

Linear Algebra for Computer Vision, Robotics, and Machine Learning

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

© Jean Gallier

January 1, 2020

Preface

In recent years, computer vision, robotics, machine learning, and data science have been some of the key areas that have contributed to major advances in technology. Anyone who looks at papers or books in the above areas will be baffled by a strange jargon involving exotic terms such as kernel PCA, ridge regression, lasso regression, support vector machines (SVM), Lagrange multipliers, KKT conditions, *etc.* Do support vector machines chase cattle to catch them with some kind of super lasso? No! But one will quickly discover that behind the jargon which always comes with a new field (perhaps to keep the outsiders out of the club), lies a lot of “classical” linear algebra and techniques from optimization theory. And there comes the main challenge: in order to understand and use tools from machine learning, computer vision, and so on, one needs to have a firm background in linear algebra and optimization theory. To be honest, some probability theory and statistics should also be included, but we already have enough to contend with.

Many books on machine learning struggle with the above problem. How can one understand what are the dual variables of a ridge regression problem if one doesn’t know about the Lagrangian duality framework? Similarly, how is it possible to discuss the dual formulation of SVM without a firm understanding of the Lagrangian framework?

The easy way out is to sweep these difficulties under the rug. If one is just a consumer of the techniques we mentioned above, the cookbook recipe approach is probably adequate. But this approach doesn’t work for someone who really wants to do serious research and make significant contributions. To do so, we believe that one must have a solid background in linear algebra and optimization theory.

This is a problem because it means investing a great deal of time and energy studying these fields, but we believe that perseverance will be amply rewarded.

Our main goal is to present fundamentals of linear algebra and optimization theory, keeping in mind applications to machine learning, robotics, and computer vision. This work consists of two volumes, the first one being linear algebra, the second one optimization theory and applications, especially to machine learning.

This first volume covers “classical” linear algebra, up to and including the primary decomposition and the Jordan form. Besides covering the standard topics, we discuss a few topics that are important for applications. These include:

1. Haar bases and the corresponding Haar wavelets.
2. Hadamard matrices.

3. Affine maps (see Section 5.4).
4. Norms and matrix norms (Chapter 8).
5. Convergence of sequences and series in a normed vector space. The matrix exponential e^A and its basic properties (see Section 8.8).
6. The group of unit quaternions, $\mathbf{SU}(2)$, and the representation of rotations in $\mathbf{SO}(3)$ by unit quaternions (Chapter 15).
7. An introduction to algebraic and spectral graph theory.
8. Applications of SVD and pseudo-inverses, in particular, principal component analysis, for short PCA (Chapter 21).
9. Methods for computing eigenvalues and eigenvectors, with a main focus on the QR algorithm (Chapter 17).

Four topics are covered in more detail than usual. These are

1. Duality (Chapter 10).
2. Dual norms (Section 13.7).
3. The geometry of the orthogonal groups $\mathbf{O}(n)$ and $\mathbf{SO}(n)$, and of the unitary groups $\mathbf{U}(n)$ and $\mathbf{SU}(n)$.
4. The spectral theorems (Chapter 16).

Except for a few exceptions we provide complete proofs. We did so to make this book self-contained, but also because we believe that no deep knowledge of this material can be acquired without working out some proofs. However, our advice is to skip some of the proofs upon first reading, especially if they are long and intricate.

The chapters or sections marked with the symbol \otimes contain material that is typically more specialized or more advanced, and they can be omitted upon first (or second) reading.

Acknowledgement: We would like to thank Christine Allen-Blanchette, Kostas Daniilidis, Carlos Esteves, Spyridon Leonardos, Stephen Phillips, João Sedoc, Stephen Shatz, Jianbo Shi, Marcelo Siqueira, and C.J. Taylor for reporting typos and for helpful comments. Special thanks to Gilbert Strang. We learned much from his books which have been a major source of inspiration. Thanks to Steven Boyd and James Demmel whose books have been an invaluable source of information. The first author also wishes to express his deepest gratitude to Philippe G. Ciarlet who was his teacher and mentor in 1970-1972 while he was a student at ENPC in Paris. Professor Ciarlet was by far his best teacher. He also knew how to instill in his students the importance of intellectual rigor, honesty, and modesty. He still has his typewritten notes on measure theory and integration, and on numerical linear algebra. The latter became his wonderful book Ciarlet [14], from which we have borrowed heavily.

Contents

1	Introduction	11
2	Vector Spaces, Bases, Linear Maps	15
2.1	Motivations: Linear Combinations, Linear Independence, Rank	15
2.2	Vector Spaces	27
2.3	Indexed Families; the Sum Notation $\sum_{i \in I} a_i$	34
2.4	Linear Independence, Subspaces	40
2.5	Bases of a Vector Space	46
2.6	Matrices	53
2.7	Linear Maps	58
2.8	Linear Forms and the Dual Space	65
2.9	Summary	68
2.10	Problems	70
3	Matrices and Linear Maps	77
3.1	Representation of Linear Maps by Matrices	77
3.2	Composition of Linear Maps and Matrix Multiplication	82
3.3	Change of Basis Matrix	87
3.4	The Effect of a Change of Bases on Matrices	90
3.5	Summary	94
3.6	Problems	94
4	Haar Bases, Haar Wavelets, Hadamard Matrices	101
4.1	Introduction to Signal Compression Using Haar Wavelets	101
4.2	Haar Matrices, Scaling Properties of Haar Wavelets	103
4.3	Kronecker Product Construction of Haar Matrices	108
4.4	Multiresolution Signal Analysis with Haar Bases	110
4.5	Haar Transform for Digital Images	112
4.6	Hadamard Matrices	119
4.7	Summary	121
4.8	Problems	121
5	Direct Sums, Rank-Nullity Theorem, Affine Maps	125
5.1	Direct Products	125

5.2	Sums and Direct Sums	126
5.3	The Rank-Nullity Theorem; Grassmann's Relation	131
5.4	Affine Maps	137
5.5	Summary	144
5.6	Problems	145
6	Determinants	153
6.1	Permutations, Signature of a Permutation	153
6.2	Alternating Multilinear Maps	158
6.3	Definition of a Determinant	162
6.4	Inverse Matrices and Determinants	170
6.5	Systems of Linear Equations and Determinants	173
6.6	Determinant of a Linear Map	175
6.7	The Cayley–Hamilton Theorem	176
6.8	Permanents	181
6.9	Summary	183
6.10	Further Readings	185
6.11	Problems	185
7	Gaussian Elimination, LU, Cholesky, Echelon Form	191
7.1	Motivating Example: Curve Interpolation	191
7.2	Gaussian Elimination	195
7.3	Elementary Matrices and Row Operations	200
7.4	LU -Factorization	203
7.5	$PA = LU$ Factorization	209
7.6	Proof of Theorem 7.5 \otimes	217
7.7	Dealing with Roundoff Errors; Pivoting Strategies	223
7.8	Gaussian Elimination of Tridiagonal Matrices	224
7.9	SPD Matrices and the Cholesky Decomposition	226
7.10	Reduced Row Echelon Form	235
7.11	RREF, Free Variables, Homogeneous Systems	241
7.12	Uniqueness of RREF	244
7.13	Solving Linear Systems Using RREF	246
7.14	Elementary Matrices and Columns Operations	253
7.15	Transvections and Dilatations \otimes	254
7.16	Summary	259
7.17	Problems	261
8	Vector Norms and Matrix Norms	273
8.1	Normed Vector Spaces	273
8.2	Matrix Norms	284
8.3	Subordinate Norms	289
8.4	Inequalities Involving Subordinate Norms	296

8.5	Condition Numbers of Matrices	298
8.6	An Application of Norms: Inconsistent Linear Systems	307
8.7	Limits of Sequences and Series	308
8.8	The Matrix Exponential	311
8.9	Summary	314
8.10	Problems	316
9	Iterative Methods for Solving Linear Systems	323
9.1	Convergence of Sequences of Vectors and Matrices	323
9.2	Convergence of Iterative Methods	326
9.3	Methods of Jacobi, Gauss–Seidel, and Relaxation	328
9.4	Convergence of the Methods	336
9.5	Convergence Methods for Tridiagonal Matrices	339
9.6	Summary	343
9.7	Problems	344
10	The Dual Space and Duality	347
10.1	The Dual Space E^* and Linear Forms	347
10.2	Pairing and Duality Between E and E^*	354
10.3	The Duality Theorem and Some Consequences	359
10.4	The Bidual and Canonical Pairings	364
10.5	Hyperplanes and Linear Forms	366
10.6	Transpose of a Linear Map and of a Matrix	367
10.7	Properties of the Double Transpose	372
10.8	The Four Fundamental Subspaces	374
10.9	Summary	377
10.10	Problems	378
11	Euclidean Spaces	381
11.1	Inner Products, Euclidean Spaces	381
11.2	Orthogonality and Duality in Euclidean Spaces	390
11.3	Adjoint of a Linear Map	397
11.4	Existence and Construction of Orthonormal Bases	400
11.5	Linear Isometries (Orthogonal Transformations)	407
11.6	The Orthogonal Group, Orthogonal Matrices	410
11.7	The Rodrigues Formula	412
11.8	QR -Decomposition for Invertible Matrices	415
11.9	Some Applications of Euclidean Geometry	420
11.10	Summary	421
11.11	Problems	423
12	QR-Decomposition for Arbitrary Matrices	435
12.1	Orthogonal Reflections	435

12.2	QR -Decomposition Using Householder Matrices	440
12.3	Summary	450
12.4	Problems	450
13	Hermitian Spaces	457
13.1	Hermitian Spaces, Pre-Hilbert Spaces	457
13.2	Orthogonality, Duality, Adjoint of a Linear Map	466
13.3	Linear Isometries (Also Called Unitary Transformations)	471
13.4	The Unitary Group, Unitary Matrices	473
13.5	Hermitian Reflections and QR -Decomposition	476
13.6	Orthogonal Projections and Involutions	481
13.7	Dual Norms	484
13.8	Summary	491
13.9	Problems	492
14	Eigenvectors and Eigenvalues	497
14.1	Eigenvectors and Eigenvalues of a Linear Map	497
14.2	Reduction to Upper Triangular Form	505
14.3	Location of Eigenvalues	509
14.4	Conditioning of Eigenvalue Problems	512
14.5	Eigenvalues of the Matrix Exponential	515
14.6	Summary	517
14.7	Problems	518
15	Unit Quaternions and Rotations in $\mathbf{SO}(3)$	529
15.1	The Group $\mathbf{SU}(2)$ and the Skew Field \mathbb{H} of Quaternions	529
15.2	Representation of Rotation in $\mathbf{SO}(3)$ By Quaternions in $\mathbf{SU}(2)$	531
15.3	Matrix Representation of the Rotation r_q	536
15.4	An Algorithm to Find a Quaternion Representing a Rotation	538
15.5	The Exponential Map $\exp: \mathfrak{su}(2) \rightarrow \mathbf{SU}(2)$	541
15.6	Quaternion Interpolation \otimes	543
15.7	Nonexistence of a “Nice” Section from $\mathbf{SO}(3)$ to $\mathbf{SU}(2)$	545
15.8	Summary	547
15.9	Problems	548
16	Spectral Theorems	551
16.1	Introduction	551
16.2	Normal Linear Maps: Eigenvalues and Eigenvectors	551
16.3	Spectral Theorem for Normal Linear Maps	557
16.4	Self-Adjoint and Other Special Linear Maps	562
16.5	Normal and Other Special Matrices	568
16.6	Rayleigh–Ritz Theorems and Eigenvalue Interlacing	571
16.7	The Courant–Fischer Theorem; Perturbation Results	576

16.8	Summary	579
16.9	Problems	580
17	Computing Eigenvalues and Eigenvectors	587
17.1	The Basic QR Algorithm	589
17.2	Hessenberg Matrices	595
17.3	Making the QR Method More Efficient Using Shifts	601
17.4	Krylov Subspaces; Arnoldi Iteration	606
17.5	GMRES	610
17.6	The Hermitian Case; Lanczos Iteration	611
17.7	Power Methods	612
17.8	Summary	614
17.9	Problems	615
18	Graphs and Graph Laplacians; Basic Facts	617
18.1	Directed Graphs, Undirected Graphs, Weighted Graphs	620
18.2	Laplacian Matrices of Graphs	627
18.3	Normalized Laplacian Matrices of Graphs	631
18.4	Graph Clustering Using Normalized Cuts	635
18.5	Summary	637
18.6	Problems	638
19	Spectral Graph Drawing	641
19.1	Graph Drawing and Energy Minimization	641
19.2	Examples of Graph Drawings	644
19.3	Summary	648
20	Singular Value Decomposition and Polar Form	651
20.1	Properties of $f^* \circ f$	651
20.2	Singular Value Decomposition for Square Matrices	655
20.3	Polar Form for Square Matrices	658
20.4	Singular Value Decomposition for Rectangular Matrices	661
20.5	Ky Fan Norms and Schatten Norms	664
20.6	Summary	665
20.7	Problems	665
21	Applications of SVD and Pseudo-Inverses	669
21.1	Least Squares Problems and the Pseudo-Inverse	669
21.2	Properties of the Pseudo-Inverse	676
21.3	Data Compression and SVD	681
21.4	Principal Components Analysis (PCA)	683
21.5	Best Affine Approximation	694
21.6	Summary	698

21.7	Problems	699
22	Annihilating Polynomials; Primary Decomposition	703
22.1	Basic Properties of Polynomials; Ideals, GCD's	705
22.2	Annihilating Polynomials and the Minimal Polynomial	710
22.3	Minimal Polynomials of Diagonalizable Linear Maps	711
22.4	Commuting Families of Linear Maps	714
22.5	The Primary Decomposition Theorem	717
22.6	Jordan Decomposition	724
22.7	Nilpotent Linear Maps and Jordan Form	726
22.8	Summary	732
22.9	Problems	733
	Bibliography	735

Chapter 1

Introduction

As we explained in the preface, this first volume covers “classical” linear algebra, up to and including the primary decomposition and the Jordan form. Besides covering the standard topics, we discuss a few topics that are important for applications. These include:

1. Haar bases and the corresponding Haar wavelets, a fundamental tool in signal processing and computer graphics.
2. Hadamard matrices which have applications in error correcting codes, signal processing, and low rank approximation.
3. Affine maps (see Section 5.4). These are usually ignored or treated in a somewhat obscure fashion. Yet they play an important role in computer vision and robotics. There is a clean and elegant way to define affine maps. One simply has to define *affine combinations*. Linear maps preserve linear combinations, and similarly affine maps preserve affine combinations.
4. Norms and matrix norms (Chapter 8). These are used extensively in optimization theory.
5. Convergence of sequences and series in a normed vector space. Banach spaces (see Section 8.7). The matrix exponential e^A and its basic properties (see Section 8.8). In particular, we prove the Rodrigues formula for rotations in $\mathbf{SO}(3)$ and discuss the surjectivity of the exponential map $\exp: \mathfrak{so}(3) \rightarrow \mathbf{SO}(3)$, where $\mathfrak{so}(3)$ is the real vector space of 3×3 skew symmetric matrices (see Section 11.7). We also show that $\det(e^A) = e^{\text{tr}(A)}$ (see Section 14.5).
6. The group of unit quaternions, $\mathbf{SU}(2)$, and the representation of rotations in $\mathbf{SO}(3)$ by unit quaternions (Chapter 15). We define a homomorphism $r: \mathbf{SU}(2) \rightarrow \mathbf{SO}(3)$ and prove that it is surjective and that its kernel is $\{-I, I\}$. We compute the rotation matrix R_q associated with a unit quaternion q , and give an algorithm to construct a quaternion from a rotation matrix. We also show that the exponential map

$\exp: \mathfrak{su}(2) \rightarrow \mathbf{SU}(2)$ is surjective, where $\mathfrak{su}(2)$ is the real vector space of skew-Hermitian 2×2 matrices with zero trace. We discuss quaternion interpolation and prove the famous *slerp interpolation formula* due to Ken Shoemake.

7. An introduction to algebraic and spectral graph theory. We define the graph Laplacian and prove some of its basic properties (see Chapter 18). In Chapter 19, we explain how the eigenvectors of the graph Laplacian can be used for graph drawing.
8. Applications of SVD and pseudo-inverses, in particular, principal component analysis, for short PCA (Chapter 21).
9. Methods for computing eigenvalues and eigenvectors are discussed in Chapter 17. We first focus on the *QR* algorithm due to Rutishauser, Francis, and Kublanovskaya. See Sections 17.1 and 17.3. We then discuss how to use an *Arnoldi iteration*, in combination with the QR algorithm, to approximate eigenvalues for a matrix A of large dimension. See Section 17.4. The special case where A is a symmetric (or Hermitian) tridiagonal matrix, involves a *Lanczos iteration*, and is discussed in Section 17.6. In Section 17.7, we present power iterations and inverse (power) iterations.

Five topics are covered in more detail than usual. These are

1. Matrix factorizations such as LU , $PA = LU$, Cholesky, and reduced row echelon form (rref). Deciding the solvability of a linear system $Ax = b$, and describing the space of solutions when a solution exists. See Chapter 7.
2. Duality (Chapter 10).
3. Dual norms (Section 13.7).
4. The geometry of the orthogonal groups $\mathbf{O}(n)$ and $\mathbf{SO}(n)$, and of the unitary groups $\mathbf{U}(n)$ and $\mathbf{SU}(n)$.
5. The spectral theorems (Chapter 16).

Most texts omit the proof that the $PA = LU$ factorization can be obtained by a simple modification of Gaussian elimination. We give a complete proof of Theorem 7.5 in Section 7.6. We also prove the uniqueness of the rref of a matrix; see Proposition 7.19.

At the most basic level, duality corresponds to transposition. But duality is really the bijection between subspaces of a vector space E (say finite-dimensional) and subspaces of linear forms (subspaces of the dual space E^*) established by two maps: the first map assigns to a subspace V of E the subspace V^0 of linear forms that vanish on V ; the second map assigns to a subspace U of linear forms the subspace U^0 consisting of the vectors in E on which all linear forms in U vanish. The above maps define a bijection such that $\dim(V) + \dim(V^0) = \dim(E)$, $\dim(U) + \dim(U^0) = \dim(E)$, $V^{00} = V$, and $U^{00} = U$.

Another important fact is that if E is a finite-dimensional space with an inner product $u, v \mapsto \langle u, v \rangle$ (or a Hermitian inner product if E is a complex vector space), then there is a canonical isomorphism between E and its dual E^* . This means that every linear form $f \in E^*$ is uniquely represented by some vector $u \in E$, in the sense that $f(v) = \langle v, u \rangle$ for all $v \in E$. As a consequence, every linear map f has an adjoint f^* such that $\langle f(u), v \rangle = \langle u, f^*(v) \rangle$ for all $u, v \in E$.

Dual norms show up in convex optimization; see Boyd and Vandenberghe [11].

Because of their importance in robotics and computer vision, we discuss in some detail the groups of isometries $\mathbf{O}(E)$ and $\mathbf{SO}(E)$ of a vector space with an inner product. The isometries in $\mathbf{O}(E)$ are the linear maps such that $f \circ f^* = f^* \circ f = \text{id}$, and the direct isometries in $\mathbf{SO}(E)$, also called rotations, are the isometries in $\mathbf{O}(E)$ whose determinant is equal to $+1$. We also discuss the hermitian counterparts $\mathbf{U}(E)$ and $\mathbf{SU}(E)$.

We prove the spectral theorems not only for real symmetric matrices, but also for real and complex normal matrices.

We stress the importance of linear maps. Matrices are of course invaluable for computing and one needs to develop skills for manipulating them. But matrices are used to represent a linear map over a basis (or two bases), and the same linear map has different matrix representations. In fact, we can view the various normal forms of a matrix (Schur, SVD, Jordan) as a suitably convenient choice of bases.

We have listed most of the `Matlab` functions relevant to numerical linear algebra and have included `Matlab` programs implementing most of the algorithms discussed in this book.

