2,\mathbb{R})/\mathbf{SO}(2) \cong H.$$

This time we can view $\mathbf{SL}(2,\mathbb{R})$ as the result of glueing circles to the upper half plane. This is not so easy to visualize. There is a better way to visualize the topology of $\mathbf{SL}(2,\mathbb{R})$ by making it act on the open disk D. We will return to this action in a little while.

(c) Now consider the action of $\mathbf{SL}(2,\mathbb{C})$ on $\mathbb{C} \cup \{\infty\} \cong S^2$ given in Example C.8. As it is transitive, let us find the stabilizer of $z = 0$. We must have

$$\frac{b}{d} = 0,$$

and as $ad - bc = 1$, we must have $b = 0$ and $ad = 1$. Thus the stabilizer of 0 is the subgroup $\mathbf{SL}(2,\mathbb{C})_0$ of $\mathbf{SL}(2,\mathbb{C})$ consisting of all matrices of the form

$$\begin{pmatrix} a & 0 \\ c & a^{-1} \end{pmatrix}, \quad \text{where} \quad a \in \mathbb{C} - \{0\} \quad \text{and} \quad c \in \mathbb{C}.$$

We get
$$\mathbf{SL}(2,\mathbb{C})/\mathbf{SL}(2,\mathbb{C})_0 \cong \mathbb{C}\cup\{\infty\} \cong S^2,$$
but this is not very illuminating.

(d) In Example C.11 we considered the action $\cdot\colon \mathbf{GL}(n)\times\mathbf{SPD}(n)\to\mathbf{SPD}(n)$ of $\mathbf{GL}(n)$ on $\mathbf{SPD}(n)$, the set of symmetric positive definite matrices. As this action is transitive, let us find the stabilizer of I. For any $A\in\mathbf{GL}(n)$, the matrix A stabilizes I iff
$$AIA^\top = AA^\top = I.$$
Therefore the stabilizer of I is $\mathbf{O}(n)$, and we find that
$$\mathbf{GL}(n)/\mathbf{O}(n) = \mathbf{SPD}(n).$$

Observe that if $\mathbf{GL}^+(n)$ denotes the subgroup of $\mathbf{GL}(n)$ consisting of all matrices with a strictly positive determinant, then we have an action $\cdot\colon \mathbf{GL}^+(n)\times\mathbf{SPD}(n)\to\mathbf{SPD}(n)$ of $\mathbf{GL}^+(n)$ on $\mathbf{SPD}(n)$. This action is transitive and we find that the stabilizer of I is $\mathbf{SO}(n)$; consequently, we get
$$\mathbf{GL}^+(n)/\mathbf{SO}(n) = \mathbf{SPD}(n).$$

(e) In Example C.12 we considered the action $\cdot\colon \mathbf{SO}(n+1)\times\mathbb{RP}^n\to\mathbb{RP}^n$ of $\mathbf{SO}(n+1)$ on the (real) projective space $\mathbb{RP}^n$. As this action is transitive, let us find the stabilizer of the line $L=[e_1]$, where $e_1=(1,0,\ldots,0)$. For any $R\in\mathbf{SO}(n+1)$, the line L is fixed iff either $R(e_1)=e_1$ or $R(e_1)=-e_1$, since e_1 and $-e_1$ define the same line. As R is orthogonal with $\det(R)=1$, this means that R is of the form
$$R=\begin{pmatrix}\alpha & 0\\ 0 & S\end{pmatrix},\quad \text{with}\quad \alpha=\pm 1\quad\text{and}\quad \det(S)=\alpha.$$
But, S must be orthogonal, so we conclude $S\in\mathbf{O}(n)$. Therefore, the stabilizer of $L=[e_1]$ is isomorphic to the group $\mathbf{O}(n)$, and we find that
$$\mathbf{SO}(n+1)/\mathbf{O}(n)\cong\mathbb{RP}^n.$$

Strictly speaking, $\mathbf{O}(n)$ is not a subgroup of $\mathbf{SO}(n+1)$, so the above equation does not make sense. We should write
$$\mathbf{SO}(n+1)/\widetilde{\mathbf{O}}(n)\cong\mathbb{RP}^n,$$
where $\widetilde{\mathbf{O}}(n)$ is the subgroup of $\mathbf{SO}(n+1)$ consisting of all matrices of the form
$$\begin{pmatrix}\alpha & 0\\ 0 & S\end{pmatrix},\quad \text{with}\quad S\in\mathbf{O}(n),\ \alpha=\pm 1\quad\text{and}\quad \det(S)=\alpha.$$
This group is also denoted $S(\mathbf{O}(1)\times\mathbf{O}(n))$. However, the common practice is to write $\mathbf{O}(n)$ instead of $S(\mathbf{O}(1)\times\mathbf{O}(n))$.

We should mention that $\mathbb{RP}^3$ and $\mathbf{SO}(3)$ are homeomorphic spaces. This is shown using the quaternions; for example, see Gallier [37], Chapter 8.

A similar argument applies to the action $\cdot\colon \mathbf{SU}(n+1)\times\mathbb{CP}^n \to \mathbb{CP}^n$ of $\mathbf{SU}(n+1)$ on the (complex) projective space $\mathbb{CP}^n$. We find that

$$\mathbf{SU}(n+1)/\mathbf{U}(n) \cong \mathbb{CP}^n.$$

Again, the above is a bit sloppy as $\mathbf{U}(n)$ is not a subgroup of $\mathbf{SU}(n+1)$. To be rigorous, we should use the subgroup $\widetilde{\mathbf{U}}(n)$ consisting of all matrices of the form

$$\begin{pmatrix} \alpha & 0 \\ 0 & S \end{pmatrix}, \quad \text{with} \quad S \in \mathbf{U}(n),\ |\alpha| = 1 \quad \text{and} \quad \det(S) = \overline{\alpha}.$$

This group is also denoted $S(\mathbf{U}(1)\times\mathbf{U}(n))$. The common practice is to write $\mathbf{U}(n)$ instead of $S(\mathbf{U}(1)\times\mathbf{U}(n))$. In particular, when $n = 1$, we find that

$$\mathbf{SU}(2)/\mathbf{U}(1) \cong \mathbb{CP}^1.$$

But, we know that $\mathbf{SU}(2) \cong S^3$, and clearly $\mathbf{U}(1) \cong S^1$. So, again, we find that $S^3/S^1 \cong \mathbb{CP}^1$ (we know more, namely, $S^3/S^1 \cong S^2 \cong \mathbb{CP}^1$.)

Observe that $\mathbb{CP}^n$ can also be viewed as the orbit space of the action $\cdot\colon S^1\times S^{2n+1} \to S^{2n+1}$ given by

$$\lambda\cdot(z_1,\ldots,z_{n+1}) = (\lambda z_1,\ldots,\lambda z_{n+1}),$$

where $S^1 = \mathbf{U}(1)$ (the group of complex numbers of modulus 1) and S^{2n+1} is identified with Σ^n.

We now return to Case (b) to give a better picture of $\mathbf{SL}(2,\mathbb{R})$. Instead of having $\mathbf{SL}(2,\mathbb{R})$ act on the upper half plane, we define an action of $\mathbf{SL}(2,\mathbb{R})$ on the open unit disk D as we did in Example C.9. Technically, it is easier to consider the group $\mathbf{SU}(1,1)$, which is isomorphic to $\mathbf{SL}(2,\mathbb{R})$, and to make $\mathbf{SU}(1,1)$ act on D. The group $\mathbf{SU}(1,1)$ is the group of 2×2 complex matrices of the form

$$\begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix}, \quad \text{with} \quad a\overline{a} - b\overline{b} = 1.$$

The reader should check that if we let

$$g = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix},$$

then the map from $\mathbf{SL}(2,\mathbb{R})$ to $\mathbf{SU}(1,1)$ given by

$$A \mapsto gAg^{-1}$$

is an isomorphism. Observe that the scaled Möbius transformation associated with g is

$$z \mapsto \frac{z-i}{z+i},$$

which is the holomorphic isomorphism mapping H to D mentionned earlier! We can define a bijection between $\mathbf{SU}(1,1)$ and $S^1 \times D$ given by

$$\begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix} \mapsto (a/|a|, b/a).$$

We conclude that $\mathbf{SL}(2,\mathbb{R}) \cong \mathbf{SU}(1,1)$ is topologically an open solid torus (i.e., with the surface of the torus removed). It is possible to further classify the elements of $\mathbf{SL}(2,\mathbb{R})$ into three categories and to have geometric interpretations of these as certain regions of the torus. For details, the reader should consult Carter, Segal and Macdonald [17] or Duistermatt and Kolk [29] (Chapter 1, Section 1.2).

The group $\mathbf{SU}(1,1)$ acts on D by interpreting any matrix in $\mathbf{SU}(1,1)$ as a Möbius tranformation; that is,

$$\begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix} \mapsto \left(z \mapsto \frac{az+b}{\overline{b}z+\overline{a}} \right).$$

The reader should check that these transformations preserve D.

Both the upper half-plane and the open disk are models of Lobachevsky's non-Euclidean geometry (where the parallel postulate fails). They are also models of hyperbolic spaces (Riemannian manifolds with constant negative curvature, see Gallot, Hulin and Lafontaine [41], Chapter III). According to Dubrovin, Fomenko, and Novikov [28] (Chapter 2, Section 13.2), the open disk model is due to Poincaré and the upper half-plane model to Klein, although Poincaré was the first to realize that the upper half-plane is a hyperbolic space.

C.4 The Grassmann and Stiefel Manifolds

In this section we introduce two very important homogeneous manifolds, the Grassmann manifolds and the Stiefel manifolds. The Grassmann manifolds are generalizations of projective spaces (real and complex), while the Stiefel manifold are generalizations of $\mathbf{O}(n)$. Both of these manifolds are examples of reductive homogeneous spaces. We begin by defining the Grassmann manifolds $G(k,n)$.

First consider the real case.

Definition C.18. Given any $n \geq 1$, for any k with $0 \leq k \leq n$, the set $G(k,n)$ of all linear k-dimensional subspaces of $\mathbb{R}^n$ (also called *k-planes*) is called a *Grassmannian* (or *Grassmann manifold*).

Any k-dimensional subspace U of $\mathbb{R}^n$ is spanned by k linearly independent vectors $u_1, \ldots, u_k$ in $\mathbb{R}^n$; write $U = \mathrm{span}(u_1, \ldots, u_k)$. We can define an action $\cdot \colon \mathbf{O}(n) \times G(k, n) \to G(k, n)$ as follows: For any $R \in \mathbf{O}(n)$, for any $U = \mathrm{span}(u_1, \ldots, u_k)$, let

$$R \cdot U = \mathrm{span}(Ru_1, \ldots, Ru_k).$$

We have to check that the above is well defined. If $U = \mathrm{span}(v_1, \ldots, v_k)$ for any other k linearly independent vectors $v_1, \ldots, v_k$, we have

$$v_i = \sum_{j=1}^{k} a_{ij} u_j, \quad 1 \leq i \leq k,$$

for some $a_{ij} \in \mathbb{R}$, and so

$$Rv_i = \sum_{j=1}^{k} a_{ij} Ru_j, \quad 1 \leq i \leq k,$$

which shows that

$$\mathrm{span}(Ru_1, \ldots, Ru_k) = \mathrm{span}(Rv_1, \ldots, Rv_k);$$

that is, the above action is well defined.

We claim this action is transitive. This is because if U and V are any two k-planes, we may assume that $U = \mathrm{span}(u_1, \ldots, u_k)$ and $V = \mathrm{span}(v_1, \ldots, v_k)$, where the u_i's form an orthonormal family and similarly for the v_i's. Then we can extend these families to orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ on $\mathbb{R}^n$, and w.r.t. the orthonormal basis $(u_1, \ldots, u_n)$, the matrix of the linear map sending u_i to v_i is orthogonal. Hence $G(k, n)$ is a homogeneous space.

In order to represent $G(k, n)$ as a quotient space, Proposition C.14 implies it is enough to find the stabilizer of any k-plane. Pick $U = \mathrm{span}(e_1, \ldots, e_k)$, where $(e_1, \ldots, e_n)$ is the canonical basis of $\mathbb{R}^n$ (i.e., $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, with the 1 in the ith position). Any $R \in \mathbf{O}(n)$ stabilizes U iff R maps $e_1, \ldots, e_k$ to k linearly independent vectors in the subspace $U = \mathrm{span}(e_1, \ldots, e_k)$, i.e., R is of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$

where S is $k \times k$ and T is $(n-k) \times (n-k)$. Moreover, as R is orthogonal, S and T must be orthogonal, that is $S \in \mathbf{O}(k)$ and $T \in \mathbf{O}(n-k)$. We deduce that the stabilizer of U is isomorphic to $\mathbf{O}(k) \times \mathbf{O}(n-k)$ and we find that

$$\mathbf{O}(n)/(\mathbf{O}(k) \times \mathbf{O}(n-k)) \cong G(k, n).$$

It turns out that this makes $G(k, n)$ into a smooth manifold of dimension

$$\frac{n(n-1)}{2} - \frac{k(k-1)}{2} - \frac{(n-k)(n-k-1)}{2} = k(n-k)$$

called a *Grassmannian*.

The restriction of the action of $\mathbf{O}(n)$ on $G(k,n)$ to $\mathbf{SO}(n)$ yields an action $\cdot\colon \mathbf{SO}(n) \times G(k,n) \to G(k,n)$ of $\mathbf{SO}(n)$ on $G(k,n)$. Then it is easy to see that this action is transitive and that the stabilizer of the subspace U is isomorphic to the subgroup $S(\mathbf{O}(k) \times \mathbf{O}(n-k))$ of $\mathbf{SO}(n)$ consisting of the rotations of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$

with $S \in \mathbf{O}(k)$, $T \in \mathbf{O}(n-k)$ and $\det(S)\det(T) = 1$. Thus, we also have

$$\mathbf{SO}(n)/S(\mathbf{O}(k) \times \mathbf{O}(n-k)) \cong G(k,n).$$

If we recall the projection map of Example C.12 in Section C.2, namely $pr\colon \mathbb{R}^{n+1} - \{0\} \to \mathbb{RP}^n$, by definition, a *$k$-plane* in $\mathbb{RP}^n$ is the image under pr of any $(k+1)$-plane in $\mathbb{R}^{n+1}$. So, for example, a line in $\mathbb{RP}^n$ is the image of a 2-plane in $\mathbb{R}^{n+1}$, and a hyperplane in $\mathbb{RP}^n$ is the image of a hyperplane in $\mathbb{R}^{n+1}$. The advantage of this point of view is that the k-planes in $\mathbb{RP}^n$ are arbitrary; that is, they do not have to go through "the origin" (which does not make sense, anyway!). Then we see that we can interpret the Grassmannian, $G(k+1, n+1)$, as a space of "parameters" for the k-planes in $\mathbb{RP}^n$. For example, $G(2, n+1)$ parametrizes the lines in $\mathbb{RP}^n$. In this viewpoint, $G(k+1, n+1)$ is usually denoted $\mathbb{G}(k,n)$.

It can be proved (using some exterior algebra) that $G(k,n)$ can be embedded in $\mathbb{RP}^{\binom{n}{k}-1}$. Much more is true. For example, $G(k,n)$ is a projective variety, which means that it can be defined as a subset of $\mathbb{RP}^{\binom{n}{k}-1}$ equal to the zero locus of a set of homogeneous equations. There is even a set of quadratic equations known as the *Plücker equations* defining $G(k,n)$. In particular, when $n = 4$ and $k = 2$, we have $G(2,4) \subseteq \mathbb{RP}^5$, and $G(2,4)$ is defined by a single equation of degree 2. The Grassmannian $G(2,4) = \mathbb{G}(1,3)$ is known as the *Klein quadric*. This hypersurface in $\mathbb{RP}^5$ parametrizes the lines in $\mathbb{RP}^3$.

Complex Grassmannians are defined in a similar way, by replacing $\mathbb{R}$ by $\mathbb{C}$ and $\mathbf{O}(n)$ by $\mathbf{U}(n)$ throughout. The complex Grassmannian $G_{\mathbb{C}}(k,n)$ is a complex manifold as well as a real manifold, and we have

$$\mathbf{U}(n)/(\mathbf{U}(k) \times \mathbf{U}(n-k)) \cong G_{\mathbb{C}}(k,n).$$

As in the case of the real Grassmannians, the action of $\mathbf{U}(n)$ on $G_{\mathbb{C}}(k,n)$ yields an action of $\mathbf{SU}(n)$ on $G_{\mathbb{C}}(k,n)$, and we get

$$\mathbf{SU}(n)/S(\mathbf{U}(k) \times \mathbf{U}(n-k)) \cong G_{\mathbb{C}}(k,n),$$

where $S(\mathbf{U}(k) \times \mathbf{U}(n-k))$ is the subgroup of $\mathbf{SU}(n)$ consisting of all matrices $R \in \mathbf{SU}(n)$ of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$

with $S \in \mathbf{U}(k)$, $T \in \mathbf{U}(n-k)$ and $\det(S)\det(T) = 1$.

Closely related to Grassmannians are the *Stiefel manifolds* $S(k, n)$. Again we begin with the real case.

Definition C.19. For any $n \geq 1$ and any k with $1 \leq k \leq n$, the set $S(k, n)$ of all orthonormal k-frames, that is, of k-tuples of orthonormal vectors $(u_1, \ldots, u_k)$ with $u_i \in \mathbb{R}^n$, is called a *Stiefel manifold*.

Obviously, $S(1, n) = S^{n-1}$ and $S(n, n) = \mathbf{O}(n)$, so assume $k \leq n - 1$. There is a natural action $\cdot\colon \mathbf{SO}(n) \times S(k, n) \to S(k, n)$ of $\mathbf{SO}(n)$ on $S(k, n)$ given by

$$R \cdot (u_1, \ldots, u_k) = (Ru_1, \ldots, Ru_k).$$

This action is transitive, because if $(u_1, \ldots, u_k)$ and $(v_1, \ldots, v_k)$ are any two orthonormal k-frames, then they can be extended to orthonormal bases (for example, by Gram-Schmidt) $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ with the same orientation (since we can pick u_n and v_n so that our bases have the same orientation), and there is a unique orthogonal transformation $R \in \mathbf{SO}(n)$ such that $Ru_i = v_i$ for $i = 1, \ldots, n$.

In order to apply Proposition C.14, we need to find the stabilizer of the orthonormal k-frame $(e_1, \ldots, e_k)$ consisting of the first canonical basis vectors of $\mathbb{R}^n$. A matrix $R \in \mathbf{SO}(n)$ stabilizes $(e_1, \ldots, e_k)$ iff it is of the form

$$R = \begin{pmatrix} I_k & 0 \\ 0 & S \end{pmatrix}$$

where $S \in \mathbf{SO}(n-k)$. Therefore, for $1 \leq k \leq n-1$, we have

$$\mathbf{SO}(n)/\mathbf{SO}(n-k) \cong S(k, n).$$

This makes $S(k, n)$ a smooth manifold of dimension

$$\frac{n(n-1)}{2} - \frac{(n-k)(n-k-1)}{2} = nk - \frac{k(k+1)}{2} = k(n-k) + \frac{k(k-1)}{2}.$$

Remark: It should be noted that we can define another type of Stiefel manifolds, denoted by $V(k, n)$, using linearly independent k-tuples $(u_1, \ldots, u_k)$ that do not necessarily form an orthonormal system. In this case, there is an action $\cdot\colon \mathbf{GL}(n, \mathbb{R}) \times V(k, n) \to V(k, n)$, and the stabilizer H of the first k canonical basis vectors $(e_1, \ldots, e_k)$ is a closed subgroup of $\mathbf{GL}(n, \mathbb{R})$, but it doesn't have a simple description (see Warner [102], Chapter 3). We get an isomorphism

$$V(k, n) \cong \mathrm{GL}(n, \mathbb{R})/H.$$

The version of the Stiefel manifold $S(k, n)$ using orthonormal frames is sometimes denoted by $V^0(k, n)$ (Milnor and Stasheff [72] use the notation $V_k^0(\mathbb{R}^n)$). Beware that the notation is not standardized. Certain authors use $V(k, n)$ for what we denote by $S(k, n)$!

Complex Stiefel manifolds are defined in a similar way by replacing $\mathbb{R}$ by $\mathbb{C}$ and $\mathbf{SO}(n)$ by $\mathbf{SU}(n)$. For $1 \leq k \leq n-1$, the complex Stiefel manifold $S_{\mathbb{C}}(k,n)$ is isomorphic to the quotient

$$\mathbf{SU}(n)/\mathbf{SU}(n-k) \cong S_{\mathbb{C}}(k,n).$$

If $k=1$, we have $S_{\mathbb{C}}(1,n) = S^{2n-1}$, and if $k=n$, we have $S_{\mathbb{C}}(n,n) = \mathbf{U}(n)$.

The Grassmannians can also be viewed as quotient spaces of the Stiefel manifolds. Every orthonomal k-frame $(u_1,\ldots,u_k)$ can be represented by an $n \times k$ matrix Y over the canonical basis of $\mathbb{R}^n$, and such a matrix Y satisfies the equation

$$Y^\top Y = I.$$

We have a right action $\cdot \colon S(k,n) \times \mathbf{O}(k) \to S(k,n)$ given by

$$Y \cdot R = YR,$$

for any $R \in \mathbf{O}(k)$. This action is well defined since

$$(YR)^\top YR = R^\top Y^\top Y R = I.$$

However, this action is not transitive (unless $k=1$), but the orbit space $S(k,n)/\mathbf{O}(k)$ is isomorphic to the Grassmannian $G(k,n)$, so we can write

$$G(k,n) \cong S(k,n)/\mathbf{O}(k).$$

Similarly, the complex Grassmannian is isomorphic to the orbit space $S_{\mathbb{C}}(k,n)/\mathbf{U}(k)$:

$$G_{\mathbb{C}}(k,n) \cong S_{\mathbb{C}}(k,n)/\mathbf{U}(k).$$

Appendix D

Hilbert Spaces

D.1 The Projection Lemma, Duality

If E is a complex vector space, recall that a map $\langle -, - \rangle \colon E \times E \to \mathbb{C}$ is a *hermitian form* if it satisfies the following properties for all $x, y, x_1, x_2, y_1, y_2 \in E$ and all $\lambda \in \mathbb{C}$: it is *sesquilinear*, which means that

$$\begin{aligned}
\langle x_1 + x_2, y \rangle &= \langle x_1, y \rangle + \langle x_2, y \rangle \\
\langle x, y_1 + y_2 \rangle &= \langle x, y_1 \rangle + \langle x, y_2 \rangle \\
\langle \lambda x, y \rangle &= \lambda \langle x, y \rangle \\
\langle x, \lambda y \rangle &= \overline{\lambda} \langle x, y \rangle,
\end{aligned}$$

and satisfies the *hermitian property*,

$$\langle y, x \rangle = \overline{\langle x, y \rangle}.$$

The hermitian property implies that $\langle x, x \rangle \in \mathbb{R}$ for all $x \in E$.

A hermitian form $\langle -, - \rangle \colon E \times E \to \mathbb{C}$ is *positive* if

$$\langle x, x \rangle \geq 0 \qquad \text{for all } x \in E.$$

A positive hermitian form satisfies the *Cauchy-Schwarz inequality*:

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle, \qquad \text{for all } x, y \in E.$$

A positive hermitian form is *positive definite* if for all $x \in E$,

$$\langle x, x \rangle = 0 \quad \text{implies that } x = 0,$$

or equivalently,

$$\langle x, x \rangle > 0 \qquad \text{for all } x \neq 0.$$

A positive definite hermitian form on E is often called a *hermitian inner product* on E, and E is called a *hermitian space* (sometimes a *pre-Hilbert space*).

Given a Hermitian space $\langle E, \langle -, - \rangle \rangle$, the function $\| \, \| \colon E \to \mathbb{R}$ defined such that $\|u\| = \sqrt{\langle u, u \rangle}$, is a norm on E. Thus, E is a normed vector space. If E is also complete, then it is a very interesting space.

In a hermitian space $\langle E, \langle -, - \rangle \rangle$, the inner product $\langle -, - \rangle$ can be recovered from the norm $\| \, \|$ using the following *polarization identities*: In the complex case,

$$\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2 + i \, \|x + iy\|^2 - i \, \|x - iy\|^2),$$

and in the real case,

$$\langle x, y \rangle = \frac{1}{2}(\|x + y\|^2 - \|x\|^2 - \|y\|^2).$$

Recall that completeness has to do with the convergence of Cauchy sequences. A normed vector space $\langle E, \| \, \| \rangle$ is automatically a metric space under the metric d defined such that $d(u, v) = \|v - u\|$ (see Chapter B for the definition of a normed vector space and of a metric space, or Lang [62, 63], or Dixmier [27]). Given a metric space E with metric d, a sequence $(a_n)_{n \geq 1}$ of elements $a_n \in E$ is a *Cauchy sequence* iff for every $\epsilon > 0$, there is some $N \geq 1$ such that

$$d(a_m, a_n) < \epsilon \quad \text{for all} \quad m, n \geq N.$$

We say that E is *complete* iff every Cauchy sequence converges to a limit (which is unique, since a metric space is Hausdorff).

Every finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ is complete. For example, one can show by induction that given any basis $(e_1, \ldots, e_n)$ of E, the linear map $h \colon \mathbb{C}^n \to E$ defined such that

$$h((z_1, \ldots, z_n)) = z_1 e_1 + \cdots + z_n e_n$$

is a homeomorphism (using the *sup*-norm on $\mathbb{C}^n$). One can also use the fact that any two norms on a finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ are equivalent (see Theorem B.3, or Lang [63], Dixmier [27], Schwartz [84]).

However, if E has infinite dimension, it may not be complete. When a Hermitian space is complete, a number of the properties that hold for finite dimensional Hermitian spaces also hold for infinite dimensional spaces. For example, any closed subspace has an orthogonal complement, and in particular, a finite dimensional subspace has an orthogonal complement. Hermitian spaces that are also complete play an important role in analysis. Since they were first studied by Hilbert, they are called Hilbert spaces.

Definition D.1. A (complex) Hermitian space $\langle E, \langle -, - \rangle \rangle$ which is a complete normed vector space under the norm $\| \, \|$ induced by $\langle -, - \rangle$ is called a *Hilbert space*. A real Euclidean space $\langle E, \langle -, - \rangle \rangle$ which is complete under the norm $\| \, \|$ induced by $\langle -, - \rangle$ is called a *real Hilbert space*.

All the results in this section hold for complex Hilbert spaces as well as for real Hilbert spaces. We state all results for the complex case only, since they also apply to the real case, and since the proofs in the complex case need a little more care.

Example D.1. The space ℓ^2 of all countably infinite sequences $x = (x_i)_{i\in\mathbb{N}}$ of complex numbers such that $\sum_{i=0}^{\infty} |x_i|^2 < \infty$ is a Hilbert space. It will be shown later that the map $\langle -, - \rangle \colon \ell^2 \times \ell^2 \to \mathbb{C}$ defined such that

$$\langle (x_i)_{i\in\mathbb{N}}, (y_i)_{i\in\mathbb{N}} \rangle = \sum_{i=0}^{\infty} x_i \overline{y_i}$$

is well defined, and that ℓ^2 is a Hilbert space under $\langle -, - \rangle$. In fact, we will prove a more general result (Proposition D.14).

Example D.2. The set $\mathcal{C}^{\infty}[a, b]$ of smooth functions $f\colon [a, b] \to \mathbb{C}$ is a Hermitian space under the Hermitian form

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx,$$

but it is not a Hilbert space because it is not complete. It is possible to construct its completion $\mathrm{L}^2([a, b])$, which turns out to be the space of Lebesgue square-integrable functions on $[a, b]$.

A simple adaptation of the completion theorem for normed vector spaces (Theorem A.72) shows that every hermitian space has a Hilbert space completion.

Theorem D.1. *If $(E, \langle -, - \rangle)$ is a hermitian space, then the Banach space $(\widehat{E}, \| \|_{\widehat{E}})$, completion of the normed vector space $(E, \| \|)$ where $\|x\| = \sqrt{\langle x, x \rangle}$ for all $x \in E$, is Hilbert space with the inner product $\langle -, - \rangle_h$ given by*

$$\langle x, y \rangle_h = \frac{1}{4}(\|x + y\|_{\widehat{E}}^2 - \|x - y\|_{\widehat{E}}^2 + i\,\|x + iy\|_{\widehat{E}}^2 - i\,\|x - iy\|_{\widehat{E}}^2),$$

and in the real case,

$$\langle x, y \rangle_h = \frac{1}{2}(\|x + y\|_{\widehat{E}}^2 - \|x\|_{\widehat{E}}^2 - \|y\|_{\widehat{E}}^2)$$

for all $x, y \in \widehat{E}$. Furthermore, the linear map $\varphi\colon E \to \widehat{E}$ given by Theorem A.72 is inner-product preserving.

Proof. Since E is dense in E, $E \times E$ is dense $\widehat{E} \times \widehat{E}$, and since the map

$$(x, y) \mapsto \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2 + i\,\|x + iy\|^2 - i\,\|x - iy\|^2),$$

and in the real case,

$$(x, y) = \frac{1}{2}(\|x + y\|^2 - \|x\|^2 - \|y\|^2),$$

is uniformly continuous, by Theorem A.61, these maps have unique continuous extensions, and if we define

$$\langle x, y\rangle_h = \frac{1}{4}(\|x+y\|_{\widehat{E}}^2 - \|x-y\|_{\widehat{E}}^2 + i\,\|x+iy\|_{\widehat{E}}^2 - i\,\|x-iy\|_{\widehat{E}}^2),$$

and in the real case,

$$\langle x, y\rangle_h \mapsto \frac{1}{2}(\|x+y\|_{\widehat{E}}^2 - \|x\|_{\widehat{E}}^2 - \|y\|_{\widehat{E}}^2),$$

for all $x, y \in \widehat{E}$, it is easy to check that we obtain positive definite hermitian forms with associated norm $\| \, \|_{\widehat{E}}$, so $\widehat{E}$ a Hilbert space with this inner product. For another proof, see Bourbaki [12]. $\square$

One of the most important facts about finite-dimensional Hermitian (and Euclidean) spaces is that they have orthonormal bases. This implies that, up to isomorphism, every finite-dimensional Hermitian space is isomorphic to $\mathbb{C}^n$ (for some $n \in \mathbb{N}$) and that the inner product is given by

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = \sum_{i=1}^{n} x_i \overline{y_i}.$$

Furthermore, every subspace W has an orthogonal complement $W^{\perp}$, and the inner product induces a natural duality between E and E^*, where E^* is the space of linear forms on E.

When E is a Hilbert space, E may be infinite dimensional, often of uncountable dimension. Thus, we can't expect that E always have an orthonormal basis. However, if we modify the notion of basis so that a "Hilbert basis" is an orthogonal family that is also dense in E, i.e., every $v \in E$ is the limit of a sequence of finite combinations of vectors from the Hilbert basis, then we can recover most of the "nice" properties of finite-dimensional Hermitian spaces. For instance, if $(u_k)_{k\in K}$ is a Hilbert basis, for every $v \in E$, we can define the Fourier coefficients $c_k = \langle v, u_k\rangle / \|u_k\|$, and then, v is the "sum" of its Fourier series $\sum_{k\in K} c_k u_k$. However, the cardinality of the index set K can be very large, and it is necessary to define what it means for a family of vectors indexed by K to be summable. We will do this in Section D.2. It turns out that every Hilbert space is isomorphic to a space of the form $\ell^2(K)$, where $\ell^2(K)$ is a generalization of the space of Example D.1 (see Theorem D.19, usually called the Riesz–Fischer theorem).

Our first goal is to prove that a closed subspace of a Hilbert space has an orthogonal complement. We also show that duality holds if we redefine the dual E' of E to be the space of *continuous* linear maps on E. Our presentation closely follows Bourbaki [12]. We also were inspired by Rudin [79], Lang [62, 63], Schwartz [84, 83], and Dixmier [27]. In fact, we highly recommend Dixmier [27] as a clear and simple text on the basics of topology and analysis. We first prove the so-called projection lemma.

Recall that in a metric space E, a subset X of E is *closed* iff for every convergent sequence (x_n) of points $x_n \in X$, the limit $x = \lim_{n\to\infty} x_n$ also belongs to X. The *closure* $\overline{X}$ *of* X is

the set of all limits of convergent sequences (x_n) of points $x_n \in X$. Obviously, $X \subseteq \overline{X}$. We say that the subset X of E is *dense in* E iff $E = \overline{X}$, the closure of X, which means that every $a \in E$ is the limit of some sequence (x_n) of points $x_n \in X$. Convex sets will again play a crucial role.

First, we state the following easy "parallelogram inequality", whose proof is left as an exercise.

Proposition D.2. *If E is a Hermitian space, for any two vectors $u, v \in E$, we have*

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

From the above, we get the following proposition:

Proposition D.3. *If E is a Hermitian space, given any $d, \delta \in \mathbb{R}$ such that $0 \leq \delta < d$, let*

$$B = \{u \in E \mid \|u\| < d\} \quad \textit{and} \quad C = \{u \in E \mid \|u\| \leq d + \delta\}.$$

For any convex set such A that $A \subseteq C - B$, we have

$$\|v - u\| \leq \sqrt{12 d\delta},$$

for all $u, v \in A$ (see Figure D.1).

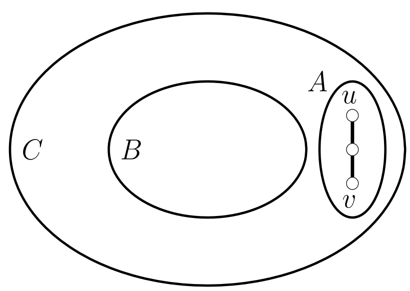


Figure D.1: Inequality of Proposition D.3

Proof. Since A is convex, $\frac{1}{2}(u + v) \in A$ if $u, v \in A$, and thus, $\|\frac{1}{2}(u + v)\| \geq d$. From the parallelogram inequality written in the form

$$\left\|\frac{1}{2}(u + v)\right\|^2 + \left\|\frac{1}{2}(u - v)\right\|^2 = \frac{1}{2}\left(\|u\|^2 + \|v\|^2\right),$$

since $\delta < d$, we get

$$\left\|\frac{1}{2}(u - v)\right\|^2 = \frac{1}{2}\left(\|u\|^2 + \|v\|^2\right) - \left\|\frac{1}{2}(u + v)\right\|^2 \leq (d + \delta)^2 - d^2 = 2d\delta + \delta^2 \leq 3d\delta,$$

from which

$$\|v - u\| \leq \sqrt{12 d\delta}.$$

$\square$

If X is a nonempty subset of a metric space (E, d), for any $a \in E$, recall that we define the *distance* $d(a, X)$ *of* a *to* X as

$$d(a, X) = \inf_{b \in X} d(a, b).$$

Also, the *diameter* $\delta(X)$ *of* X is defined by

$$\delta(X) = \sup\{d(a, b) \mid a, b \in X\}.$$

It is possible that $\delta(X) = \infty$. We leave the following standard two facts as an exercise (see Dixmier [27]):

Proposition D.4. *Let E be a metric space.*

(1) For every subset $X \subseteq E$, $\delta(X) = \delta(\overline{X})$.

(2) If E is a complete metric space, for every sequence (F_n) of closed nonempty subsets of E such that $F_{n+1} \subseteq F_n$, if $\lim_{n \to \infty} \delta(F_n) = 0$, then $\bigcap_{n=1}^{\infty} F_n$ consists of a single point.

We are now ready to prove the crucial projection lemma.

Proposition D.5. *(Projection lemma) Let E be a Hilbert space.*

(1) For any nonempty convex and closed subset $X \subseteq E$, for any $u \in E$, there is a unique vector $p_X(u) \in X$ such that

$$\|u - p_X(u)\| = \inf_{v \in X} \|u - v\| = d(u, X).$$

(2) The vector $p_X(u)$ is the unique vector $w \in E$ satisfying the following property (see Figure D.2):

$$w \in X \quad \text{and} \quad \Re \langle u - w, z - w \rangle \leq 0 \quad \text{for all } z \in X. \tag{$*$}$$

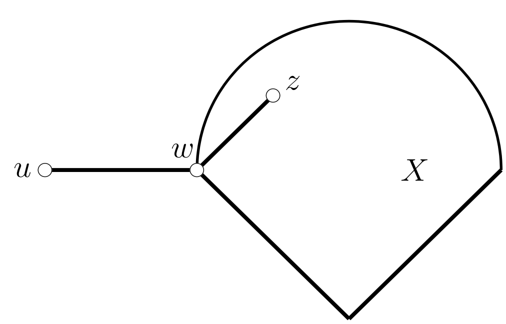


Figure D.2: Inequality of Proposition D.5

Proof. (1) Let $d = \inf_{v \in X} \|u - v\| = d(u, X)$. We define a sequence X_n of subsets of X as follows: for every $n \geq 1$,

$$X_n = \left\{ v \in X \mid \|u - v\| \leq d + \frac{1}{n} \right\}.$$

It is immediately verified that each X_n is nonempty (by definition of d), convex, and that $X_{n+1} \subseteq X_n$. Also, by Proposition D.3, we have

$$\sup\{\|w - v\| \mid v, w \in X_n\} \leq \sqrt{12d/n},$$

and thus, $\bigcap_{n \geq 1} X_n$ contains at most one point. We will prove that $\bigcap_{n \geq 1} X_n$ contains exactly one point, namely, $p_X(u)$. For this, define a sequence $(w_n)_{n \geq 1}$ by picking some $w_n \in X_n$ for every $n \geq 1$. We claim that $(w_n)_{n \geq 1}$ is a Cauchy sequence. Given any $\epsilon > 0$, if we pick N such that

$$N > \frac{12d}{\epsilon^2},$$

since $(X_n)_{n \geq 1}$ is a monotonic decreasing sequence, for all $m, n \geq N$, we have

$$\|w_m - w_n\| \leq \sqrt{12d/N} < \epsilon,$$

as desired. Since E is complete, the sequence $(w_n)_{n \geq 1}$ has a limit w, and since $w_n \in X$ and X is closed, we must have $w \in X$. Also observe that

$$\|u - w\| \leq \|u - w_n\| + \|w_n - w\|,$$

and since w is the limit of $(w_n)_{n \geq 1}$ and

$$\|u - w_n\| \leq d + \frac{1}{n},$$

given any $\epsilon > 0$, there is some n large enough so that

$$\frac{1}{n} < \frac{\epsilon}{2} \quad \text{and} \quad \|w_n - w\| \leq \frac{\epsilon}{2},$$

and thus

$$\|u - w\| \leq d + \epsilon.$$

Since the above holds for every $\epsilon > 0$, we have $\|u - w\| = d$. Thus, $w \in X_n$ for all $n \geq 1$, which proves that $\bigcap_{n \geq 1} X_n = \{w\}$. Now, any $z \in X$ such that $\|u - z\| = d(u, X) = d$ also belongs to every X_n, and thus $z = w$, proving the uniqueness of w, which we denote as $p_X(u)$.

(2) Let $w \in X$. Since X is convex, $z = (1 - \lambda)p_X(u) + \lambda w \in X$ for every λ, $0 \leq \lambda \leq 1$. Then, we have

$$\|u - z\| \geq \|u - p_X(u)\|$$

for all λ, $0 \leq \lambda \leq 1$, and since

$$\begin{aligned}\|u - z\|^2 &= \|u - p_X(u) - \lambda(w - p_X(u))\|^2 \\ &= \|u - p_X(u)\|^2 + \lambda^2\|w - p_X(u)\|^2 - 2\lambda\Re\,\langle u - p_X(u), w - p_X(u)\rangle\,,\end{aligned}$$

for all λ, $0 < \lambda \leq 1$, we get

$$\Re\,\langle u - p_X(u), w - p_X(u)\rangle = \frac{1}{2\lambda}\,\left(\|u - p_X(u)\|^2 - \|u - z\|^2\right) + \frac{\lambda}{2}\|w - p_X(u)\|^2,$$

and since this holds for every λ, $0 < \lambda \leq 1$ and

$$\|u - z\| \geq \|u - p_X(u)\|,$$

we have

$$\Re\,\langle u - p_X(u), w - p_X(u)\rangle \leq 0.$$

Conversely, assume that $w \in X$ satisfies the condition

$$\Re\,\langle u - w, z - w\rangle \leq 0$$

for all $z \in X$. For all $z \in X$, we have

$$\|u - z\|^2 = \|u - w\|^2 + \|z - w\|^2 - 2\Re\,\langle u - w, z - w\rangle \geq \|u - w\|^2,$$

which implies that $\|u - w\| = d(u, X) = d$, and from (1), that $w = p_X(u)$. $\square$

The vector $p_X(u)$ is called the *projection of* u *onto* X, and the map $p_X\colon E \to X$ is called the *projection of* E *onto* X. In the case of a real Hilbert space, there is an intuitive geometric interpretation of the condition

$$\langle u - p_X(u), z - p_X(u)\rangle \leq 0$$

for all $z \in X$. If we restate the condition as

$$\langle u - p_X(u), p_X(u) - z\rangle \geq 0$$

for all $z \in X$, this says that the absolute value of the measure of the angle between the vectors $u - p_X(u)$ and $p_X(u) - z$ is at most $\pi/2$. This makes sense, since X is convex, and points in X must be on the side opposite to the "tangent space" to X at $p_X(u)$, which is orthogonal to $u - p_X(u)$. Of course, this is only an intuitive description, since the notion of tangent space has not been defined!

The map $p_X\colon E \to X$ is continuous, as shown below.

Proposition D.6. *Let E be a Hilbert space. For any nonempty convex and closed subset $X \subseteq E$, the map $p_X\colon E \to X$ is continuous.*

Proof. For any two vectors $u, v \in E$, let $x = p_X(u) - u$, $y = p_X(v) - p_X(u)$, and $z = v - p_X(v)$. Clearly,

$$v - u = x + y + z,$$

and from Proposition D.5 (2), we also have

$$\Re \langle x, y \rangle \geq 0 \quad \text{and} \quad \Re \langle z, y \rangle \geq 0,$$

from which we get

$$\begin{aligned}
\|v - u\|^2 &= \|x + y + z\|^2 = \|x + z + y\|^2 \\
&= \|x + z\|^2 + \|y\|^2 + 2\Re \langle x, y \rangle + 2\Re \langle z, y \rangle \\
&\geq \|y\|^2 = \|p_X(v) - p_X(u)\|^2.
\end{aligned}$$

However, $\|p_X(v) - p_X(u)\| \leq \|v - u\|$ obviously implies that p_X is continuous. $\square$

We can now prove the following important proposition.

Proposition D.7. *Let E be a Hilbert space.*

(1) For any closed subspace $V \subseteq E$, we have $E = V \oplus V^{\perp}$, and the map $p_V \colon E \to V$ is linear and continuous.

(2) For any $u \in E$, the projection $p_V(u)$ is the unique vector $w \in E$ such that

$$w \in V \quad \textit{and} \quad \langle u - w, z \rangle = 0 \quad \textit{for all } z \in V.$$

Proof. (1) First, we prove that $u - p_V(u) \in V^{\perp}$ for all $u \in E$. For any $v \in V$, since V is a subspace, $z = p_V(u) + \lambda v \in V$ for all $\lambda \in \mathbb{C}$, and since V is convex and nonempty (since it is a subspace), and closed by hypothesis, by Proposition D.5 (2), we have

$$\Re(\overline{\lambda} \langle u - p_V(u), v \rangle) = \Re(\langle u - p_V(u), \lambda v \rangle = \Re \langle u - p_V(u), z - p_V(u) \rangle \leq 0$$

for all $\lambda \in \mathbb{C}$. In particular, the above holds for $\lambda = \langle u - p_V(u), v \rangle$, which yields

$$|\langle u - p_V(u), v \rangle| \leq 0,$$

and thus, $\langle u - p_V(u), v \rangle = 0$. As a consequence, $u - p_V(u) \in V^{\perp}$ for all $u \in E$. Since $u = p_V(u) + u - p_V(u)$ for every $u \in E$, we have $E = V + V^{\perp}$. On the other hand, since $\langle -, - \rangle$ is positive definite, $V \cap V^{\perp} = \{0\}$, and thus $E = V \oplus V^{\perp}$.

We already proved in Proposition D.6 that $p_V \colon E \to V$ is continuous. Also, since

$$p_V(\lambda u + \mu v) - (\lambda p_V(u) + \mu p_V(v)) = p_V(\lambda u + \mu v) - (\lambda u + \mu v) + \lambda(u - p_V(u)) + \mu(v - p_V(v)),$$

for all $u, v \in E$, and since the left-hand side term belongs to V, and from what we just showed, the right-hand side term belongs to $V^{\perp}$, we have

$$p_V(\lambda u + \mu v) - (\lambda p_V(u) + \mu p_V(v)) = 0,$$

showing that p_V is linear.

(2) This is basically obvious from (1). We proved in (1) that $u - p_V(u) \in V^{\perp}$, which is exactly the condition

$$\langle u - p_V(u), z \rangle = 0$$

for all $z \in V$. Conversely, if $w \in V$ satisfies the condition

$$\langle u - w, z \rangle = 0$$

for all $z \in V$, since $w \in V$, every vector $z \in V$ is of the form $y - w$, with $y = z + w \in V$, and thus, we have

$$\langle u - w, y - w \rangle = 0$$

for all $y \in V$, which implies the condition of Proposition D.5 (2):

$$\Re \langle u - w, y - w \rangle \leq 0$$

for all $y \in V$. By Proposition D.5, $w = p_V(u)$ is the projection of u onto V. $\square$

Definition D.2. Let E be a Hilbert space. For any closed subspace $V \subseteq E$, the linear and continuous map $p_V \colon E \to V$ given by Proposition D.7 is called the *orthogonal projection* of E onto V (recall that $E = V \oplus V^{\perp}$).

Let us illustrate the power of Proposition D.7 on the following "least squares" problem. Given a real $m \times n$-matrix A and some vector $b \in \mathbb{R}^m$, we would like to solve the linear system

$$Ax = b$$

in the least-squares sense, which means that we would like to find some solution $x \in \mathbb{R}^n$ that minimizes the Euclidean norm $\|Ax - b\|$ of the error $Ax - b$. It is actually not clear that the problem has a solution, but it does! The problem can be restated as follows: Is there some $x \in \mathbb{R}^n$ such that

$$\|Ax - b\| = \inf_{y \in \mathbb{R}^n} \|Ay - b\|,$$

or equivalently, is there some $z \in \mathrm{Im}\,(A)$ such that

$$\|z - b\| = d(b, \mathrm{Im}\,(A)),$$

where $\mathrm{Im}\,(A) = \{Ay \in \mathbb{R}^m \mid y \in \mathbb{R}^n\}$, the image of the linear map induced by A. Since $\mathrm{Im}\,(A)$ is a closed subspace of $\mathbb{R}^m$, because we are in finite dimension, Proposition D.7 tells us that there is a unique $z \in \mathrm{Im}\,(A)$ such that

$$\|z - b\| = \inf_{y \in \mathbb{R}^n} \|Ay - b\|,$$

and thus, the problem always has a solution since $z \in \mathrm{Im}\,(A)$, and since there is at least some $x \in \mathbb{R}^n$ such that $Ax = z$ (by definition of $\mathrm{Im}\,(A)$). Note that such an x is not necessarily unique. Furthermore, Proposition D.7 also tells us that $z \in \mathrm{Im}\,(A)$ is the solution of the equation

$$\langle z - b, w \rangle = 0 \quad \text{for all } w \in \mathrm{Im}\,(A),$$

or equivalently, that $x \in \mathbb{R}^n$ is the solution of

$$\langle Ax - b, Ay \rangle = 0 \quad \text{for all } y \in \mathbb{R}^n,$$

which is equivalent to

$$\langle A^\top(Ax - b), y \rangle = 0 \quad \text{for all } y \in \mathbb{R}^n,$$

and thus, since the inner product is positive definite, to $A^\top(Ax - b) = 0$, i.e.,

$$A^\top Ax = A^\top b.$$

Therefore, the solutions of the original least-squares problem are precisely the solutions of the the so-called *normal equations*

$$A^\top Ax = A^\top b,$$

discovered by Gauss and Legendre around 1800. We also proved that the normal equations always have a solution.

Computationally, it is best not to solve the normal equations directly, and instead, to use methods such as the QR-decomposition (applied to A) or the SVD-decomposition (in the form of the pseudo-inverse). We will come back to this point later on.

As another corollary of Proposition D.7, for any continuous nonnull linear map $h\colon E \to \mathbb{C}$, the null space

$$H = \mathrm{Ker}\, h = \{u \in E \mid h(u) = 0\} = h^{-1}(0)$$

is a closed hyperplane H, and thus, $H^\perp$ is a subspace of dimension one such that $E = H \oplus H^\perp$. This suggests defining the dual space of E as the set of all continuous maps $h\colon E \to \mathbb{C}$.

Remark: If $h\colon E \to \mathbb{C}$ is a linear map which is **not** continuous, then it can be shown that the hyperplane $H = \mathrm{Ker}\, h$ is dense in E! Thus, $H^\perp$ is reduced to the trivial subspace $\{0\}$. This goes against our intuition of what a hyperplane in $\mathbb{R}^n$ (or $\mathbb{C}^n$) is, and warns us not to trust our "physical" intuition too much when dealing with infinite dimensions.

Definition D.3. Given two vector spaces E and F over the complex field $\mathbb{C}$, a function $f\colon E \to F$ is *semilinear* if

$$\begin{aligned} f(u + v) &= f(u) + f(v), \\ f(\lambda u) &= \overline{\lambda} f(u), \end{aligned}$$

for all $u, v \in E$ and all $\lambda \in \mathbb{C}$.

Instead of defining semilinear maps, we can define the vector space $\overline{E}$ as the vector space with the same carrier set E whose addition is the same as that of E, but whose multiplication by a complex number λ is given by

$$(\lambda, u) \mapsto \overline{\lambda}u.$$

Then it is easy to check that a function $f\colon E \to \mathbb{C}$ is semilinear iff $f\colon \overline{E} \to \mathbb{C}$ is linear.

A fundamental fact about a *finite-dimensional* hermitian space is that the hermitian inner product induces a bijection (i.e., independent of the choice of bases) between the vector space E and its dual space E^*.

Given a Hermitian space E, for any vector $u \in E$, let $\varphi^l_u\colon E \to \mathbb{C}$ be the map defined such that

$$\varphi^l_u(v) = \overline{\langle u, v\rangle}, \quad \text{for all } v \in E.$$

Similarly, for any vector $v \in E$, let $\varphi^r_v\colon E \to \mathbb{C}$ be the map defined such that

$$\varphi^r_v(u) = \langle u, v\rangle, \quad \text{for all } u \in E.$$

Since the Hermitian product is linear in its first argument u, the map φ^r_v is a linear form in E^*, and since it is semilinear in its second argument v, the map φ^l_u is also a linear form in E^*. Thus, we have two maps $\flat^l\colon E \to E^*$ and $\flat^r\colon E \to E^*$, defined such that

$$\flat^l(u) = \varphi^l_u, \quad \text{and} \quad \flat^r(v) = \varphi^r_v.$$

Actually, $\varphi^l_u = \varphi^r_u$ and $\flat^l = \flat^r$. Indeed, for all $u, v \in E$, we have

$$\begin{aligned} \flat^l(u)(v) &= \varphi^l_u(v) \\ &= \overline{\langle u, v\rangle} \\ &= \langle v, u\rangle \\ &= \varphi^r_u(v) \\ &= \flat^r(u)(v). \end{aligned}$$

Therefore, we use the notation φ_u for both φ^l_u and φ^r_u, and $\flat$ for both $\flat^l$ and $\flat^r$.

Theorem D.8. *let E be a Hermitian space E. The map $\flat\colon E \to E^*$ defined such that*

$$\flat(u) = \varphi_u \quad \textit{for all } u \in E$$

is semilinear and injective. When E is also of finite dimension, the map $\flat\colon \overline{E} \to E^*$ is a canonical isomorphism.

Proof. That $\flat\colon E \to E^*$ is a semilinear map follows immediately from the fact that $\flat = \flat^r$, and that the Hermitian product is semilinear in its second argument. If $\varphi_u = \varphi_v$, then $\varphi_u(w) = \varphi_v(w)$ for all $w \in E$, which by definition of φ_u and φ_v means that

$$\langle w, u\rangle = \langle w, v\rangle$$

for all $w \in E$, which by semilinearity on the right is equivalent to

$$\langle w, v - u \rangle = 0 \quad \text{for all } w \in E,$$

which implies that $u = v$, since the Hermitian product is positive definite. Thus, $\flat\colon E \to E^*$ is injective. Finally, when E is of finite dimension n, E^* is also of dimension n, and then $\flat\colon E \to E^*$ is bijective. Since $\flat$ is semilinar, the map $\flat\colon \overline{E} \to E^*$ is an isomorphism. $\square$

However, if E is infinite dimensional, the map $\flat\colon E \to E^*$ is not surjective, since the linear forms of the form $u \mapsto \langle u, v \rangle$ (for some fixed vector $v \in E$) are continuous (the inner product is continuous), but there are linear forms that are not continuous.

We now show that by redefining the dual space of a Hilbert space as the set of continuous linear forms on E, we recover Theorem D.8.

Definition D.4. Given a Hilbert space E, we define the *dual space E' of E* as the vector space of all continuous linear forms $h\colon E \to \mathbb{C}$. Maps in E' are also called *bounded linear operators, bounded linear functionals, or simply, operators or functionals*.

Theorem D.8 is generalized to Hilbert spaces as follows.

Theorem D.9. *(Riesz representation theorem) Let E be a Hilbert space. Then, the map $\flat\colon E \to E'$ defined such that*

$$\flat(v) = \varphi_v,$$

is semilinear, continuous, and bijective.

Proof. The proof is basically identical to the proof of Theorem D.8, except that a different argument is required for the surjectivity of $\flat\colon E \to E'$, since E may not be finite dimensional. For any nonnull linear operator $h \in E'$, the hyperplane $H = \operatorname{Ker} h = h^{-1}(0)$ is a closed subspace of E, and by Proposition D.7, $H^{\perp}$ is a subspace of dimension one such that $E = H \oplus H^{\perp}$. Then, picking any nonnull vector $w \in H^{\perp}$, observe that H is also the kernel of the linear operator φ_w, with

$$\varphi_w(u) = \langle u, w \rangle,$$

and thus, since any two nonzero linear forms defining the same hyperplane must be proportional, there is some nonzero scalar $\lambda \in \mathbb{C}$ such that $h = \lambda\varphi_w$. But then, $h = \varphi_{\overline{\lambda} w}$, proving that $\flat\colon E \to E'$ is surjective. $\square$

Theorem D.9 is known as the *Riesz representation theorem, or "Little Riesz Theorem."* It shows that the inner product on a Hilbert space induces a natural linear isomorphism between E and its dual E'.

Remarks:

(1) Actually, the map $\flat\colon E \to E'$ turns out to be an isometry. To show this, we need to recall the notion of norm of a linear map, which we do not want to do right now.

(2) Many books on quantum mechanics use the so-called Dirac notation to denote objects in the Hilbert space E and operators in its dual space E'. In the Dirac notation, an element of E is denoted as $|x\rangle$, and an element of E' is denoted as $\langle t|$. The scalar product is denoted as $\langle t| \cdot |x\rangle$. This uses the isomorphism between E and E', except that the inner product is assumed to be semi-linear on the left, rather than on the right.

Theorem D.9 allows us to define the adjoint of a continuous linear map, as in the Hermitian case.

Proposition D.10. *Given a Hilbert space E, for every continuous linear map $f\colon E \to E$, there is a unique linear map $f^*\colon E \to E$, such that*

$$\langle f^*(u), v\rangle = \langle u, f(v)\rangle$$

for all $u, v \in E$. The map f^* is called the adjoint of f.

Proposition D.11 will show that if f is continuous, then f^* is also continuous. As in the Hermitian case, given two Hilbert spaces E and F, for any continuous linear map $f\colon E \to F$, such that

$$\langle f(u), v\rangle_2 = \langle u, f^*(v)\rangle_1$$

for all $u \in E$ and all $v \in F$. The linear map f^* is also called the adjoint of f.

The following results will be needed in Section 11.5.

Proposition D.11. *Let E a Hilbert space. For every continuous linear map $f\colon E \to E$, we have*

$$\begin{aligned}\|f^*\| &= \|f\|\\ \|f^* \circ f\| &= \|f\|^2\\ \|f \circ f^*\| &= \|f^* \circ f\|.\end{aligned}$$

In the above equations, we use the operator norm induced by the inner product on E. The first equation implies that f^* is continuous.

Proof. Since f^* is the adjoint of f we have

$$\langle f(x), y\rangle = \langle x, f^*(y)\rangle \quad \text{for all } x, y \in E.$$

By the Cauchy-Schwarz inrquality and properties of the operator norm,

$$|\langle x, f^*(y)\rangle| = |\langle f(x), y\rangle| \leq \|f(x)\| \, \|y\| \leq \|f\| \, \|x\| \, \|y\|.$$

If we let $x = f^*(y)$, we obtain

$$\|f^*(y)\|^2 \leq \|f\| \, \|f^*(y)\| \, \|y\|,$$

which implies that

$$\|f^*(y)\| \leq \|f\| \, \|y\| \,, \quad \text{for all } y \in E,$$

so by definition of the operator norm $\|f^*\|$,

$$\|f^*\| \leq \|f\| \,.$$

Repeating the same argument with f^* substituted for f and the fact that $(f^*)^* = f$ we get $\|f\| \leq \|f^*\|$, and so $\|f^*\| = \|f\|$.

Since $(f^*)^* = f$, the map f is the adjoint of f^* and we have

$$\langle f^*(f(x)), x\rangle = \langle f(x), f(x)\rangle = \|f(x)\|^2 \quad \text{for all } x \in E,$$

so by the Cauchy-Schwarz inequality,

$$\|f(x)\|^2 \leq \|f^*(f(x))\| \, \|x\| \,.$$

Since we are using the operator norm,

$$\|f(x)\|^2 \leq \|f^*(f(x))\| \, \|x\| \leq \|f^* \circ f\| \, \|x\|^2 \quad \text{for all } x \in E,$$

which implies (first take square roots) that

$$\|f\|^2 \leq \|f^* \circ f\| \,.$$

However, by a well-known property of the operator norm and the fact that $\|f^*\| = \|f\|$, we have

$$\|f^* \circ f\| \leq \|f^*\| \, \|f\| = \|f\|^2 \,.$$

Therefore, $\|f^* \circ f\| = \|f\|^2$.

The above equation with f replaced by f^* yields $\|f \circ f^*\| = \|f^*\|^2$, and since $\|f^*\| = \|f\|$, we obtain $\|f \circ f^*\| = \|f^*\|^2 = \|f\|^2 = \|f^* \circ f\|$, which is the third equation. $\square$

As a corollary of Proposition D.11, if f is self-adjoint, that is, $f^* = f$, then

$$\|f \circ f\| = \|f\| \,.$$

D.2 Total Orthogonal Families (Hilbert Bases), Fourier Coefficients

We conclude our quick tour of Hilbert spaces by showing that the notion of orthogonal basis can be generalized to Hilbert spaces. However, the useful notion is not the usual notion of a basis, but a notion which is an abstraction of the concept of Fourier series. Every element of a Hilbert space is the "sum" of its Fourier series.

Definition D.5. Given a Hilbert space E, a family $(u_k)_{k\in K}$ of nonnull vectors is an *orthogonal family* iff the u_k are pairwise orthogonal, i.e., $\langle u_i, u_j\rangle = 0$ for all $i \neq j$ $(i, j \in K)$, and an *orthonormal family* iff $\langle u_i, u_j\rangle = \delta_{i,j}$, for all $i, j \in K$. A *total orthogonal family (or system)* or *Hilbert basis* is an orthogonal family that is dense in E. This means that for every $v \in E$, for every $\epsilon > 0$, there is some finite subset $I \subseteq K$ and some family $(\lambda_i)_{i\in I}$ of complex numbers, such that

$$\Big\| v - \sum_{i\in I} \lambda_i u_i \Big\| < \epsilon.$$

Given an orthogonal family $(u_k)_{k\in K}$, for every $v \in E$, for every $k \in K$, the scalar $c_k = \langle v, u_k\rangle / \|u_k\|^2$ is called the *k-th Fourier coefficient of v over $(u_k)_{k\in K}$*.

Remark: The terminology Hilbert basis is misleading, because a Hilbert basis $(u_k)_{k\in K}$ is not necessarily a basis in the algebraic sense. Indeed, in general, $(u_k)_{k\in K}$ does not span E. Intuitively, it takes linear combinations of the u_k's with infinitely many nonnull coefficients to span E. Technically, this is achieved in terms of limits. In order to avoid the confusion between bases in the algebraic sense and Hilbert bases, some authors refer to algebraic bases as *Hamel bases* and to total orthogonal families (or Hilbert bases) as *Schauder bases*.

Given an orthogonal family $(u_k)_{k\in K}$, for any finite subset I of K, we often call sums of the form $\sum_{i\in I} \lambda_i u_i$ *partial sums of Fourier series*, and if these partial sums converge to a limit denoted as $\sum_{k\in K} c_k u_k$, we call $\sum_{k\in K} c_k u_k$ a *Fourier series*.

However, we have to make sense of such sums! Indeed, when K is unordered or uncountable, the notion of limit or sum has not been defined. This can be done as follows (for more details, see Dixmier [27]):

Definition D.6. Given a normed vector space E (say, a Hilbert space), for any nonempty index set K, we say that a family $(u_k)_{k\in K}$ of vectors in E is *summable with sum $v \in E$* iff for every $\epsilon > 0$, there is some finite subset I of K, such that,

$$\Big\| v - \sum_{j\in J} u_j \Big\| < \epsilon$$

for every finite subset J with $I \subseteq J \subseteq K$. We say that the family $(u_k)_{k\in K}$ is *summable* iff there is some $v \in E$ such that $(u_k)_{k\in K}$ is summable with sum v. A family $(u_k)_{k\in K}$ is a *Cauchy family* iff for every $\epsilon > 0$, there is a finite subset I of K, such that,

$$\Big\| \sum_{j\in J} u_j \Big\| < \epsilon$$

for every finite subset J of K with $I \cap J = \emptyset$,

If $(u_k)_{k\in K}$ is summable with sum v, we usually denote v as $\sum_{k\in K} u_k$. The following technical proposition will be needed:

Proposition D.12. *Let E be a complete normed vector space (say, a Hilbert space).*

(1) For any nonempty index set K, a family $(u_k)_{k\in K}$ is summable iff it is a Cauchy family.

(2) Given a family $(r_k)_{k\in K}$ of nonnegative reals $r_k \geq 0$, if there is some real number $B > 0$ such that $\sum_{i\in I} r_i < B$ for every finite subset I of K, then $(r_k)_{k\in K}$ is summable and $\sum_{k\in K} r_k = r$, where r is least upper bound of the set of finite sums $\sum_{i\in I} r_i$ ($I \subseteq K$).

Proof. (1) If $(u_k)_{k\in K}$ is summable, for every finite subset I of K, let

$$u_I = \sum_{i\in I} u_i \quad \text{and} \quad u = \sum_{k\in K} u_k$$

For every $\epsilon > 0$, there is some finite subset I of K such that

$$\|u - u_L\| < \epsilon/2$$

for all finite subsets L such that $I \subseteq L \subseteq K$. For every finite subset J of K such that $I \cap J = \emptyset$, since $I \subseteq I \cup J \subseteq K$ and $I \cup J$ is finite, we have

$$\|u - u_{I\cup J}\| < \epsilon/2 \quad \text{and} \quad \|u - u_I\| < \epsilon/2,$$

and since

$$\|u_{I\cup J} - u_I\| \leq \|u_{I\cup J} - u\| + \|u - u_I\|$$

and $u_{I\cup J} - u_I = u_J$ since $I \cap J = \emptyset$, we get

$$\|u_J\| = \|u_{I\cup J} - u_I\| < \epsilon,$$

which is the condition for $(u_k)_{k\in K}$ to be a Cauchy family.

Conversely, assume that $(u_k)_{k\in K}$ is a Cauchy family. We define inductively a decreasing sequence (X_n) of subsets of E, each of diameter at most $1/n$, as follows: For $n = 1$, since $(u_k)_{k\in K}$ is a Cauchy family, there is some finite subset J_1 of K such that

$$\|u_J\| < 1/2$$

for every finite subset J of K with $J_1 \cap J = \emptyset$. We pick some finite subset J_1 with the above property, and we let $I_1 = J_1$ and

$$X_1 = \{u_I \mid I_1 \subseteq I \subseteq K,\ I \text{ finite}\}.$$

For $n \geq 1$, there is some finite subset J_{n+1} of K such that

$$\|u_J\| < 1/(2n + 2)$$

for every finite subset J of K with $J_{n+1} \cap J = \emptyset$. We pick some finite subset J_{n+1} with the above property, and we let $I_{n+1} = I_n \cup J_{n+1}$ and

$$X_{n+1} = \{u_I \mid I_{n+1} \subseteq I \subseteq K,\ I \text{ finite}\}.$$

Since $I_n \subseteq I_{n+1}$, it is obvious that $X_{n+1} \subseteq X_n$ for all $n \geq 1$. We need to prove that each X_n has diameter at most $1/n$. Since J_n was chosen such that

$$\|u_J\| < 1/(2n)$$

for every finite subset J of K with $J_n \cap J = \emptyset$, and since $J_n \subseteq I_n$, it is also true that

$$\|u_J\| < 1/(2n)$$

for every finite subset J of K with $I_n \cap J = \emptyset$ (since $I_n \cap J = \emptyset$ and $J_n \subseteq I_n$ implies that $J_n \cap J = \emptyset$). Then, for every two finite subsets J, L such that $I_n \subseteq J, L \subseteq K$, we have

$$\|u_{J-I_n}\| < 1/(2n) \quad \text{and} \quad \|u_{L-I_n}\| < 1/(2n),$$

and since

$$\|u_J - u_L\| \leq \|u_J - u_{I_n}\| + \|u_{I_n} - u_L\| = \|u_{J-I_n}\| + \|u_{L-I_n}\|,$$

we get

$$\|u_J - u_L\| < 1/n,$$

which proves that $\delta(X_n) \leq 1/n$. Now, if we consider the sequence of closed sets $(\overline{X_n})$, we still have $\overline{X_{n+1}} \subseteq \overline{X_n}$, and by Proposition D.4, $\delta(\overline{X_n}) = \delta(X_n) \leq 1/n$, which means that $\lim_{n\to\infty} \delta(\overline{X_n}) = 0$, and by Proposition D.4, $\bigcap_{n=1}^{\infty} \overline{X_n}$ consists of a single element u. We claim that u is the sum of the family $(u_k)_{k\in K}$.

For every $\epsilon > 0$, there is some $n \geq 1$ such that $n > 2/\epsilon$, and since $u \in \overline{X_m}$ for all $m \geq 1$, there is some finite subset J_0 of K such that $I_n \subseteq J_0$ and

$$\|u - u_{J_0}\| < \epsilon/2,$$

where I_n is the finite subset of K involved in the definition of X_n. However, since $\delta(X_n) \leq 1/n$, for every finite subset J of K such that $I_n \subseteq J$, we have

$$\|u_J - u_{J_0}\| \leq 1/n < \epsilon/2,$$

and since

$$\|u - u_J\| \leq \|u - u_{J_0}\| + \|u_{J_0} - u_J\|,$$

we get

$$\|u - u_J\| < \epsilon$$

for every finite subset J of K with $I_n \subseteq J$, which proves that u is the sum of the family $(u_k)_{k\in K}$.

(2) Since every finite sum $\sum_{i\in I} r_i$ is bounded by the uniform bound B, the set of these finite sums has a least upper bound $r \leq B$. For every $\epsilon > 0$, since r is the least upper bound of the finite sums $\sum_{i\in I} r_i$ (where I finite, $I \subseteq K$), there is some finite $I \subseteq K$ such that

$$\left| r - \sum_{i\in I} r_i \right| < \epsilon,$$

and since $r_k \geq 0$ for all $k \in K$, we have

$$\sum_{i\in I} r_i \leq \sum_{j\in J} r_j$$

whenever $I \subseteq J$, which shows that

$$\left| r - \sum_{j\in J} r_j \right| \leq \left| r - \sum_{i\in I} r_i \right| < \epsilon$$

for every finite subset J such that $I \subseteq J \subseteq K$, proving that $(r_k)_{k\in K}$ is summable with sum $\sum_{k\in K} r_k = r$. $\square$

Remark: The notion of summability implies that the sum of a family $(u_k)_{k\in K}$ is independent of any order on K. In this sense, it is a kind of "commutative summability". More precisely, it is easy to show that for every bijection $\varphi\colon K \to K$ (intuitively, a reordering of K), the family $(u_k)_{k\in K}$ is summable iff the family $(u_l)_{l\in\varphi(K)}$ is summable, and if so, they have the same sum.

The following proposition gives some of the main properties of Fourier coefficients. Among other things, at most countably many of the Fourier coefficient may be nonnull, and the partial sums of a Fourier series converge. Given an orthogonal family $(u_k)_{k\in K}$, we let $U_k = \mathbb{C}u_k$, and $p_{U_k}\colon E \to U_k$ is the projection of E onto U_k.

Proposition D.13. *Let E be a Hilbert space, $(u_k)_{k\in K}$ an orthogonal family in E, and V the closure of the subspace generated by $(u_k)_{k\in K}$. The following properties hold:*

(1) For every $v \in E$, for every finite subset $I \subseteq K$, we have

$$\sum_{i\in I} |c_i|^2 \leq \|v\|^2,$$

where the c_k are the Fourier coefficients of v.

(2) For every vector $v \in E$, if $(c_k)_{k\in K}$ are the Fourier coefficients of v, the following conditions are equivalent:

(2a) $v \in V$

(2b) The family $(c_k u_k)_{k\in K}$ is summable and $v = \sum_{k\in K} c_k u_k$.

(2c) The family $(|c_k|^2)_{k\in K}$ is summable and $\|v\|^2 = \sum_{k\in K} |c_k|^2$;

(3) The family $(|c_k|^2)_{k\in K}$ is summable, and we have the Bessel inequality:

$$\sum_{k\in K} |c_k|^2 \leq \|v\|^2.$$

As a consequence, at most countably many of the c_k may be nonzero. The family $(c_k u_k)_{k\in K}$ forms a Cauchy family, and thus, the Fourier series $\sum_{k\in K} c_k u_k$ converges in E to some vector $u = \sum_{k\in K} c_k u_k$. Furthermore, $u = p_V(v)$.

Proof. (1) Let

$$u_I = \sum_{i \in I} c_i u_i$$

for any finite subset I of K. We claim that $v - u_I$ is orthogonal to u_i for every $i \in I$. Indeed,

$$\begin{aligned}\langle v - u_I, u_i \rangle &= \left\langle v - \sum_{j \in I} c_j u_j, u_i \right\rangle \\ &= \langle v, u_i \rangle - \sum_{j \in I} c_j \langle u_j, u_i \rangle \\ &= \langle v, u_i \rangle - c_i \|u_i\|^2 \\ &= \langle v, u_i \rangle - \langle v, u_i \rangle = 0,\end{aligned}$$

since $\langle u_j, u_i \rangle = 0$ for all $i \neq j$ and $c_i = \langle v, u_i \rangle / \|u_i\|^2$. As a consequence, we have

$$\begin{aligned}\|v\|^2 &= \left\| v - \sum_{i \in I} c_i u_i + \sum_{i \in I} c_i u_i \right\|^2 \\ &= \left\| v - \sum_{i \in I} c_i u_i \right\|^2 + \left\| \sum_{i \in I} c_i u_i \right\|^2 \\ &= \left\| v - \sum_{i \in I} c_i u_i \right\|^2 + \sum_{i \in I} |c_i|^2,\end{aligned}$$

since the u_i are pairwise orthogonal, that is,

$$\|v\|^2 = \left\| v - \sum_{i \in I} c_i u_i \right\|^2 + \sum_{i \in I} |c_i|^2.$$

Thus,

$$\sum_{i \in I} |c_i|^2 \leq \|v\|^2,$$

as claimed.

(2) We prove the chain of implications $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

$(a) \Rightarrow (b)$: If $v \in V$, since V is the closure of the subspace spanned by $(u_k)_{k \in K}$, for every $\epsilon > 0$, there is some finite subset I of K and some family $(\lambda_i)_{i \in I}$ of complex numbers, such that

$$\left\| v - \sum_{i \in I} \lambda_i u_i \right\| < \epsilon.$$

Now, for every finite subset J of K such that $I \subseteq J$, we have

$$\begin{aligned}\left\| v - \sum_{i \in I} \lambda_i u_i \right\|^2 &= \left\| v - \sum_{j \in J} c_j u_j + \sum_{j \in J} c_j u_j - \sum_{i \in I} \lambda_i u_i \right\|^2 \\ &= \left\| v - \sum_{j \in J} c_j u_j \right\|^2 + \left\| \sum_{j \in J} c_j u_j - \sum_{i \in I} \lambda_i u_i \right\|^2,\end{aligned}$$

since $I \subseteq J$ and the u_j (with $j \in J$) are orthogonal to $v - \sum_{j\in J} c_j u_j$ by the argument in (1), which shows that

$$\Big\| v - \sum_{j\in J} c_j u_j \Big\| \leq \Big\| v - \sum_{i\in I} \lambda_i u_i \Big\| < \epsilon,$$

and thus, that the family $(c_k u_k)_{k\in K}$ is summable with sum v, so that

$$v = \sum_{k\in K} c_k u_k.$$

$(b) \Rightarrow (c)$: If $v = \sum_{k\in K} c_k u_k$, then for every $\epsilon > 0$, there some finite subset I of K, such that

$$\Big\| v - \sum_{j\in J} c_j u_j \Big\| < \sqrt{\epsilon},$$

for every finite subset J of K such that $I \subseteq J$, and since we proved in (1) that

$$\|v\|^2 = \Big\| v - \sum_{j\in J} c_j u_j \Big\|^2 + \sum_{j\in J} |c_j|^2,$$

we get

$$\|v\|^2 - \sum_{j\in J} |c_j|^2 < \epsilon,$$

which proves that $(|c_k|^2)_{k\in K}$ is summable with sum $\|v\|^2$.

$(c) \Rightarrow (a)$: Finally, if $(|c_k|^2)_{k\in K}$ is summable with sum $\|v\|^2$, for every $\epsilon > 0$, there is some finite subset I of K such that

$$\|v\|^2 - \sum_{j\in J} |c_j|^2 < \epsilon^2$$

for every finite subset J of K such that $I \subseteq J$, and again, using the fact that

$$\|v\|^2 = \Big\| v - \sum_{j\in J} c_j u_j \Big\|^2 + \sum_{j\in J} |c_j|^2,$$

we get

$$\Big\| v - \sum_{j\in J} c_j u_j \Big\| < \epsilon,$$

which proves that $(c_k u_k)_{k\in K}$ is summable with sum $\sum_{k\in K} c_k u_k = v$, and $v \in V$.

(3) Since $\sum_{i\in I} |c_i|^2 \leq \|v\|^2$ for every finite subset I of K, by Proposition D.12, the family $(|c_k|^2)_{k\in K}$ is summable. The Bessel inequality

$$\sum_{k\in K} |c_k|^2 \leq \|v\|^2$$

is an obvious consequence of the inequality $\sum_{i\in I} |c_i|^2 \leq \|v\|^2$ (for every finite $I \subseteq K$). Now, for every natural number $n \geq 1$, if K_n is the subset of K consisting of all c_k such that $|c_k| \geq 1/n$, the number of elements in K_n is at most

$$\sum_{k\in K_n} |nc_k|^2 \leq n^2 \sum_{k\in K} |c_k|^2 \leq n^2\|v\|^2,$$

which is finite, and thus, at most a countable number of the c_k may be nonzero.

Since $(|c_k|^2)_{k\in K}$ is summable with sum c, for every $\epsilon > 0$, there is some finite subset I of K such that

$$\sum_{j\in J} |c_j|^2 < \epsilon^2$$

for every finite subset J of K such that $I \cap J = \emptyset$. Since

$$\Big\|\sum_{j\in J} c_j u_j\Big\|^2 = \sum_{j\in J} |c_j|^2,$$

we get

$$\Big\|\sum_{j\in J} c_j u_j\Big\| < \epsilon.$$

This proves that $(c_k u_k)_{k\in K}$ is a Cauchy family, which, by Proposition D.12, implies that $(c_k u_k)_{k\in K}$ is summable, since E is complete. Thus, the Fourier series $\sum_{k\in K} c_k u_k$ is summable, with its sum denoted $u \in V$.

Since $\sum_{k\in K} c_k u_k$ is summable with sum u, for every $\epsilon > 0$, there is some finite subset I_1 of K such that

$$\Big\|u - \sum_{j\in J} c_j u_j\Big\| < \epsilon$$

for every finite subset J of K such that $I_1 \subseteq J$. By the triangle inequality, for every finite subset I of K,

$$\Big\|u - v\Big\| \leq \Big\|u - \sum_{i\in I} c_i u_i\Big\| + \Big\|\sum_{i\in I} c_i u_i - v\Big\|.$$

By (2), every $w \in V$ is the sum of its Fourier series $\sum_{k\in K} \lambda_k u_k$, and for every $\epsilon > 0$, there is some finite subset I_2 of K such that

$$\Big\|w - \sum_{j\in J} \lambda_j u_j\Big\| < \epsilon$$

for every finite subset J of K such that $I_2 \subseteq J$. By the triangle inequality, for every finite subset I of K,

$$\Big\|v - \sum_{i\in I} \lambda_i u_i\Big\| \leq \|v - w\| + \Big\|w - \sum_{i\in I} \lambda_i u_i\Big\|.$$

Letting $I = I_1 \cup I_2$, since we showed in (2) that

$$\Big\| v - \sum_{i \in I} c_i u_i \Big\| \leq \Big\| v - \sum_{i \in I} \lambda_i u_i \Big\|$$

for every finite subset I of K, we get

$$\begin{aligned} \|u - v\| &\leq \Big\| u - \sum_{i \in I} c_i u_i \Big\| + \Big\| \sum_{i \in I} c_i u_i - v \Big\| \\ &\leq \Big\| u - \sum_{i \in I} c_i u_i \Big\| + \Big\| \sum_{i \in I} \lambda_i u_i - v \Big\| \\ &\leq \Big\| u - \sum_{i \in I} c_i u_i \Big\| + \|v - w\| + \Big\| w - \sum_{i \in I} \lambda_i u_i \Big\|, \end{aligned}$$

and thus

$$\|u - v\| \leq \|v - w\| + 2\epsilon.$$

Since this holds for every $\epsilon > 0$, we have

$$\|u - v\| \leq \|v - w\|$$

for all $w \in V$, i.e. $\|v - u\| = d(v, V)$, with $u \in V$, which proves that $u = p_V(v)$. $\square$

D.3 The Hilbert Space $\ell^2(K)$ and the Riesz–Fischer Theorem

Proposition D.13 suggests looking at the space of sequences $(z_k)_{k \in K}$ (where $z_k \in \mathbb{C}$) such that $(|z_k|^2)_{k \in K}$ is summable. Indeed, such spaces are Hilbert spaces, and it turns out that every Hilbert space is isomorphic to one of those. Such spaces are the infinite-dimensional version of the spaces $\mathbb{C}^n$ under the usual Euclidean norm.

Definition D.7. Given any nonempty index set K, the space $\ell^2(K)$ is the set of all sequences $(z_k)_{k \in K}$, where $z_k \in \mathbb{C}$, such that $(|z_k|^2)_{k \in K}$ is summable, i.e., $\sum_{k \in K} |z_k|^2 < \infty$.

Remarks:

(1) When K is a finite set of cardinality n, $\ell^2(K)$ is isomorphic to $\mathbb{C}^n$.

(2) When $K = \mathbb{N}$, the space $\ell^2(\mathbb{N})$ is the space ℓ^2 from Example D.1. It is a Hilbert space, and we now prove this fact for any index set K.

Proposition D.14. *Given any nonempty index set K, the space $\ell^2(K)$ is a Hilbert space under the Hermitian product*

$$\langle (x_k)_{k\in K}, (y_k)_{k\in K}\rangle = \sum_{k\in K} x_k\overline{y_k}.$$

The subspace consisting of sequences $(z_k)_{k\in K}$ such that $z_k = 0$, except perhaps for finitely many k, is a dense subspace of $\ell^2(K)$.

Proof. First, we need to prove that $\ell^2(K)$ is a vector space. Assume that $(x_k)_{k\in K}$ and $(y_k)_{k\in K}$ are in $\ell^2(K)$. This means that $(|x_k|^2)_{k\in K}$ and $(|y_k|^2)_{k\in K}$ are summable, which, in view of Proposition D.12, is equivalent to the existence of some positive bounds A and B such that $\sum_{i\in I}|x_i|^2 < A$ and $\sum_{i\in I}|y_i|^2 < B$, for every finite subset I of K. To prove that $(|x_k + y_k|^2)_{k\in K}$ is summable, it is sufficient to prove that there is some $C > 0$ such that $\sum_{i\in I}|x_i + y_i|^2 < C$ for every finite subset I of K. However, the parallelogram inequality implies that

$$\sum_{i\in I}|x_i + y_i|^2 \leq \sum_{i\in I} 2(|x_i|^2 + |y_i|^2) \leq 2(A + B),$$

for every finite subset I of K, and we conclude by Proposition D.12. Similarly, for every $\lambda \in \mathbb{C}$,

$$\sum_{i\in I}|\lambda x_i|^2 \leq \sum_{i\in I}|\lambda|^2|x_i|^2 \leq |\lambda|^2 A,$$

and $(\lambda_k x_k)_{k\in K}$ is summable. Therefore, $\ell^2(K)$ is a vector space.

By the Cauchy-Schwarz inequality,

$$\sum_{i\in I}|x_i\overline{y_i}| \leq \sum_{i\in I}|x_i||y_i| \leq \Big(\sum_{i\in I}|x_i|^2\Big)^{1/2}\Big(\sum_{i\in I}|y_i|^2\Big)^{1/2} \leq \sum_{i\in I}(|x_i|^2 + |y_i|^2)/2 \leq (A + B)/2,$$

for every finite subset I of K. Here, we used the fact that

$$4CD \leq (C + D)^2,$$

which is equivalent to

$$(C - D)^2 \geq 0.$$

By Proposition D.12, $(|x_k\overline{y_k}|)_{k\in K}$ is summable. The customary language is that $(x_k\overline{y_k})_{k\in K}$ is absolutely summable. However, it is a standard fact that this implies that $(x_k\overline{y_k})_{k\in K}$ is summable (For every $\epsilon > 0$, there is some finite subset I of K such that

$$\sum_{j\in J}|x_j\overline{y_j}| < \epsilon$$

for every finite subset J of K such that $I \cap J = \emptyset$, and thus

$$|\sum_{j\in J} x_j\overline{y_j}| \leq \sum_{i\in J}|x_j\overline{y_j}| < \epsilon,$$

proving that $(x_k\overline{y_k})_{k\in K}$ is a Cauchy family, and thus summable). We still have to prove that $\ell^2(K)$ is complete.

Consider a sequence $((\lambda_k^n)_{k\in K})_{n\geq 1}$ of sequences $(\lambda_k^n)_{k\in K} \in \ell^2(K)$, and assume that it is a Cauchy sequence. This means that for every $\epsilon > 0$, there is some $N \geq 1$ such that

$$\sum_{k\in K} |\lambda_k^m - \lambda_k^n|^2 < \epsilon^2$$

for all $m, n \geq N$. For every fixed $k \in K$, this implies that

$$|\lambda_k^m - \lambda_k^n| < \epsilon$$

for all $m, n \geq N$, which shows that $(\lambda_k^n)_{n\geq 1}$ is a Cauchy sequence in $\mathbb{C}$. Since $\mathbb{C}$ is complete, the sequence $(\lambda_k^n)_{n\geq 1}$ has a limit $\lambda_k \in \mathbb{C}$. We claim that $(\lambda_k)_{k\in K} \in \ell^2(K)$ and that this is the limit of $((\lambda_k^n)_{k\in K})_{n\geq 1}$.

Given any $\epsilon > 0$, the fact that $((\lambda_k^n)_{k\in K})_{n\geq 1}$ is a Cauchy sequence implies that there is some $N \geq 1$ such that for every finite subset I of K, we have

$$\sum_{i\in I} |\lambda_i^m - \lambda_i^n|^2 < \epsilon/4$$

for all $m, n \geq N$. Let $p = |I|$. Then,

$$|\lambda_i^m - \lambda_i^n| < \frac{\sqrt{\epsilon}}{2\sqrt{p}}$$

for every $i \in I$. Since λ_i is the limit of $(\lambda_i^n)_{n\geq 1}$, we can find some n large enough so that

$$|\lambda_i^n - \lambda_i| < \frac{\sqrt{\epsilon}}{2\sqrt{p}}$$

for every $i \in I$. Since

$$|\lambda_i^m - \lambda_i| \leq |\lambda_i^m - \lambda_i^n| + |\lambda_i^n - \lambda_i|,$$

we get

$$|\lambda_i^m - \lambda_i| < \frac{\sqrt{\epsilon}}{\sqrt{p}},$$

and thus,

$$\sum_{i\in I} |\lambda_i^m - \lambda_i|^2 < \epsilon,$$

for all $m \geq N$. Since the above holds for every finite subset I of K, by Proposition D.12, we get

$$\sum_{k\in K} |\lambda_k^m - \lambda_k|^2 < \epsilon,$$

for all $m \geq N$. This proves that $(\lambda_k^m - \lambda_k)_{k\in K} \in \ell^2(K)$ for all $m \geq N$, and since $\ell^2(K)$ is a vector space and $(\lambda_k^m)_{k\in K} \in \ell^2(K)$ for all $m \geq 1$, we get $(\lambda_k)_{k\in K} \in \ell^2(K)$. However,

$$\sum_{k\in K} |\lambda_k^m - \lambda_k|^2 < \epsilon$$

for all $m \geq N$, means that the sequence $(\lambda_k^m)_{k\in K}$ converges to $(\lambda_k)_{k\in K} \in \ell^2(K)$. The fact that the subspace consisting of sequences $(z_k)_{k\in K}$ such that $z_k = 0$ except perhaps for finitely many k is a dense suspace of $\ell^2(K)$ is left as an easy exercise. $\square$

Remark: The subspace consisting of all sequences $(z_k)_{k\in K}$ such that $z_k = 0$, except perhaps for finitely many k, provides an example of a subspace which is not closed in $\ell^2(K)$. Indeed, this space is strictly contained in $\ell^2(K)$, since there are countable sequences of nonnull elements in $\ell^2(K)$ (why?).

We just need two more propositions before being able to prove that every Hilbert space is isomorphic to some $\ell^2(K)$.

Proposition D.15. *Let E be a Hilbert space, and $(u_k)_{k\in K}$ an orthogonal family in E. The following properties hold:*

(1) For every family $(\lambda_k)_{k\in K} \in \ell^2(K)$, the family $(\lambda_k u_k)_{k\in K}$ is summable. Furthermore, $v = \sum_{k\in K} \lambda_k u_k$ is the only vector such that $c_k = \lambda_k$ for all $k \in K$, where the c_k are the Fourier coefficients of v.

(2) For any two families $(\lambda_k)_{k\in K} \in \ell^2(K)$ and $(\mu_k)_{k\in K} \in \ell^2(K)$, if $v = \sum_{k\in K} \lambda_k u_k$ and $w = \sum_{k\in K} \mu_k u_k$, we have the following equation, also called Parseval identity:

$$\langle v, w \rangle = \sum_{k\in K} \lambda_k \overline{\mu_k}.$$

Proof. (1) The fact that $(\lambda_k)_{k\in K} \in \ell^2(K)$ means that $(|\lambda_k|^2)_{k\in K}$ is summable. The proof given in Proposition D.13 (3) applies to the family $(|\lambda_k|^2)_{k\in K}$ (instead of $(|c_k|^2)_{k\in K}$), and yields the fact that $(\lambda_k u_k)_{k\in K}$ is summable. Letting $v = \sum_{k\in K} \lambda_k u_k$, recall that $c_k = \langle v, u_k \rangle / \|u_k\|^2$. Pick some $k \in K$. Since $\langle -, - \rangle$ is continuous, for every $\epsilon > 0$, there is some $\eta > 0$ such that

$$|\langle v, u_k \rangle - \langle w, u_k \rangle| < \epsilon \|u_k\|^2$$

whenever

$$\|v - w\| < \eta.$$

However, since for every $\eta > 0$, there is some finite subset I of K such that

$$\left\| v - \sum_{j\in J} \lambda_j u_j \right\| < \eta$$

for every finite subset J of K such that $I \subseteq J$, we can pick $J = I \cup \{k\}$, and letting $w = \sum_{j\in J} \lambda_j u_j$, we get

$$\left| \langle v, u_k \rangle - \left\langle \sum_{j\in J} \lambda_j u_j, u_k \right\rangle \right| < \epsilon \|u_k\|^2.$$

However,

$$\langle v, u_k \rangle = c_k \|u_k\|^2 \quad \text{and} \quad \left\langle \sum_{j\in J} \lambda_j u_j, u_k \right\rangle = \lambda_k \|u_k\|^2,$$

and thus, the above proves that $|c_k - \lambda_k| < \epsilon$ for every $\epsilon > 0$, and thus, that $c_k = \lambda_k$.

(2) Since $\langle -, - \rangle$ is continuous, for every $\epsilon > 0$, there are some $\eta_1 > 0$ and $\eta_2 > 0$, such that

$$| \langle x, y \rangle | < \epsilon$$

whenever $\|x\| < \eta_1$ and $\|y\| < \eta_2$. Since $v = \sum_{k\in K} \lambda_k u_k$ and $w = \sum_{k\in K} \mu_k u_k$, there is some finite subset I_1 of K such that

$$\left\| v - \sum_{j\in J} \lambda_j u_j \right\| < \eta_1$$

for every finite subset J of K such that $I_1 \subseteq J$, and there is some finite subset I_2 of K such that

$$\left\| w - \sum_{j\in J} \mu_j u_j \right\| < \eta_2$$

for every finite subset J of K such that $I_2 \subseteq J$. Letting $I = I_1 \cup I_2$, we get

$$\left| \left\langle v - \sum_{i\in I} \lambda_i u_i, w - \sum_{i\in I} \mu_i u_i \right\rangle \right| < \epsilon.$$

Furthermore,

$$\begin{aligned}
\langle v, w \rangle &= \left\langle v - \sum_{i\in I} \lambda_i u_i + \sum_{i\in I} \lambda_i u_i, w - \sum_{i\in I} \mu_i u_i + \sum_{i\in I} \mu_i u_i \right\rangle \\
&= \left\langle v - \sum_{i\in I} \lambda_i u_i, w - \sum_{i\in I} \mu_i u_i \right\rangle + \sum_{i\in I} \lambda_i \overline{\mu_i},
\end{aligned}$$

since the u_i are orthogonal to $v - \sum_{i\in I} \lambda_i u_i$ and $w - \sum_{i\in I} \mu_i u_i$ for all $i \in I$. This proves that for every $\epsilon > 0$, there is some finite subset I of K such that

$$\left| \langle v, w \rangle - \sum_{i\in I} \lambda_i \overline{\mu_i} \right| < \epsilon.$$

We already know from Proposition D.14 that $(\lambda_k\overline{\mu_k})_{k\in K}$ is summable, and since $\epsilon > 0$ is arbitrary, we get

$$\langle v, w\rangle = \sum_{k\in K} \lambda_k\overline{\mu_k}.$$

□

The next proposition states properties characterizing Hilbert bases (total orthogonal families).

Proposition D.16. *Let E be a Hilbert space, and let $(u_k)_{k\in K}$ be an orthogonal family in E. The following properties are equivalent:*

(1) The family $(u_k)_{k\in K}$ is a total orthogonal family.

(2) For every vector $v \in E$, if $(c_k)_{k\in K}$ are the Fourier coefficients of v, then the family $(c_k u_k)_{k\in K}$ is summable and $v = \sum_{k\in K} c_k u_k$.

(3) For every vector $v \in E$, we have the Parseval identity:

$$\|v\|^2 = \sum_{k\in K} |c_k|^2.$$

(4) For every vector $u \in E$, if $\langle u, u_k\rangle = 0$ for all $k \in K$, then $u = 0$.

Proof. The equivalence of (1), (2), and (3), is an immediate consequence of Proposition D.13 and Proposition D.15.

(4) If $(u_k)_{k\in K}$ is a total orthogonal family and $\langle u, u_k\rangle = 0$ for all $k \in K$, since $u = \sum_{k\in K} c_k u_k$ where $c_k = \langle u, u_k\rangle/\|u_k\|^2$, we have $c_k = 0$ for all $k \in K$, and $u = 0$.

Conversely, assume that the closure V of $(u_k)_{k\in K}$ is different from E. Then, by Proposition D.7, we have $E = V \oplus V^{\perp}$, where $V^{\perp}$ is the orthogonal complement of V, and $V^{\perp}$ is nontrivial since $V \neq E$. As a consequence, there is some nonnull vector $u \in V^{\perp}$. But then, u is orthogonal to every vector in V, and in particular,

$$\langle u, u_k\rangle = 0$$

for all $k \in K$, which, by assumption, implies that $u = 0$, contradicting the fact that $u \neq 0$. □

Remarks:

(1) If E is a Hilbert space and $(u_k)_{k\in K}$ is a total orthogonal family in E, there is a simpler argument to prove that $u = 0$ if $\langle u, u_k\rangle = 0$ for all $k \in K$, based on the continuity of $\langle -, -\rangle$. The argument is to prove that the assumption implies that $\langle v, u\rangle = 0$ for all $v \in E$. Since $\langle -, -\rangle$ is positive definite, this implies that $u = 0$. By continuity of $\langle -, -\rangle$, for every $\epsilon > 0$, there is some $\eta > 0$ such that for every finite subset I of K, for every family $(\lambda_i)_{i\in I}$, for every $v \in E$,

$$\left|\langle v, u\rangle - \left\langle \sum_{i\in I} \lambda_i u_i, u\right\rangle\right| < \epsilon$$

whenever

$$\left\|v - \sum_{i\in I} \lambda_i u_i\right\| < \eta.$$

Since $(u_k)_{k\in K}$ is dense in E, for every $v \in E$, there is some finite subset I of K and some family $(\lambda_i)_{i\in I}$ such that

$$\left\|v - \sum_{i\in I} \lambda_i u_i\right\| < \eta,$$

and since by assumption, $\left\langle \sum_{i\in I} \lambda_i u_i, u\right\rangle = 0$, we get

$$|\langle v, u\rangle| < \epsilon.$$

Since this holds for every $\epsilon > 0$, we must have $\langle v, u\rangle = 0$

(2) If V is any nonempty subset of E, the kind of argument used in the previous remark can be used to prove that $V^\perp$ is closed (even if V is not), and that $V^{\perp\perp}$ is the closure of V.

We will now prove that every Hilbert space has some Hilbert basis. This requires using a fundamental theorem from set theory known as *Zorn's Lemma*, which we quickly review.

Given any set X with a partial ordering $\leq$, recall that a nonempty subset C of X is a *chain* if it is totally ordered (i.e., for all $x, y \in C$, either $x \leq y$ or $y \leq x$). A nonempty subset Y of X is *bounded* iff there is some $b \in X$ such that $y \leq b$ for all $y \in Y$. Some $m \in X$ is *maximal* iff for every $x \in X$, $m \leq x$ implies that $x = m$. We can now state Zorn's Lemma. For more details, see Rudin [79], Lang [61], or Artin [3].

Proposition D.17. *Given any nonempty partially ordered set X, if every (nonempty) chain in X is bounded, then X has some maximal element.*

We can now prove the existence of Hilbert bases. We define a partial order on families $(u_k)_{k\in K}$ as follows: For any two families $(u_k)_{k\in K_1}$ and $(v_k)_{k\in K_2}$, we say that

$$(u_k)_{k\in K_1} \leq (v_k)_{k\in K_2}$$

iff $K_1 \subseteq K_2$ and $u_k = v_k$ for all $k \in K_1$. This is clearly a partial order.

Proposition D.18. *Let E be a Hilbert space. Given any orthogonal family $(u_k)_{k\in K}$ in E, there is a total orthogonal family $(u_l)_{l\in L}$ containing $(u_k)_{k\in K}$.*

Proof. Consider the set $\mathcal{S}$ of all orthogonal families greater than or equal to the family $B = (u_k)_{k\in K}$. We claim that every chain in $\mathcal{S}$ is bounded. Indeed, if $C = (C_l)_{l\in L}$ is a chain in $\mathcal{S}$, where $C_l = (u_{k,l})_{k\in K_l}$, the union family

$$(u_k)_{k\in\bigcup_{l\in L} K_l}, \text{ where } u_k = u_{k,l} \text{ whenever } k \in K_l,$$

is clearly an upper bound for C, and it is immediately verified that it is an orthogonal family. By Zorn's Lemma D.17, there is a maximal family $(u_l)_{l\in L}$ containing $(u_k)_{k\in K}$. If $(u_l)_{l\in L}$ is not dense in E, then its closure V is strictly contained in E, and by Proposition D.7, the orthogonal complement $V^\perp$ of V is nontrivial since $V \neq E$. As a consequence, there is some nonnull vector $u \in V^\perp$. But then, u is orthogonal to every vector in $(u_l)_{l\in L}$, and we can form an orthogonal family strictly greater than $(u_l)_{l\in L}$ by adding u to this family, contradicting the maximality of $(u_l)_{l\in L}$. Therefore, $(u_l)_{l\in L}$ is dense in E, and thus, it is a Hilbert basis. ☐

Remark: It is possible to prove that all Hilbert bases for a Hilbert space E have index sets K of the same cardinality. For a proof, see Bourbaki [12].

At last, we can prove that every Hilbert space is isomorphic to some Hilbert space $\ell^2(K)$ for some suitable K.

Theorem D.19. *(Riesz–Fischer) For every Hilbert space E, there is some nonempty set K such that E is isomorphic to the Hilbert space $\ell^2(K)$. More specifically, for any Hilbert basis $(u_k)_{k\in K}$ of E, the maps $f\colon \ell^2(K) \to E$ and $g\colon E \to \ell^2(K)$ defined such that*

$$f\left((\lambda_k)_{k\in K}\right) = \sum_{k\in K} \lambda_k u_k \quad \textit{and} \quad g(u) = \left(\langle u, u_k\rangle/\|u_k\|^2\right)_{k\in K} = (c_k)_{k\in K},$$

are bijective linear isometries such that $g \circ f = \mathrm{id}$ and $f \circ g = \mathrm{id}$.

Proof. By Proposition D.15 (1), the map f is well defined, and it it clearly linear. By Proposition D.13 (3), the map g is well defined, and it is also clearly linear. By Proposition D.13 (2b), we have

$$f(g(u)) = u = \sum_{k\in K} c_k u_k,$$

and by Proposition D.15 (1), we have

$$g(f\left((\lambda_k)_{k\in K}\right)) = (\lambda_k)_{k\in K},$$

and thus $g \circ f = \mathrm{id}$ and $f \circ g = \mathrm{id}$. By Proposition D.15 (2), the linear map g is an isometry. Therefore, f is a linear bijection and an isometry between $\ell^2(K)$ and E, with inverse g. ☐

Remark: The surjectivity of the map $g\colon E \to \ell^2(K)$ is known as the *Riesz–Fischer* theorem.

Having done all this hard work, we sketch how these results apply to Fourier series. Again, we refer the readers to Rudin [79] or Lang [62, 63] for a comprehensive exposition.

Let $\mathcal{C}(T)$ denote the set of all periodic continuous functions $f\colon [-\pi, \pi] \to \mathbb{C}$ with period 2π. There is a Hilbert space $\mathrm{L}^2(T)$ containing $\mathcal{C}(T)$ and such that $\mathcal{C}(T)$ is dense in $\mathrm{L}^2(T)$, whose inner product is given by

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

The Hilbert space $\mathrm{L}^2(T)$ is the space of *Lebesgue square-integrable periodic functions* (of period 2π).

It turns out that the family $(e^{ikx})_{k\in\mathbb{Z}}$ is a total orthogonal family in $\mathrm{L}^2(T)$, because it is already dense in $\mathcal{C}(T)$ (for instance, see Rudin [79]). Then, the Riesz–Fischer theorem says that for every family $(c_k)_{k\in\mathbb{Z}}$ of complex numbers such that

$$\sum_{k\in\mathbb{Z}} |c_k|^2 < \infty,$$

there is a unique function $f \in \mathrm{L}^2(T)$ such that f is equal to its Fourier series

$$f(x) = \sum_{k\in\mathbb{Z}} c_k e^{ikx},$$

where the Fourier coefficients c_k of f are given by the formula

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt.$$

The Parseval theorem says that

$$\sum_{k=-\infty}^{+\infty} c_k\overline{d_k} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\overline{g(t)}dt$$

for all $f, g \in \mathrm{L}^2(T)$, where c_k and d_k are the Fourier coefficients of f and g.

Thus, there is an isomorphism between the two Hilbert spaces $\mathrm{L}^2(T)$ and $\ell^2(\mathbb{Z})$, which is the deep reason why the Fourier coefficients "work". Theorem D.19 implies that the Fourier series $\sum_{k\in\mathbb{Z}} c_k e^{ikx}$ of a function $f \in \mathrm{L}^2(T)$ converges to f in the L^2-sense, i.e., in the mean-square sense. This does not necessarily imply that the Fourier series converges to f pointwise! This is a subtle issue, and for more on this subject, the reader is referred to Lang [62, 63] or Schwartz [86, 87].

We can also consider the set $\mathcal{C}([-1,1])$ of continuous functions $f\colon [-1,1] \to \mathbb{C}$. There is a Hilbert space $\mathrm{L}^2([-1,1])$ containing $\mathcal{C}([-1,1])$ and such that $\mathcal{C}([-1,1])$ is dense in $\mathrm{L}^2([-1,1])$, whose inner product is given by

$$\langle f, g\rangle = \int_{-1}^{1} f(x)\overline{g(x)}dx.$$

The Hilbert space $\mathrm{L}^2([-1,1])$ is the space of *Lebesgue square-integrable functions* over $[-1,1]$. The Legendre polynomials $P_n(x)$ form a Hilbert basis of $\mathrm{L}^2([-1,1])$.

Recall that if we let f_n be the function

$$f_n(x) = (x^2-1)^n,$$

$P_n(x)$ is defined as follows:

$$P_0(x) = 1, \quad \text{and} \quad P_n(x) = \frac{1}{2^n n!}\, f_n^{(n)}(x),$$

where $f_n^{(n)}$ is the nth derivative of f_n. The reason for the leading coefficient is to get $P_n(1) = 1$. It can be shown with much efforts that

$$P_n(x) = \sum_{0\leq k\leq n/2} (-1)^k \,\frac{(2(n-k))!}{2^n(n-k)!k!(n-2k)!}\, x^{n-2k}.$$

Appendix E

Well-Ordered Sets, Ordinals, Cardinals, Alephs

The purpose of this chapter is to define the notions of ordinal, cardinal and alephs, and to review some of their main properties. Intuitively the ordinals are the equivalence classes of well-ordered sets under the equivalence relation of order-isomorphism (the order-types). This idea goes back to Cantor; see Levy [65] for a thorough discussion of this approach. However, such a definition does not make sense because the collection of well-ordered sets is not a set. To circumvent this difficulty, following Von Neumann, we can define an ordinal as a certain special type of set.

E.1 Well-Ordered Sets

We begin by reviewing the notions of partial orders, total orders, strict partial orders, and strict total orders. Given a set X and a binary relation $\preceq \subseteq X \times X$ on X, we write $x \preceq y$ for $(x, y) \in \preceq$ and $x \not\preceq y$ for $\neg(x \preceq y)$.

Definition E.1. Given a set X, a binary relation $\leq$ on X is a *partial order* if it satisfies the following properties:

(1) The relation $\leq$ is *reflexive*, which means that for all x, if $x \in X$, then $x \leq x$.

(2) The relation $\leq$ is *transitive*, which means that for all x, y, z, if $x, y, z \in X$, $x \leq y$ and $y \leq z$, then $x \leq z$.

(3) The relation $\leq$ is *antisymmetric*, which means that for all x, y, if $x, y \in X$, $x \leq y$ and $y \leq x$, then $x = y$. The pair $(X, \leq)$ is called a *partially ordered set*.

A binary relation $\leq$ on X is a *total order (or simple order)* if it is a partial order and if if it is *strongly connected*, which means that for all x, y, if $x, y \in X$, then either $x \leq y$ or $y \leq x$. The pair $(X, \leq)$ is called a *totally ordered set*.

The empty set (with the empty relation) is trivially a partially and a totally ordered set.

Example E.1.

(1) Given any nonempty set X, the inclusion relation $Y \subseteq Z$ on subsets Y and Z of X is a partial order which is not a total order if X has at least two elements.

(2) The set $\mathbb{N}$ of natural numbers with its usual ordering is a totally ordered set.

(3) The set $\mathbb{Z}$ of integers with its usual ordering is a totally ordered set.

(4) The relation $\ll$ on $\mathbb{N} \times \mathbb{N}$ defined such that for all $(m_1, n_1), (m_2, n_2) \in \mathbb{N} \times \mathbb{N}$,

$$(m_1, n_1) \ll (m_2, n_2) \quad \text{iff} \quad \begin{cases} m_1 = m_2 \text{ and } n_1 = n_2, \text{ or} \\ m_1 < m_2, \text{ or} \\ m_1 = m_2 \text{ and } n_1 < n_2 \end{cases}$$

is a total order.

Definition E.2. Given a set X, a binary relation $\leq$ on X is a *strict partial order* if it satisfies the following properties:

(1) The relation $\leq$ is *asymmetric*, which means that for all x, y, if $x, y \in X$, then either $x \not\leq y$ or $y \not\leq x$, equivalently $\neg((x \leq y) \wedge (y \leq x))$.

(2) The relation $\leq$ is *transitive*, which means that for all x, y, z, if $x, y, z \in X$, $x \leq y$ and $y \leq z$, then $x \leq z$. The pair $(X, \leq)$ is called a *strictly partially ordered set*.

A binary relation $\leq$ on X is a *strict total order (or strict simple order)* if it is a strict partial order and if it is *connected*, which means that for all x, y, if $x, y \in X$ and $x \neq y$, then $x \leq y$ or $y \leq x$. The pair $(X, \leq)$ is called a *strictly totally ordered set*.

The empty set (with the empty relation) is trivially a strictly partially and a strictly totally ordered set.

Example E.2.

(1) Given any nonempty set X, the strict inclusion relation $Y \subseteq Z$ and $Y \neq Z$ on subsets Y and Z of X is a strict partial order which is not a strict total order if X has at least two elements.

(2) The set $\mathbb{N}$ of natural numbers with the strict ordering $m < n$ (namely $m \leq n$ and $m \neq n$) is a strictly totally ordered set.

(3) The set $\mathbb{Z}$ of integers with its strict ordering $m < n$ (namely $m \leq n$ and $m \neq n$) is a strictly totally ordered set.

(4) The relation $\ll$ on $\mathbb{N} \times \mathbb{N}$ defined such that for all $(m_1, n_1), (m_2, n_2) \in \mathbb{N} \times \mathbb{N}$,

$$(m_1, n_1) \ll (m_2, n_2) \quad \text{iff} \quad \begin{cases} m_1 < n_1, \text{ or} \\ m_1 = n_1 \text{ and } m_2 < n_2 \end{cases}$$

is a strict total order.

Definition E.3. Given a set X, a partial order $\leq$ on X is a *well-order* if every nonempty subset Y of X has a smallest element, which can be expressed as follows: for all Y, if $Y \neq \emptyset$ and $Y \subseteq X$, then there is some $x \in Y$ such that for y, if $y \in Y$, then $x \leq y$. The pair $(X, \leq)$ is called a *well-ordered set*.

A strict partial order $\leq$ on X is a *strictly well-order* if every nonempty subset Y of X has a smallest element. The pair $(X, \leq)$ is called a *strictly well-ordered set*.

The empty set (with the empty relation) is trivially a well-ordered set and strictly well-ordered set. If a well-ordered set is nonempty, then by picking $Y = \{x, y\}$ for any $x, y \in X$, since Y must have a smallest element, we see that either $x \leq y$ or $y \leq x$, that is, a well-ordered set is totally ordered. The same reasoning shows that a strictly well-ordered set is strictly totally ordered.

Example E.3.

(1) The partial order of Example E.1 is not a well-order (in fact, it is not a total order).

(2) The set $\mathbb{N}$ is well-ordered under its natural ordering.

(3) The set $\mathbb{Z}$ is not well-ordered under its natural ordering. For example, the subset $\{n \in \mathbb{Z} \mid n \leq 0\}$ does not have a smallest element.

(4) The set $\mathbb{N} \times \mathbb{N}$ under the total order of Example E.1 is well-ordered.

Proposition E.1. *Let $(X, \leq)$ be a partially ordered set. The relation $<$ on X given by*

$$x < y \quad \textit{iff} \quad x \leq y \ \textit{and} \ x \neq y$$

is a strict partial order on X. If $(X, \leq)$ is a totally ordered set, then the relation $<$ on X defined above is a strict total order. If $(X, \leq)$ is a well-ordered set, then the relation $<$ on X defined above is a strict well-order.

Proof. Assume that $(X, \leq)$ is a partially ordered set. The relation $<$ is transitive because if $x < y$ and $y < z$, then $x \leq y$, $y \leq z$, $x \neq y$ and $y \neq z$, so by transitivity of $\leq$ we have $x \leq z$. If $x = z$, then $y \leq z$ is equivalent to $y \leq x$, and since $x \leq y$, and $\leq$ is antisymmetric, we get $x = y$, a contradiction. The relation $<$ is asymmetric, because if $x < y$ and $y < x$, then $x \leq y$, $y \leq x$ and $x \neq y$, but since $\leq$ is antisymmetric, $x = y$, a contradiction.

The other statements are left as exercises to the reader. ∎

We say that $(X, <)$ is the strictly partially ordered set associated with the partially ordered set $(X, \leq)$, *etc.*

A detailed exposition of the above results and much more can be found in Suppes [95].

The importance of well-orders has to do with the fact that they support a powerful induction principle.

Definition E.4. For any partially ordered set $(E, \leq)$, for any $x \in E$, the subset $s(x) = \{y \in E \mid y < x\} = \{y \in E \mid y \leq x,\ y \neq x\}$ is called an *initial segment* of E.

Theorem E.2. *Let $(E, \leq)$ be a well-ordered set. For any subset A of E, if for every $a \in E$,*

$$\text{if } a \in A \text{ whenever } b \in A \text{ for all } b \in E \text{ such that } b < a,$$

then $A = E$. Equivalently, for all $a \in E$, if $s(a) \subseteq A$ implies that $a \in A$, then $A = E$.

Proof. Suppose by contradiction that $A \neq E$. Then the subset $E - A$ is nonempty, and since E is well-ordered, it has a least element $b \notin A$. We claim that $s(b) \subseteq A$. Indeed, $y \in s(b)$ iff $y < b$, but then we can't have $y \in E - A$, because this would contradict the fact that b is the smallest element of $E - A$, so $y \in A$. Since $s(b) \subseteq A$, by hypothesis $b \in A$, a contradiction. $\square$

Theorem E.2 immediately implies the following induction principle.

Theorem E.3. *Let $(E, \leq)$ be a well-ordered set and let $P(x)$ be a first-order formula with free variable x. For every $a \in E$, if $P(a)$ holds whenever $P(b)$ holds for all $b \in E$ such that $b < a$, then $P(x)$ holds for all $x \in E$.*

Theorem E.3 follows immediately from Theorem E.2 by setting $A = \{a \in E \mid P(a) = \mathbf{true}\}$. The induction principle in Theorem E.3 is sometimes called *transfinite induction on a well-ordered set*. It is a generalization of complete induction on $\mathbb{N}$.

Definition E.5. Let $(X_1, \leq_1)$ and $(X_2, \leq_2)$ be two partially ordered sets. A function $f \colon X_1 \to X_2$ is an *(order) isomorphism* if it is a bijection and if

$$x \leq_1 y \quad \text{iff} \quad f(x) \leq_2 f(y), \quad \text{for all } x, y \in X_1.$$

The same definition applies if $(X_1, \leq_1)$ and $(X_2, \leq_2)$ are two strictly partially ordered sets, and if the orderings are total or well-orders.

Note that a well-ordered set may be isomorphic to a proper subset of itself. For example, $(\mathbb{N}, \leq)$ is isomorphic to $(2\mathbb{N}, \leq)$ (where $2\mathbb{N} = \{2n \mid n \in \mathbb{N}\}$). However, we have the following important results.

Proposition E.4. *Let $(E, \leq)$ be a well-ordered set. If $f \colon E \to E$ is a function such that for all $x, y \in E$, if $x \neq y$ and $x \leq y$ implies that $f(x) \neq f(y)$ and $f(x) \leq f(y)$, then*

$$x \leq f(x) \quad \text{for all } x \in E.$$

Using Proposition E.4 we can prove the following result.

Proposition E.5. *Let $(E_1, \leq_1)$ and $(E_2, \leq_2)$ be two well-ordered sets. If $f\colon E_1 \to E_2$ and $g\colon E_1 \to E_2$ are isomorphisms, then $f = g$.*

As a corollary of Proposition E.5 it can be shown that if $(E, \leq)$ is a well-ordered set, then there is no isomorphism between E and any initial segment $s(x) = \{y \in E \mid y < x\}$, for any $x \in E$.

E.2 Ordinals

Technically, the definition of an ordinal depends on the precise axiomatic definition chosen for set theory (in first-order logic), specifically whether individual constants other than the symbol $\emptyset$ (the empty set) are allowed. Suppes [95] allows such individual symbols. For simplicity we follow Krivine [60] who does not allow such symbols. What this means is that the sets under consideration only have other sets as members, building up from the empty set.

Definition E.6. An *ordinal* is a set α such that

(1) The membership relation $x \in y$ on α (with $x, y \in \alpha$) is a strict well-order.

(2) For every x, if $x \in \alpha$, then $x \subseteq \alpha$. By definition of the inclusion relation, this means that for all x, y, if $y \in x$ and $x \in \alpha$, then $y \in \alpha$. Sometimes it is said that α is a *transitive set*.[1]

Remark: One of the axioms of set theory, the *sum axiom, also called the union axiom*, states that for every set X, the collection of all y such that $y \in x$ for some $x \in X$ is a set, denoted $\bigcup X$ or $\bigcup_{x \in X} x$. Then Condition (2) of Definition E.6 is equivalent to the condition

$$\bigcup \alpha \subseteq \alpha.$$

This condition is used in Suppes [95] and Levy [65].

We see that Definition E.6 implies that an ordinal is a set of sets of sets, *etc.*

For example, $\emptyset$, $\{\emptyset\}$, $\{\emptyset, \{\emptyset\}\}$, $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ are ordinals, and more generally, if α is an ordinal, then $\alpha \cup \{\alpha\}$ is also an ordinal denoted α^+. The ordinal $\emptyset$ is also denoted by 0.

This is the method used by Von Neumann to define the natural numbers. The number 0 is represented by the empty set, 1 is represented by $\{\emptyset\} = \{0\}$, 2 is represented by $\{\emptyset, \{\emptyset\}\} = \{0, 1\}$, 3 is represented by $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{0, 1, 2\}$, and if α represents a natural number, then $\alpha^+ = \alpha \cup \{\alpha\}$ represents the natural number $\alpha + 1$. For this reason, we also denote α^+ as $\alpha + 1$.

[1] This use of the word transitive is unfortunate since it differs from its meaning in Definition E.1(2).

We now list (mostly) without proof the most important properties of ordinals. Proofs can be found in Suppes [95] and Krivine [60]. A more advanced, rigorous and very thorough presentation can be found in Levy [65].

Proposition E.6. *Let α be an ordinal.*

(1) For any $\xi \in \alpha$, we have $s(\xi) = \{\eta \in \alpha \mid \eta \in \xi\} = \xi$.

(2) If $\xi \in \alpha$, then ξ is an ordinal.

Proposition E.7. *For every ordinal α, we have $\alpha \notin \alpha$.*

Proof. For any $\xi \in \alpha$, since the membership relation $\in$ on α is a strict order, we have $\xi \notin \xi$. Then if $\alpha \in \alpha$, we also have $\alpha \notin \alpha$, a contradiction. $\square$

Using Theorem E.3 the following result can be shown.

Proposition E.8. *For any two ordinals α, β, if there is an isomorphism between α and β (each equipped with the strict order of membership), then $\alpha = \beta$.*

Proposition E.9. *For any two ordinals α, β, either $\alpha = \beta$, $\alpha \in \beta$, or $\beta \in \alpha$, and these three cases are mutually exclusive.*

Proposition E.9 implies that for any two ordinals α, β, we have $\alpha \subseteq \beta$ iff $\alpha = \beta$ or $\alpha \in \beta$. It follows that the relation $\alpha \subseteq \beta$ is a total order on the ordinals, and we also write $\alpha \leq \beta$ instead of $\alpha \subseteq \beta$ and $\alpha < \beta$ for $\alpha \in \beta$. Observe that the relation $\alpha \in \beta$ is the strict total order associated with the total order $\subseteq$.

Proposition E.10. *For any ordinal α, the ordinal $\alpha^+ = \alpha \cup \{\alpha\}$ is the smallest ordinal strictly greater than α.*

Proposition E.11. *For any set S of ordinals, the set $\beta = \bigcup_{\alpha \in S} \alpha = \bigcup S$ is an ordinal which is the least upper bound of the set S.*

Proposition E.12. *For any set S of ordinals, the membership relation on S is a strict well-order. As a consequence, for any ordinal α, the ordinals $\beta < \alpha$ form a strictly well-ordered set (under inclusion).*

Proposition E.13. *(Burali–Forti paradox) The collection of all ordinals is not a set.*

Proof. Assume that the collection of all ordinals is a set α. Then by Proposition E.12, the set α is strictly well-ordered. Also, by definition of the set α, if $\beta \in \alpha$, then β is an ordinal, and since by Proposition E.6(2), every $\xi \in \beta$ is an ordinal, we have $\xi \in \alpha$ (since α is the set of all ordinals), so $\beta \subseteq \alpha$. Then by definition of an ordinal, α is an ordinal, and since α is the set of all ordinals, $\alpha \in \alpha$, contradicting Proposition E.7. $\square$

Proposition E.14 confirms that the concept of ordinal captures the idea that the ordinals are the "order-types" of well-ordered sets.

Proposition E.14. *For every well-ordered set $(S, \leq)$, there is a unique ordinal α and a unique isomorphism between $(S, <)$ and α (where $(S, <)$ is the strictly well-ordered set associated with the well-ordered set $(S, \leq)$ and α is strictly well-ordered by the membership relation).*

Proposition E.14 is proven using Theorem E.3.

Finite and infinite ordinals are defined as follows.

Definition E.7. An ordinal α is *finite* if either $\alpha = \emptyset$ or for every $\beta \subseteq \alpha$ with $\beta \neq \emptyset$, there is some ordinal ξ such that $\beta = \xi + 1$. An *infinite ordinal* is an ordinal that is not finite.

Remark: Definition E.7 is the definition found in Levy [65] and Krivine [60]. A different definition is used in Suppes [95].

So far we don't know if infinite ordinals exist! The axiom of infinity asserts that infinite ordinals exist.

Axiom of Infinity. There exists an infinite ordinal.

It can be shown that the axiom of infinity is equivalent to the fact that the collection of finite ordinals is a set (which is an ordinal), denoted ω; see Krivine [60].

Remark: In an axiomatic presentation of the axioms of Zermelo–Frankel set theory it is customary to state a version of the axiom of infinity which does not involve the notion of ordinal. It can be shown that this version of the axiom of infinity is equivalent to the above version about ordinals. For this classical approach, see Suppes [95] and Levy [65]. Since it is not our intention to give an axiomatic presentation of Zermelo–Frankel set theory, the above version of the axiom of infinity is preferable.

Definition E.8. Under the axiom of infinity, the set of all finite ordinals is an ordinal denoted ω.

In the Von Neumann approach, the natural numbers are identified with the finite ordinals. Thus ω is the set of natural numbers and it is also denoted $\mathbb{N}$ by most mathematicians. The ordinal ω is not a finite ordinal. It is the smallest infinite ordinal because if ξ is an infinite ordinal such that $\xi \in \omega$, then ξ is a finite ordinal (ω is the set of all finite ordinals), a contradiction.

Definition E.9. An ordinal $\alpha \neq \emptyset$ is a *limit ordinal* if for all $\beta \in \alpha$, we also have $\beta + 1 \in \alpha$.

It is easy to see that an ordinal $\alpha \neq \emptyset$ is a limit ordinal iff there is no ordinal β such that $\alpha = \beta + 1$ iff

$$\alpha = \bigcup \alpha = \bigcup_{\beta \in \alpha} \beta$$

Furthermore, it can be shown that every limit ordinal is infinite and that the axiom of infinity is equivalent to the existence of a limit ordinal; see Krivine [60].

E.3 Cardinals, Alephs ($\aleph_\alpha$) and Beths ($\beth_\alpha$)

Having defined the ordinals, we can define cardinals and the cardinality of a set. This is where the axiom of choice shows its nose.

Definition E.10. A *cardinal* is an ordinal $\mathfrak{a}$ such that if β is any ordinal in bijection with $\mathfrak{a}$, then $\mathfrak{a} \subseteq \beta$.

A cardinal is often referred to as an *initial ordinal*. It appears that the universal notation adopted to denote cardinals is to use lower case German letters ("Fraktur" font), $\mathfrak{a}, \mathfrak{b}$, *etc*. This convention is convenient since if we denote ordinals by lower case Greek letters (as it is customary), then we have a visual mechanism to distinguish between ordinals and cardinals. As we will see shortly, cardinals are also denoted using the Hebrew letter aleph with an ordinal subscript ($\aleph_\alpha$).

Proposition E.15. *Every finite ordinal is a cardinal.*

Definition E.11. The smallest infinite ordinal ω is a cardinal, which is also denoted $\aleph_0$.

As we will see later, there is no largest cardinal, but this is not easy to prove; see Suppes [95] (Section 7.3, Theorem 60).

Assume that the **axiom of choice holds**. An easy-going version of the axiom of choice is that for any two nonempty sets X and Y, for any surjection $f\colon X \to Y$, there is some injection $g\colon Y \to X$ such that $f \circ g = \mathrm{id}_Y$.

Theorem E.16. *(Zermelo) Every set has some well-ordering.*

A proof of Theorem E.16 can be found in all set theory texts, in particular Suppes [95] Krivine [60]. Theorem E.16 is one of many results equivalent to the famous axiom of choice. If you think Theorem E.16 is obvious, try finding a well-ordering on the power set $\mathcal{P}(\mathbb{N})$ of the set $\mathbb{N}$ of natural numbers.

Now, *if we accept the axiom of choice*, since by Theorem E.16 every set X has some well-order (not unique if X has at least two elements), by Proposition E.14, there is a bijection between X and some ordinal α. Then it is not hard to show that the ordinals β that are in bijection with X form a set (because if γ is an ordinal in bijection with the power set $\mathcal{P}(X)$, then $\beta \in \gamma$), so by Proposition E.12, there is a smallest ordinal, denoted $|X|$, among the ordinals in bijection with X.

Definition E.12. Given any set X, the smallest ordinal $|X|$ (also denoted $\mathrm{card}(X)$) in bijection with X is a cardinal called the *cardinal number (or cardinality)* of X.

It can be shown that the collection of cardinals numbers is *not* a set.

Remark: It is possible to define the notion of cardinality of a set even if we do not assume the axiom of choice. But then the cardinal $|X|$ of set X is a certain kind of set that may not be an ordinal. In fact, the cardinal $|X|$ is an ordinal iff the set $|X|$ is well-orderable. See Levy [65] (Chapter III, Section 2).

Definition E.13. The cardinality of the set $\mathbb{R}$ of real numbers is denoted by $\mathfrak{c}$ and is called the *cardinality of the continuum* (or *power of the continuum*).

It is a standard theorem of set theory that there is a bijection between $\mathcal{P}(\mathbb{N})$, the power set of the set $\mathbb{N}$ of natural numbers, and the set $\mathbb{R}$ of real numbers; see Section 6.7 of Suppes [95].

Definition E.14. For any cardinal $\mathfrak{a}$, the cardinality of the power set $\mathcal{P}(\mathfrak{a})$ of $\mathfrak{a}$ is denoted $2^{\mathfrak{a}}$.

Using the above definition, the fact that there is a bijection between $\mathcal{P}(N)$ and $\mathbb{R}$ is restated as $\mathfrak{c} = 2^{\aleph_0}$. Cantor's theorem (which says that there is no surjection of a set X onto its power set $\mathcal{P}(X)$) stated in terms of cardinals says that for any cardinal $\mathfrak{a}$, we have

$$\mathfrak{a} < 2^{\mathfrak{a}}.$$

Our next goal is to show that it is possible to provide an enumeration of the infinite cardinals indexed by the ordinals. We first proceed informally. The idea is to define the infinite cardinal $\aleph_\alpha$ for every ordinal α as follows: the cardinal $\aleph_\alpha$ is the infinite cardinal β such that the set $\{\xi \mid \xi \in \beta,\ \xi \text{ is an infinite cardinal}\}$ is isomorphic (as a strictly well-ordered set under the membership relation) to α (also with the strict order of membership). Intuitively, $\aleph_\alpha$ is the α's infinite cardinal. So $\aleph_1$ is the smallest cardinal of cardinality strictly greater than $\aleph_0$, then $\aleph_2$ is the smallest cardinal of cardinality strictly greater than $\aleph_1$, and more generally $\aleph_{\alpha+1}$ is the smallest cardinal of cardinality strictly greater than $\aleph_\alpha$. See Definition E.17 for a rigorous approach (which needs to deal with the case where α is a limit ordinal).

Then Cantor's theorem implies that

$$\aleph_{\alpha+1} \subseteq 2^{\aleph_\alpha}.$$

Whether or not the above inequality is actually an equality is a famous problem called the *generalized continuum hypothesis*. For $\alpha = 0$, famous results of Gödel and Cohen show that the statement $\aleph_1 = \mathfrak{c} = 2^{\aleph_0}$ is independent of Zermelo–Frankel set theory (with the axiom of choice).

Our definition of the *alephs* (denoted $\aleph_\alpha$) is not rigorous. It is actually possible to define rigorously the alephs *without assuming the axiom of choice*, as explained in Suppes [95].

We need to recall the following notation.

Definition E.15. Let A and B be any two sets.

(1) We write $A \approx B$ if there is a bijection from A to B. In this case we say that A and B are *equipollent*.

(2) We write $A \preceq B$ if there is a subset C of B such that $A \approx C$.

(3) We write $A \prec B$ if $A \preceq B$ and $\neg(B \preceq A)$, and we write $A \succ B$ if $B \preceq A$ and $\neg(A \preceq B)$.

Two ordinals may be equipollent and yet be very different in terms of they order structure. A simple example consists of the two ordinals ω and $\omega + 1 = \omega \cup \{\omega\}$. We can define the bijection f from $\omega + 1$ to ω given by $f(\{\omega\}) = 0$ and $f(n) = n + 1$, for any $n \in \omega$.

If $\varphi(\alpha)$ is a first-order formula in which α ranges over the ordinals, it can be shown that if there is some ordinal α such that $\varphi(\alpha)$ holds, then there is a smallest ordinal β such that $\varphi(\beta)$ holds; see Suppes [95] (Section 7.1, Theorem 5). The above fact suggests the definition of the smallest ordinal $\mu_\alpha(\varphi(\alpha))$ satisfying a first-order formula $\varphi(\alpha)$ (where α denotes an ordinal). If $\forall\alpha\neg\varphi(\alpha)$, that is, $\varphi(\alpha)$ is not satisfied by any ordinal, then we set $\mu_\alpha(\varphi(\alpha)) = 0$.

Definition E.16. Given a first-order formula $\varphi(\alpha)$ where α denotes an ordinal, the ordinal $\mu_\alpha(\varphi(\alpha))$ is defined such that for every ordinal β, we have the equivalence

$$\mu_\alpha(\varphi(\alpha)) = \beta \quad \text{iff} \quad [\varphi(\beta) \wedge \forall\gamma(\varphi(\gamma) \Longrightarrow (\beta \subseteq \gamma))] \vee [\forall\alpha\neg\varphi(\alpha) \wedge (\beta = 0)].$$

Then it can be shown that

(1) If $\varphi(\beta)$ holds for some ordinal β, then $\mu_\alpha(\varphi(\alpha)) \subseteq \beta$.

(2) If $\exists\alpha\varphi(\alpha)$ holds, then $\varphi(\mu_\alpha(\varphi(\alpha)))$ holds.

The alephs are then defined by transfinite recursion as follows.

Definition E.17. The ordinals $\aleph_\alpha$ (the *alephs*) are defined as follows:

(1) $\aleph_0 = \omega$,

(2) For any successor ordinal $\alpha + 1$,

$$\aleph_{\alpha+1} = \mu_\beta(\beta \succ \aleph_\alpha).$$

(3) For any limit ordinal α,

$$\aleph_\alpha = \bigcup_{\beta\in\alpha} \aleph_\beta.$$

Actually, we really have to justify why a recursive definition as in Definition E.17 is legitimate. To do so requires delving into axiomatic set theory more than we want to for the purpose of this appendix. Let us just say that the *axiom schema of replacement* (due to Zermelo) is required. Intuitively, this axiom says that that if $\varphi(x, y)$ is a functional relation, which means that for all x, y_1, y_2, $\varphi(x, y_1)$ and $\varphi(x, y_2)$ implies that $y_1 = y_2$, then for any set A, the image of A by φ, that is, the collection of y such that $\varphi(x, y)$ for some $x \in A$, is also a set. Then a powerful version of definition by *transfinite recursion* can be established. For details, see Suppes [95] (Chapter 7). Incidentally, this version of transfinite recursion is also used to define addition, multiplication, and exponentiation of ordinals.

Returning to the alephs, the following properties can be shown; see Suppes [95] (Chapter 7).

Actually, it is not obvious at all that for every $\aleph_\alpha$, there is some ordinal β such that $\beta \succ \aleph_\alpha$, so that in Clause (2) of Definition E.17, some nonzero ordinal is returned. The next proposition shows that this is indeed the case; see Suppes [95] (Section 7.3, Theorem 63).

Proposition E.17. *For any ordinal α, there is some ordinal β such that for every ordinal $\gamma \in \alpha$ we have $\beta \succ \aleph_\gamma$.*

Proposition E.17 implies the following result which, together with the equation $\aleph_0 = \omega$, can be used as a definition of $\aleph_\alpha$.

Proposition E.18. *If α is a nonzero ordinal, then*

$$\aleph_\alpha = \mu_\beta(\forall\gamma((\gamma \in \alpha) \Longrightarrow (\beta \succ \aleph_\gamma))).$$

Proposition E.19. *For every ordinal α, the ordinal $\aleph_\alpha$ is an infinite cardinal.*

Proposition E.20. *For any two ordinals α, β, if $\alpha \in \beta$, then $\aleph_\alpha \in \aleph_\beta$.*

Proposition E.20 implies that there is no largest aleph.

Proposition E.21. *For any ordinal α, there is no infinite cardinal β such that $\aleph_\alpha \in \beta \in \aleph_{\alpha+1}$.*

It can also be shown that every cardinal $\aleph_\alpha$ is a limit ordinal; see Levy [65]. Finally, every infinite cardinal arises as some aleph, which means that there is an "enumeration" of the infinite cardinals by the ordinals.

Theorem E.22. *For every infinite cardinal $\mathfrak{a}$, there is an ordinal α (necessarily unique by Proposition E.20) such that $\mathfrak{a} = \aleph_\alpha$.*

All the above results do *not* rely on the axiom of choice. However, the axiom of choice is needed to show that every set has a cardinal (is in bijection with a cardinal).

Remark: Let us again assume that the axiom of choice holds. Then we can restate the generalized continuum hypothesis by introducing cardinals known as the *beth's*.

Definition E.18. We define by transfinite recursion the cardinals *beth* α, denoted $\beth_\alpha$, as follows: for every ordinal α,

$$\begin{aligned}
\beth_0 &= \aleph_0 \\
\beth_{\alpha+1} &= \mathrm{card}(\mathcal{P}(\beth_\alpha)) \\
\beth_\alpha &= \bigcup_{\beta<\alpha} \beth_\beta \quad \text{if } \alpha \text{ is a limit ordinal.}
\end{aligned}$$

Observe that

$$\beth_1 = \mathfrak{c},$$

the cardinality of the continuum. We can show by transfinite induction that

$$\aleph_\alpha \leq \beth_\alpha$$

for every ordinal α, and the generalized continuum hypothesis is restated as

$$\aleph_\alpha = \beth_\alpha$$

for every ordinal α. The continuum hypothesis is restated as

$$\aleph_1 = \beth_1.$$

Infinite ordinals beyond ω are very hard to understand. A way to get a better grasp of the infinite ordinals is to generalize the operations of addition, multiplication, and exponentiation defined on the natural numbers (the finite ordinals) to infinite ordinals. This is done by generalizing the familiar recursive definitions to infinite ordinals, and the trick for doing so is to extends these recursive definitions to limit ordinals. However such developments are too peripheral to harmonic analysis to be covered here.

Bibliography

[1] Lars V. Ahlfors and Leo Sario. *Riemann Surfaces*. Princeton Math. Series, No. 2. Princeton University Press, 1960.

[2] Tom M. Apostol. *Introduction to Analytic Number Theory*. Undergraduate Texts in Mathematics. Springer, first edition, 1976.

[3] Michael Artin. *Algebra*. Prentice Hall, first edition, 1991.

[4] Marcel Berger. *Géométrie 1*. Nathan, 1990. English edition: Geometry 1, Universitext, Springer Verlag.

[5] Armand Borel. *Essays in the History of Lie Groups and Algebraic Groups*, volume 21 of *History of Mathematics*. AMS, first edition, 2001.

[6] Nicolas Bourbaki. *Elements of Mathematics. Intégration, Chapters 1–4*. Eléments de Mathématiques. Hermann, second edition, 1965.

[7] Nicolas Bourbaki. *Elements of Mathematics. Intégration, Chapters 7–8. Mesure de Haar, Convolution et Représentations*. Eléments de Mathématiques. Hermann, second edition, 1965.

[8] Nicolas Bourbaki. *Elements of Mathematics. Intégration, Chapter 5. Intégration des mesures*. Eléments de Mathématiques. Hermann, second edition, 1967.

[9] Nicolas Bourbaki. *Elements of Mathematics. Théories Spectrales, Chapters 1–2*. Eléments de Mathématiques. Hermann, first edition, 1967.

[10] Nicolas Bourbaki. *Algèbre, Chapitre 8*. Eléments de Mathématiques. Hermann, 1968.

[11] Nicolas Bourbaki. *Elements of Mathematics. Intégration, Chapter 6. Intégration vectorielle*. Eléments de Mathématiques. Hermann, first edition, 1968.

[12] Nicolas Bourbaki. *Espaces Vectoriels Topologiques*. Eléments de Mathématiques. Masson, 1981.

[13] Nicolas Bourbaki. *Topologie Générale, Chapitres 1-4*. Eléments de Mathématiques. Masson, 1990.

[14] Nicolas Bourbaki. *Topologie Générale, Chapitres 5-10*. Eléments de Mathématiques. CCLS, 1990.

[15] Glen E Bredon. *Topology and Geometry*. GTM No. 139. Springer Verlag, first edition, 1993.

[16] T. Bröcker and T. tom Dieck. *Representations of Compact Lie Groups*. GTM, Vol. 98. Springer Verlag, first edition, 1985.

[17] Roger Carter, Graeme Segal, and Ian Macdonald. *Lectures on Lie Groups and Lie Algebras*. Cambridge University Press, first edition, 1995.

[18] Claude Chevalley. *Theory of Lie Groups I*. Princeton Mathematical Series, No. 8. Princeton University Press, first edition, 1946. Eighth printing.

[19] Anton Deitmar. *A First Course in Harmonic Analysis*. UTM. Springer Verlag, first edition, 2002.

[20] Jean Dieudonné. *Special Functions and Linear Representations of Lie Groups*. Regional Conference Series in Mathematics, No. 42. AMS, first edition, 1980.

[21] Jean Dieudonné. *Éléments d'Analyse, Tome V. Groupes de Lie Compacts, Groupes de Lie Semi-Simples*. Edition Jacques Gabay, first edition, 2003.

[22] Jean Dieudonné. *Éléments d'Analyse, Tome VI. Analyse Harmonique*. Edition Jacques Gabay, first edition, 2003.

[23] Jean Dieudonné. *Éléments d'Analyse, Tome VII. Équations Fonctionnelles Linéaires. Première partie, Opérateurs Pseudo-Différentiels*. Edition Jacques Gabay, first edition, 2003.

[24] Jean Dieudonné. *Éléments d'Analyse, Tome II. Chapitres XII à XV*. Edition Jacques Gabay, first edition, 2005.

[25] Jean Dieudonné. *Éléments d'Analyse, Tome I. Chapitres I à XI*. Edition Jacques Gabay, third edition, 2010.

[26] Jacques Dixmier. *Les C^* Algèbres et Leurs Représentations*. Edition Jacques Gabay, second edition, 1969.

[27] Jacques Dixmier. *General Topology*. UTM. Springer Verlag, first edition, 1984.

[28] B.A. Dubrovin, A.T. Fomenko, and S.P. Novikov. *Modern Geometry–Methods and Applications. Part I*. GTM No. 93. Springer Verlag, second edition, 1985.

[29] J.J. Duistermaat and J.A.C. Kolk. *Lie Groups*. Universitext. Springer Verlag, first edition, 2000.

[30] Nelson Dunford and Jacob T. Schwartz. *Linear Operators, Part I: General Theory*. Wiley, first edition, 1988.

[31] Gerald A. Edgar. *Measure, Topology, and Fractal Geometry*. Undergraduate Texts in Mathematics. Springer Verlag, first edition, 1992.

[32] Gerald B. Folland. *Fourier Analysis and its Applications*. Brooks/Cole Publishing Company, fgirst edition, 1992.

[33] Gerald B. Folland. *A Course in Abstract Harmonic Analysis*. CRC Press, first edition, 1995.

[34] Gerald B. Folland. *Real Analysis. Modern Techniques and Their Applications*. Wiley-Interscience, second edition, 1999.

[35] William Fulton. *Algebraic Topology, A first course*. GTM No. 153. Springer Verlag, first edition, 1995.

[36] William Fulton and Joe Harris. *Representation Theory, A first course*. GTM No. 129. Springer Verlag, first edition, 1991.

[37] Jean H. Gallier. *Geometric Methods and Applications, For Computer Science and Engineering*. TAM, Vol. 38. Springer, second edition, 2011.

[38] Jean H. Gallier and Jocelyn Quaintance. *Differential Geometry and Lie Groups. A Computational Perspective*. Geometry and Computing 12. Springer, first edition, 2020.

[39] Jean H. Gallier and Jocelyn Quaintance. *Differential Geometry and Lie Groups. A Second Course*. Geometry and Computing 13. Springer, first edition, 2020.

[40] Jean H. Gallier and Jocelyn Quaintance. *Linear Algebra and Optimization with Applications to Machine Learning, Volume I*. World Scientific, first edition, 2020.

[41] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Universitext. Springer Verlag, second edition, 1993.

[42] Brian Hall. *Lie Groups, Lie Algebras, and Representations. An Elementary Introduction*. GTM No. 222. Springer Verlag, first edition, 2003.

[43] Marshall Hall. *Theory of Groups*. AMS Chelsea, second edition, 1976.

[44] Paul R. Halmos. *Measure Theory*. GTM, Vol. No 18. Springer-Verlag, first edition, 1978.

[45] Sigurdur Helgason. *Geometric Analysis on Symmetric Spaces*. SURV, Vol. 39. AMS, first edition, 1994.

[46] Sigurdur Helgason. *Groups and Geometric Analysis. Integral Geometry, Invariant Differential Operators and Spherical Functions.* MSM, Vol. 83. AMS, first edition, 2000.

[47] Sigurdur Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces.* GSM, Vol. 34. AMS, first edition, 2001.

[48] Edwin Hewitt and Kenneth A. Ross. *Abstract Harmonic Analysis II.* Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, first edition, 1970.

[49] Edwin Hewitt and Kenneth A. Ross. *Abstract Harmonic Analysis I.* Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, second edition, 1979.

[50] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, first edition, 1990.

[51] James E. Humphreys. *Introduction to Lie Algebras and Representation Theory.* GTM No. 9. Springer Verlag, first edition, 1972.

[52] Jean-Pierre Kahane and Pierre Gilles Lemaré-Rieusset. *Séries de Fourier et Ondelettes.* Cassini, first edition, 1998.

[53] A.A. Kirillov. *Elements of the Theory of Representations.* Springer-Verlag, first edition, 1976.

[54] A.A. Kirillov. *Lectures on the Orbit Method.* GSM, Vol. 64. AMS, first edition, 2004.

[55] A.A. Kirillov (Ed.). *Representation Theory and Noncommutative Harmonic Analysis.* Encyclopaedia of Mathematical Sciences, Vol. 22. Springer Verlag, first edition, 1994.

[56] Anthony W. Knapp. *Representation Theory of Semisimple Groups.* Princeton Landmarks in Mathematics. Princeton University Press, first edition, 2001.

[57] Anthony W. Knapp. *Lie Groups Beyond an Introduction.* Progress in Mathematics, Vol. 140. Birkhäuser, second edition, 2002.

[58] A.N. Kolmogorov and S.V. Fomin. *Introductory Real Analysis.* Dover, first edition, 1975.

[59] Yvette Kosmann-Schwarzbach. *Groups and Symmetries. From Finite Groups to Lie Groups.* Universitext. Springer Verlag, first edition, 2010.

[60] Jean-Louis Krivine. *Théorie Axiomatique des Ensembles.* PUF, first edition, 1969.

[61] Serge Lang. *Algebra.* Addison Wesley, third edition, 1993.

[62] Serge Lang. *Real and Functional Analysis.* GTM 142. Springer Verlag, third edition, 1996.

[63] Serge Lang. *Undergraduate Analysis.* UTM. Springer Verlag, second edition, 1997.

[64] Peter Lax. *Functional Analysis.* Wiley, first edition, 2002.

[65] Azriel Levy. *Basic Set Theory.* Perspectives in Mathematical Logic. Springer-Verlag, first edition, 1979.

[66] George W. Mackey. *The Scope and History of Commutative and Noncommutative Harmonic Analysis*, volume 5 of *History of Mathematics.* AMS, first edition, 1992.

[67] Ib Madsen and Jorgen Tornehave. *From Calculus to Cohomology. De Rham Cohomology and Characteristic Classes.* Cambridge University Press, first edition, 1998.

[68] Paul Malliavian, Hélène Airault, Leslie Kay, and Gérard Letac. *Integration and Probability.* GTM, Vol. No 157. Springer-Verlag, first edition, 1995.

[69] Charles-Michel Marle. *Mesures et Probabilités.* Hermann, first edition, 1974.

[70] William S. Massey. *Algebraic Topology: An Introduction.* GTM No. 56. Springer Verlag, second edition, 1987.

[71] William S. Massey. *A Basic Course in Algebraic Topology.* GTM No. 127. Springer Verlag, first edition, 1991.

[72] John W. Milnor and James D. Stasheff. *Characteristic Classes.* Annals of Math. Series, No. 76. Princeton University Press, first edition, 1974.

[73] R. Mneimné and F. Testard. *Introduction à la Théorie des Groupes de Lie Classiques.* Hermann, first edition, 1997.

[74] James R. Munkres. *Analysis on Manifolds.* Addison Wesley, 1991.

[75] James R. Munkres. *Topology.* Prentice Hall, second edition, 2000.

[76] Barrett O'Neill. *Semi-Riemannian Geometry With Applications to Relativity.* Pure and Applies Math., Vol 103. Academic Press, first edition, 1983.

[77] L. Pontryagin. *Topological Groups.* Princeton University Press, first edition, 1939.

[78] L. Pontryagin. *Topological Groups.* Gordon and Breach, second edition, 1960.

[79] Walter Rudin. *Real and Complex Analysis.* McGraw Hill, third edition, 1987.

[80] Walter Rudin. *Functional Analysis.* McGraw Hill, second edition, 1991.

[81] Arthur A. Sagle and Ralph E. Walde. *Introduction to Lie Groups and Lie Algebras.* Academic Press, first edition, 1973.

[82] Hans Samelson. *Notes on Lie Algebras.* Universitext. Springer, second edition, 1990.

[83] Laurent Schwartz. *Topologie Générale et Analyse Fonctionnelle*. Collection Enseignement des Sciences. Hermann, 1980.

[84] Laurent Schwartz. *Analyse I. Théorie des Ensembles et Topologie*. Collection Enseignement des Sciences. Hermann, 1991.

[85] Laurent Schwartz. *Analyse II. Calcul Différentiel et Equations Différentielles*. Collection Enseignement des Sciences. Hermann, 1992.

[86] Laurent Schwartz. *Analyse III. Calcul Intégral*. Collection Enseignement des Sciences. Hermann, 1993.

[87] Laurent Schwartz. *Analyse IV. Applications à la Théorie de la Mesure*. Collection Enseignement des Sciences. Hermann, 1993.

[88] Denis Serre. *Matrices, Theory and Applications*. GTM No. 216. Springer Verlag, second edition, 2010.

[89] Jean-Pierre Serre. *A Course in Arithmetic*. Graduate Text in Mathematics, No. 7. Springer, first edition, 1973.

[90] Jean-Pierre Serre. *Linear Representations of Finite Groups*. GTM No. 42. Springer, first edition, 1977.

[91] Jean-Pierre Serre. *Lie Algebras and Lie Groups*. Lecture Notes in Mathematics, No. 1500. Springer, second edition, 1992.

[92] Jean-Pierre Serre. *Complex Semisimple Lie Algebras*. Springer Monographs in Mathematics. Springer, first edition, 2000.

[93] Barry Simon. *Representations of Finite and Compact Groups*. GSM, Vol. 10. AMS, first edition, 1996.

[94] Elias M. Stein and Rami Shakarchi. *Fourier Analysis. An Introduction*. Princeton Lectures in Analysis, I. Princeton University Press, 2003.

[95] Patrick Suppes. *Axiomatic Set Theory*. Dover, 1972.

[96] Michael E. Taylor. *Noncommutative Harmonic Analysis*. Mathematical Surveys and Monographs, No. 22. AMS, first edition, 1986.

[97] Audrey Terras. *Fourier Analysis on Finite Groups and Applications*. Student Text 43. Cambridge University Press, first edition, 1999.

[98] V.S. Varadarajan. *Lie Groups, Lie Algebras, and Their Representations*. GTM No. 102. Springer Verlag, first edition, 1984.

[99] V.S. Varadarajan. *An Introduction to Harmonic Analysis on Semisimple Lie Groups*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, first edition, 1999.

[100] O. Veblen and J. W. Young. *Projective Geometry, Vol. 2*. Ginn, first edition, 1946.

[101] N.J. Vilenkin. *Special Functions and the Theory of Group Representations*. Translations of Mathematical Monographs No 22. AMS, first edition, 1968.

[102] Frank Warner. *Foundations of Differentiable Manifolds and Lie Groups*. GTM No. 94. Springer Verlag, first edition, 1983.

[103] Garth Warner. *Harmonic Analysis on Semi-Simple Lie Groups, I*. Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, first edition, 1972.

[104] Garth Warner. *Harmonic Analysis on Semi-Simple Lie Groups, II*. Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, first edition, 1972.

[105] André Weil. *L'Intégration dans les Groupes Topologiques et ses Applications*. Hermann, second edition, 1979.

Symbol Index

Index