Quadratic Optimization Problems Arising in Computer Vision

Jean Gallier
Special Thanks to Jianbo Shi and Ryan Kennedy

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

March 23, 2011

Perverse Cohomology of Rhubarbs and the Stability of the Universe

The *stability of our universe* is clearly a fundamental problem.
Unfortunately, I proved

Perverse Cohomology of Rhubarbs and the Stability of the Universe

The *stability of our universe* is clearly a fundamental problem.
Unfortunately, I proved

Theorem 1

Our universe, U, is unstable.

Perverse Cohomology of Rhubarbs and the Stability of the Universe

The *stability of our universe* is clearly a fundamental problem.
Unfortunately, I proved

Theorem 1

Our universe, U, is unstable.

Proof.

Perverse Cohomology of Rhubarbs and the Stability of the Universe

The *stability of our universe* is clearly a fundamental problem. Unfortunately, I proved

Theorem 1

Our universe, U, is unstable.

Proof.

It can be shown that the perverse cohomology group

$$H^{257}_{rhub}(U) = 10^{10^{10^{10}}} \text{ rhubarbs.}$$

Perverse Cohomology of Rhubarbs and the Stability of the Universe

The *stability of our universe* is clearly a fundamental problem.
Unfortunately, I proved

Theorem 1

Our universe, U, is unstable.

Proof.

It can be shown that the perverse cohomology group

$$H^{257}_{rhub}(U) = 10^{10^{10^{10}}} \text{ rhubarbs.}$$

This is too big, therefore, the universe is unstable. □

Perverse Cohomology of Rhubarbs and the Stability of the Universe

The *stability of our universe* is clearly a fundamental problem. Unfortunately, I proved

Theorem 1

Our universe, U, is unstable.

Proof.

It can be shown that the perverse cohomology group

$$H^{257}_{rhub}(U) = 10^{10^{10^{10}}} \text{ rhubarbs.}$$

This is too big, therefore, the universe is unstable. □

It is obvious that Theorem 1 implies that $\mathrm{P} \neq \mathrm{NP}$.

1. Quadratic Optimization Problems; What Are They?

Many problems in computer vision, and more generally, computer science, can be cast as *optimization problems*.

1. Quadratic Optimization Problems; What Are They?

Many problems in computer vision, and more generally, computer science, can be cast as *optimization problems*.

Typically, one defines an *objective function*, f, whose domain is a subset of $\mathbb{R}^n$, and one wants to

$$\begin{aligned} \text{maximize} \quad & f(x) \\ \text{subject to constraints} \quad & g_1(x) = 0 \\ & g_2(x) \leq 0 \\ & \vdots \end{aligned}$$

1. Quadratic Optimization Problems; What Are They?

Many problems in computer vision, and more generally, computer science, can be cast as *optimization problems*.

Typically, one defines an *objective function*, f, whose domain is a subset of $\mathbb{R}^n$, and one wants to

$$\begin{aligned} \text{maximize} \quad & f(x) \\ \text{subject to constraints} \quad & g_1(x) = 0 \\ & g_2(x) \leq 0 \\ & \vdots \end{aligned}$$

The constaint functions, g_1, g_2, etc., are often linear or quadratic but they can be more complicated.

We will consider optimization problems where the optimization function, f, is *quadratic function* and the constaints are *quadratic or linear*.

We will consider optimization problems where the optimization function, f, is *quadratic function* and the constaints are *quadratic or linear*.

In general, a quadratric function is of the form

$$f(x) = x^\top Ax,$$

where $x \in \mathbb{R}^n$ and A is an $n \times n$ matrix.

We will consider optimization problems where the optimization function, f, is *quadratic function* and the constaints are *quadratic or linear*.

In general, a quadratic function is of the form

$$f(x) = x^\top A x,$$

where $x \in \mathbb{R}^n$ and A is an $n \times n$ matrix.

For example, we can express $f(x, y) = 5x^2 + 4xy + 2y^2$ in terms of a matrix as

$$f(x, y) = (x, y) \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

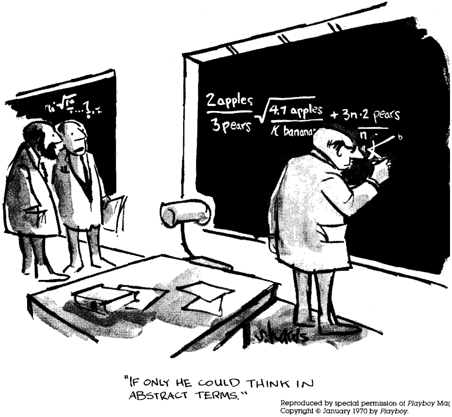

Figure: The power of abstraction

Important Fact 1.

We may assume that A is *symmetric*, which means that $A^\top = A$.

Important Fact 1.

We may assume that A is *symmetric*, which means that $A^\top = A$.

This is because we can write

$$A = H(A) + S(A),$$

where

$$H(A) = \frac{A + A^\top}{2} \quad \text{and} \quad S(A) = \frac{A - A^\top}{2}$$

Important Fact 1.

We may assume that A is *symmetric*, which means that $A^\top = A$.

This is because we can write

$$A = H(A) + S(A),$$

where

$$H(A) = \frac{A + A^\top}{2} \quad \text{and} \quad S(A) = \frac{A - A^\top}{2}$$

and $H(A)$ is *symmetric*, i.e., $H(A)^\top = H(A)$,

Important Fact 1.

We may assume that A is *symmetric*, which means that $A^\top = A$.

This is because we can write

$$A = H(A) + S(A),$$

where

$$H(A) = \frac{A + A^\top}{2} \quad \text{and} \quad S(A) = \frac{A - A^\top}{2}$$

and $H(A)$ is *symmetric*, i.e., $H(A)^\top = H(A)$,

$S(A)$ is *skew symmetric*, i.e., $S(A)^\top = -S(A)$.

If S is skew symmetric, it is easy to show that

$$x^{\top} S(A) x = 0,$$

If S is skew symmetric, it is easy to show that

$$x^\top S(A)x = 0,$$

so we get

$$f(x) = x^\top Ax = x^\top H(A)x.$$

If S is skew symmetric, it is easy to show that

$$x^{\top} S(A) x = 0,$$

so we get

$$f(x) = x^{\top} A x = x^{\top} H(A) x.$$

If A is a complex matrix, then we consider

$$A^{*} = (\overline{A})^{\top}$$

(the *transjugate*, *conjugate transpose* or *adjoint* of A)

We also have (replacing $A^\top$ by A^*)

$$A = H(A) + S(A)$$

We also have (replacing $A^\top$ by A^*)

$$A = H(A) + S(A)$$

where $H(A)$ is *Hermitian*, i.e., $H(A)^* = H(A)$,

We also have (replacing $A^\top$ by A^*)

$$A = H(A) + S(A)$$

where $H(A)$ is *Hermitian*, i.e., $H(A)^* = H(A)$,

and $S(A)$ is *skew Hermitian*, i.e., $S(A)^* = -S(A)$.

We also have (replacing $A^\top$ by A^*)

$$A = H(A) + S(A)$$

where $H(A)$ is *Hermitian*, i.e., $H(A)^* = H(A)$,

and $S(A)$ is *skew Hermitian*, i.e., $S(A)^* = -S(A)$.

Then, a quadratic function over $\mathbb{C}^n$ is of the form

$$f(x) = x^* A x,$$

with $x \in \mathbb{C}^n$.

If S is *skew Hermitian*, we have

$$(x^* S x)^* = -x^* S x,$$

If S is *skew Hermitian*, we have

$$(x^* S x)^* = -x^* S x,$$

but this only implies that the *real part* of $f(x)$ is zero that is, $f(x)$ is pure imaginary or zero.

If S is *skew Hermitian*, we have

$$(x^* S x)^* = -x^* S x,$$

but this only implies that the *real part* of $f(x)$ is zero that is, $f(x)$ is pure imaginary or zero.

However, if A *is* Hermitian, then $f(x) = x^* A x$, *is real*.

Important Fact 2.

Every $n \times n$ real symmetric matrix, A, has *real eigenvalues*, say

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

and can be *diagonalized* with respect to an *orthonormal basis* of *eigenvectors*.

Important Fact 2.

Every $n \times n$ real symmetric matrix, A, has *real eigenvalues*, say

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

and can be *diagonalized* with respect to an *orthonormal basis* of *eigenvectors*.

This means that there is a basis of orthonormal vectors, $(e_1, \ldots, e_n)$, where e_i is an *eigenvector* for λ_i, that is,

$$Ae_i = \lambda_i e_i, \qquad 1 \leq i \leq n.$$

Important Fact 2.

Every $n \times n$ real symmetric matrix, A, has *real eigenvalues*, say

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

and can be *diagonalized* with respect to an *orthonormal basis* of *eigenvectors*.

This means that there is a basis of orthonormal vectors, $(e_1, \ldots, e_n)$, where e_i is an *eigenvector* for λ_i, that is,

$$Ae_i = \lambda_i e_i, \qquad 1 \leq i \leq n.$$

The same result holds for (complex) Hermitian matrices (w.r.t. the Hermitian inner product).

The Basic Quadratic Optimization Problem

Our quadratic optimization problem is then to

$$\begin{array}{ll} \text{maximize} & x^\top A x \\ \text{subject to} & x^\top x = 1,\ x \in \mathbb{R}^n, \end{array}$$

where A is an $n \times n$ *symmetric* matrix.

The Basic Quadratic Optimization Problem

Our quadratic optimization problem is then to

$$\begin{aligned} \text{maximize} \quad & x^\top A x \\ \text{subject to} \quad & x^\top x = 1,\ x \in \mathbb{R}^n, \end{aligned}$$

where A is an $n \times n$ *symmetric* matrix.

If we diagonalize A w.r.t. an orthonormal basis of eigenvectors, $(e_1, \ldots, e_n)$, where

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

are the eigenvalues of A and if we write

$$x = x_1 e_1 + \cdots + x_n e_n,$$

then it is easy to see that

The Basic Quadratic Optimization Problem

Our quadratic optimization problem is then to

$$\begin{aligned} &\text{maximize} && x^\top A x \\ &\text{subject to} && x^\top x = 1,\ x \in \mathbb{R}^n, \end{aligned}$$

where A is an $n \times n$ *symmetric* matrix.

If we diagonalize A w.r.t. an orthonormal basis of eigenvectors, $(e_1, \ldots, e_n)$, where

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

are the eigenvalues of A and if we write

$$x = x_1 e_1 + \cdots + x_n e_n,$$

then it is easy to see that

$$f(x) = x^\top A x = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2,$$

subject to

$$x_1^1 + \cdots + x_n^2 = 1.$$

Since

$$x_1^1 + \cdots + x_n^2 = 1$$

and

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Since

$$x_1^1 + \cdots + x_n^2 = 1$$

and

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

we have

$$f(x) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2 \leq \lambda_1 (x_1^2 + \cdots + x_n^2) = \lambda_1$$

and

$$f(e_1) = \lambda_1.$$

Courant Fischer

Consequently

$$\max_{x^\top x=1} x^\top Ax = \lambda_1,$$

the *largest* eigenvalue of A, and this maximum is achieved for any *unit eigenvector* associated with λ_1.

Courant Fischer

Consequently

$$\max_{x^\top x=1} x^\top Ax = \lambda_1,$$

the *largest* eigenvalue of A, and this maximum is achieved for any *unit eigenvector* associated with λ_1.

This fact is part of the *Courant-Fischer Theorem*.

Figure: Richard Courant, 1888-1972

Figure: Richard Courant, 1888-1972

This result also holds for Hermitian matrices.

A Quadratic Optimization Problem Arising in Contour Grouping

Jianbo Shi and his students Qihui Zhu and Gang Song have investigated the problem of *contour grouping* in 2D images (ICCV 2007).

A Quadratic Optimization Problem Arising in Contour Grouping

Jianbo Shi and his students Qihui Zhu and Gang Song have investigated the problem of *contour grouping* in 2D images (ICCV 2007).

The problem is to find 1D (closed) curve-like structures in images.

A Quadratic Optimization Problem Arising in Contour Grouping

Jianbo Shi and his students Qihui Zhu and Gang Song have investigated the problem of *contour grouping* in 2D images (ICCV 2007).

The problem is to find 1D (closed) curve-like structures in images.

The goal is to find cycles linking small edges called *edgels*.

A Quadratic Optimization Problem Arising in Contour Grouping

Jianbo Shi and his students Qihui Zhu and Gang Song have investigated the problem of *contour grouping* in 2D images (ICCV 2007).

The problem is to find 1D (closed) curve-like structures in images.

The goal is to find cycles linking small edges called *edgels*.

The method uses a directed graph where the nodes are edgels and the edges connect pairs of edgels within some distance.

A Quadratic Optimization Problem Arising in Contour Grouping

Jianbo Shi and his students Qihui Zhu and Gang Song have investigated the problem of *contour grouping* in 2D images (ICCV 2007).

The problem is to find 1D (closed) curve-like structures in images.

The goal is to find cycles linking small edges called *edgels*.

The method uses a directed graph where the nodes are edgels and the edges connect pairs of edgels within some distance.

Every edge has a *weight*, W_{ij}, measuring the (directed) collinearity of two edgels using the elastic energy between these edgels.

Given a weighted directed graph, $G = (V, E, W)$, we seek a set of edges, $S \subseteq V$, (a *cut*) and an ordering, $\mathcal{O}$, on S, that maximizes a certain objective function,

Given a weighted directed graph, $G = (V, E, W)$, we seek a set of edges, $S \subseteq V$, (a *cut*) and an ordering, $\mathcal{O}$, on S, that maximizes a certain objective function,

$$C(S, \mathcal{O}, k) = \frac{1 - E_{\rm cut}(S) - I_{\rm cut}(S, \mathcal{O}, k)}{T(k)},$$

where

Given a weighted directed graph, $G = (V, E, W)$, we seek a set of edges, $S \subseteq V$, (a *cut*) and an ordering, $\mathcal{O}$, on S, that maximizes a certain objective function,

$$C(S, \mathcal{O}, k) = \frac{1 - E_{\mathrm{cut}}(S) - I_{\mathrm{cut}}(S, \mathcal{O}, k)}{T(k)},$$

where

1. $E_{\mathrm{cut}}(S)$ measures how strongly S is separated from its surrounding background (*external cut*)

Given a weighted directed graph, $G = (V, E, W)$, we seek a set of edges, $S \subseteq V$, (a *cut*) and an ordering, $\mathcal{O}$, on S, that maximizes a certain objective function,

$$C(S, \mathcal{O}, k) = \frac{1 - E_{\mathrm{cut}}(S) - I_{\mathrm{cut}}(S, \mathcal{O}, k)}{T(k)},$$

where

1. $E_{\mathrm{cut}}(S)$ measures how strongly S is separated from its surrounding background (*external cut*)
2. $I_{\mathrm{cut}}(S, \mathcal{O}, k)$ is a measure of the *entanglement* of the edges between the nodes in S (*internal cut*)

Given a weighted directed graph, $G = (V, E, W)$, we seek a set of edges, $S \subseteq V$, (a *cut*) and an ordering, $\mathcal{O}$, on S, that maximizes a certain objective function,

$$C(S, \mathcal{O}, k) = \frac{1 - E_{\mathrm{cut}}(S) - I_{\mathrm{cut}}(S, \mathcal{O}, k)}{T(k)},$$

where

1. $E_{\mathrm{cut}}(S)$ measures how strongly S is separated from its surrounding background (*external cut*)
2. $I_{\mathrm{cut}}(S, \mathcal{O}, k)$ is a measure of the *entanglement* of the edges between the nodes in S (*internal cut*)
3. $T(k)$ is the *tube size* of the cut; it depends on the *thickness factor*, k (in fact, $T(k) = k/|S|$).

Kennedy, Shi, and J.G. found a better formulation of the objective function involving a new normalization of the matrix arising from the graph G.

Kennedy, Shi, and J.G. found a better formulation of the objective function involving a new normalization of the matrix arising from the graph G.

We will only present the "old" formulation. The new formulation and new results are presented in a recent CVPR paper (2011):
Contour cuts: identifying salient contours in images by solving a Hermitian eigenvalue problem, R. Kennedy, J. Shi and J. Gallier.

Maximizing $C(S, \mathcal{O}, k)$ is a hard combinatorial problem so, Shi, Zhu and Song had the idea of converting the orginal problem to a simpler problem using a *circular embedding*.

Maximizing $C(S, \mathcal{O}, k)$ is a hard combinatorial problem so, Shi, Zhu and Song had the idea of converting the orginal problem to a simpler problem using a *circular embedding*.

The main idea is that a cycle is an image of the unit circle.

Maximizing $C(S, \mathcal{O}, k)$ is a hard combinatorial problem so, Shi, Zhu and Song had the idea of converting the orginal problem to a simpler problem using a *circular embedding*.

The main idea is that a cycle is an image of the unit circle.

Thus, we try to map the nodes of the graph onto the unit circle but nodes not in a cycle will be mapped to the origin.

Maximizing $C(S, \mathcal{O}, k)$ is a hard combinatorial problem so, Shi, Zhu and Song had the idea of converting the orginal problem to a simpler problem using a *circular embedding*.

The main idea is that a cycle is an image of the unit circle.

Thus, we try to map the nodes of the graph onto the unit circle but nodes not in a cycle will be mapped to the origin.

A point on the unit circle has coordinates

$$(\cos\theta, \sin\theta),$$

which are conveniently encoded as the complex number

$$z = \cos\theta + i\sin\theta = e^{i\theta}.$$

The nodes in a cycle will be mapped to the complex numbers

$$z_j = e^{i\theta_j}, \qquad \theta_j = \frac{2\pi j}{|S|}.$$

The nodes in a cycle will be mapped to the complex numbers

$$z_j = e^{i\theta_j}, \qquad \theta_j = \frac{2\pi j}{|S|}.$$

The *maximum jumping angle* $\theta_{\max}$ will also play a role; this is the maximum of the angle between two consecutive nodes.

Circular embedding score

Then, Shi and Zhu proved that maximizing $C(S, \mathcal{O}, k)$ is equivalent to maximizing the *circular embedding score*,

Circular embedding score

Then, Shi and Zhu proved that maximizing $C(S, \mathcal{O}, k)$ is equivalent to maximizing the *circular embedding score*,

$$C_e(r, \theta, \theta_{\max}) = \frac{1}{\theta_{\max}} \sum_{\substack{\theta_i < \theta_j \leq \theta_i + \theta_{\max} \\ r_i > 0,\, r_j > 0}} P_{ij}/|S|,$$

where

Circular embedding score

Then, Shi and Zhu proved that maximizing $C(S, \mathcal{O}, k)$ is equivalent to maximizing the *circular embedding score*,

$$C_e(r, \theta, \theta_{\max}) = \frac{1}{\theta_{\max}} \sum_{\substack{\theta_i < \theta_j \leq \theta_i + \theta_{\max} \\ r_i > 0,\, r_j > 0}} P_{ij}/|S|,$$

where

1. The matrix $P = (P_{ij})$ is obtained from the weight matrix, W, (of the graph $G = (V, E, W)$) by a suitable *normalization*

Circular embedding score

Then, Shi and Zhu proved that maximizing $C(S, \mathcal{O}, k)$ is equivalent to maximizing the *circular embedding score*,

$$C_e(r, \theta, \theta_{\max}) = \frac{1}{\theta_{\max}} \sum_{\substack{\theta_i < \theta_j \leq \theta_i + \theta_{\max} \\ r_i > 0,\, r_j > 0}} P_{ij} / |S|,$$

where

1. The matrix $P = (P_{ij})$ is obtained from the weight matrix, W, (of the graph $G = (V, E, W)$) by a suitable *normalization*
2. $r_j \in \{0, 1\}$

Circular embedding score

Then, Shi and Zhu proved that maximizing $C(S, \mathcal{O}, k)$ is equivalent to maximizing the *circular embedding score*,

$$C_e(r, \theta, \theta_{\max}) = \frac{1}{\theta_{\max}} \sum_{\substack{\theta_i < \theta_j \leq \theta_i + \theta_{\max} \\ r_i > 0,\, r_j > 0}} P_{ij}/|S|,$$

where

1. The matrix $P = (P_{ij})$ is obtained from the weight matrix, W, (of the graph $G = (V, E, W)$) by a suitable *normalization*
2. $r_j \in \{0, 1\}$
3. θ_j is an angle specifying the ordering of the nodes in the cycle

Circular embedding score

Then, Shi and Zhu proved that maximizing $C(S, \mathcal{O}, k)$ is equivalent to maximizing the *circular embedding score*,

$$C_e(r, \theta, \theta_{\max}) = \frac{1}{\theta_{\max}} \sum_{\substack{\theta_i < \theta_j \leq \theta_i + \theta_{\max} \\ r_i > 0,\ r_j > 0}} P_{ij}/|S|,$$

where

1. The matrix $P = (P_{ij})$ is obtained from the weight matrix, W, (of the graph $G = (V, E, W)$) by a suitable *normalization*
2. $r_j \in \{0, 1\}$
3. θ_j is an angle specifying the ordering of the nodes in the cycle
4. $\theta_{\max}$ is the maximum jumping angle.

This optimization problem is still hard to solve.

This optimization problem is still hard to solve. Consequently, Shi and Zhu considered a *continuous relaxation* of the probem by allowing r_j to be any real in the interval $[0, 1]$ and θ_j to be any angle (within a suitable range).

This optimization problem is still hard to solve. Consequently, Shi and Zhu considered a *continuous relaxation* of the probem by allowing r_j to be any real in the interval $[0, 1]$ and θ_j to be any angle (within a suitable range).

In the circular embedding, a node in then represented by the complex number

$$x_j = r_j e^{i\theta_j}.$$

We also introduce the *average jumping angle*

$$\Delta\theta = \overline{\theta_k - \theta_j}.$$

This optimization problem is still hard to solve. Consequently, Shi and Zhu considered a *continuous relaxation* of the probem by allowing r_j to be any real in the interval $[0, 1]$ and θ_j to be any angle (within a suitable range).

In the circular embedding, a node in then represented by the complex number

$$x_j = r_j e^{i\theta_j}.$$

We also introduce the *average jumping angle*

$$\Delta\theta = \overline{\theta_k - \theta_j}.$$

Then, it is not hard to see that the numerator of $C_e(r, \theta, \theta_{\mathrm{max}})$ is well approximated by the expression

$$\sum_{j,k} P_{jk} \cos(\theta_k - \theta_j - \Delta\theta) = \sum_{j,k} \mathrm{Re}(x_j^* x_k \cdot e^{-i\Delta\theta}).$$

Continuous Relaxation

Thus, $C_e(r, \theta, \theta_{\max})$ is well approximated by

$$\frac{1}{\theta_{\max}} \frac{\sum_{j,k} \mathrm{Re}(x_j^* x_k \cdot e^{-i\Delta\theta})}{\sum_j |x_j|^2}.$$

Continuous Relaxation

Thus, $C_e(r, \theta, \theta_{\max})$ is well approximated by

$$\frac{1}{\theta_{\max}} \frac{\sum_{j,k} \mathrm{Re}(x_j^* x_k \cdot e^{-i\Delta\theta})}{\sum_j |x_j|^2}.$$

This term can be written in terms of the matrix P as

$$C_e(r, \theta, \theta_{\max}) \approx \frac{1}{\theta_{\max}} \frac{\mathrm{Re}(x^* P x \cdot e^{-i\Delta\theta})}{x^* x},$$

where $x \in \mathbb{C}^n$ is the vector $x = (x_1, \ldots, x_n)$.

The matrix P is a real matrix but, in general, it not symmetric nor normal ($PP^* = P^*P$).

The matrix P is a real matrix but, in general, it not symmetric nor normal ($PP^* = P^*P$).

If we write $\delta = \Delta\theta$ and if we assume that $0 < \delta_{\min} \leq \delta \leq \delta_{\max}$, we would like to solve the following optimization problem:

The matrix P is a real matrix but, in general, it not symmetric nor normal ($PP^* = P^* P$).

If we write $\delta = \Delta\theta$ and if we assume that $0 < \delta_{\min} \leq \delta \leq \delta_{\max}$, we would like to solve the following optimization problem:

$$\begin{array}{ll} \text{maximize} & \mathrm{Re}(x^* e^{-i\delta} Px) \\ \text{subject to} & x^* x = 1,\ x \in \mathbb{C}^n; \\ & \delta_{\min} \leq \delta \leq \delta_{\max}. \end{array}$$

Zhu then further relaxed this problem to the problem:

$$\begin{aligned} &\text{maximize} && \mathrm{Re}(x^* e^{-i\delta} P y) \\ &\text{subject to} && x^* y = c,\ x, y \in \mathbb{C}^n; \\ &&& \delta_{\min} \leq \delta \leq \delta_{\max}. \end{aligned}$$

with $c = e^{-i\delta}$.

Zhu then further relaxed this problem to the problem:

$$\begin{aligned} &\text{maximize} && \mathrm{Re}(x^* e^{-i\delta} P y) \\ &\text{subject to} && x^* y = c,\ x, y \in \mathbb{C}^n; \\ & && \delta_{\min} \leq \delta \leq \delta_{\max}. \end{aligned}$$

with $c = e^{-i\delta}$.

However, it turns out that this problem is *too relaxed*, because the constraint $x^* y = c$ is weak; it allows x to be *very large* and y to be *very small*, and conversely.

However, this relaxation in unnecessary.

However, this relaxation in unnecessary.
Indeed, for any complex number, $z = x + iy$,

$$\mathrm{Re}(z) = x = \frac{z + \overline{z}}{2},$$

However, this relaxation in unnecessary.
Indeed, for any complex number, $z = x + iy$,

$$\mathrm{Re}(z) = x = \frac{z + \overline{z}}{2},$$

and a calculation shows that

$$\mathrm{Re}(x^* \, e^{-i\delta} P \, x) = x^* \frac{1}{2}(e^{-i\delta} P + e^{i\delta} P^\top)x.$$

However, this relaxation in unnecessary.
Indeed, for any complex number, $z = x + iy$,

$$\mathrm{Re}(z) = x = \frac{z + \overline{z}}{2},$$

and a calculation shows that

$$\mathrm{Re}(x^* \, e^{-i\delta} P \, x) = x^* \frac{1}{2}(e^{-i\delta} P + e^{i\delta} P^\top) x.$$

Note that

$$H(e^{-i\delta} P) = \frac{1}{2}(e^{-i\delta} P + e^{i\delta} P^\top)$$

is the *Hermitian part* of $e^{-i\delta} P$.

A New Formulation of the Optimization Problem

Another simple calculation shows that

$$H(e^{-i\delta}P) = \cos\delta\, H(P) - i\sin\delta\, S(P).$$

A New Formulation of the Optimization Problem

Another simple calculation shows that

$$H(e^{-i\delta}P) = \cos\delta\, H(P) - i\sin\delta\, S(P).$$

In view of the above, our original (relaxed) optimization problem can be stated as

$$\begin{aligned} &\text{maximize} && x^* H(\delta)\, x \\ &\text{subject to} && x^* x = 1,\ x \in \mathbb{C}^n; \\ & && \delta_{\min} \leq \delta \leq \delta_{\max} \end{aligned}$$

A New Formulation of the Optimization Problem

Another simple calculation shows that

$$H(e^{-i\delta}P) = \cos\delta\, H(P) - i\sin\delta\, S(P).$$

In view of the above, our original (relaxed) optimization problem can be stated as

$$\begin{aligned} &\text{maximize} && x^* H(\delta)\, x \\ &\text{subject to} && x^* x = 1,\ x \in \mathbb{C}^n; \\ & && \delta_{\min} \leq \delta \leq \delta_{\max} \end{aligned}$$

with

$$H(\delta) = H(e^{-i\delta}P) = \cos\delta\, H(P) - i\sin\delta\, S(P),$$

a *Hermitian matrix*.

The optimal value is the *largest eigenvalue*, λ_1, of $H(\delta)$, over all δ such that $\delta_{\min} \leq \delta \leq \delta_{\max}$ and it is attained for any associated complex unit eigenvector, $x = x_r + ix_i$.

The optimal value is the *largest eigenvalue*, λ_1, of $H(\delta)$, over all δ such that $\delta_{\min} \leq \delta \leq \delta_{\max}$ and it is attained for any associated complex unit eigenvector, $x = x_r + ix_i$.

Ryan Kennedy has implemented this method and has obtained good results.

The Case Where P is a Normal Matrix

When P is a normal matrix ($PP^\top = P^\top P$) it is possible to express the eigenvalues of $H(\delta)$ and the corresponding eigenvectors in terms of the (complex) eigenvalues of P and its eigenvectors.

The Case Where P is a Normal Matrix

When P is a normal matrix ($PP^\top = P^\top P$) it is possible to express the eigenvalues of $H(\delta)$ and the corresponding eigenvectors in terms of the (complex) eigenvalues of P and its eigenvectors.

If $u + iv$ is an eigenvector of P for the (complex) eigenvalue $\lambda + i\mu$, then $u + iv$ is also an eigenvector of $H(\delta)$ for the (real) eigenvalue $\cos\delta\,\lambda - \sin\delta\,\mu$.

The Case Where P is a Normal Matrix

When P is a normal matrix ($PP^{\top} = P^{\top}P$) it is possible to express the eigenvalues of $H(\delta)$ and the corresponding eigenvectors in terms of the (complex) eigenvalues of P and its eigenvectors.

If $u + iv$ is an eigenvector of P for the (complex) eigenvalue $\lambda + i\mu$, then $u + iv$ is also an eigenvector of $H(\delta)$ for the (real) eigenvalue $\cos\delta\,\lambda - \sin\delta\,\mu$.

Geometrically, this means that the eigenvalues of $H(\delta)$ vary on circles, plotted as a function of δ.

The Case Where P is a Normal Matrix

When P is a normal matrix ($PP^\top = P^\top P$) it is possible to express the eigenvalues of $H(\delta)$ and the corresponding eigenvectors in terms of the (complex) eigenvalues of P and its eigenvectors.

If $u + iv$ is an eigenvector of P for the (complex) eigenvalue $\lambda + i\mu$, then $u + iv$ is also an eigenvector of $H(\delta)$ for the (real) eigenvalue $\cos\delta\,\lambda - \sin\delta\,\mu$.

Geometrically, this means that the eigenvalues of $H(\delta)$ vary on circles, plotted as a function of δ.

The next four Figures were produced by Ryan Kennedy.

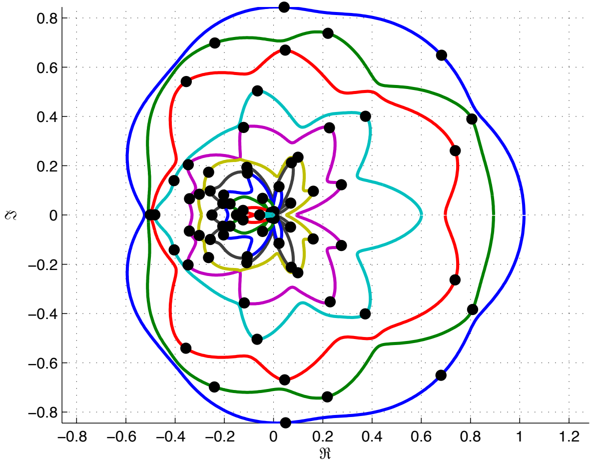

Figure: The eigenvalues of a matrix $H(\delta)$ which is not normal

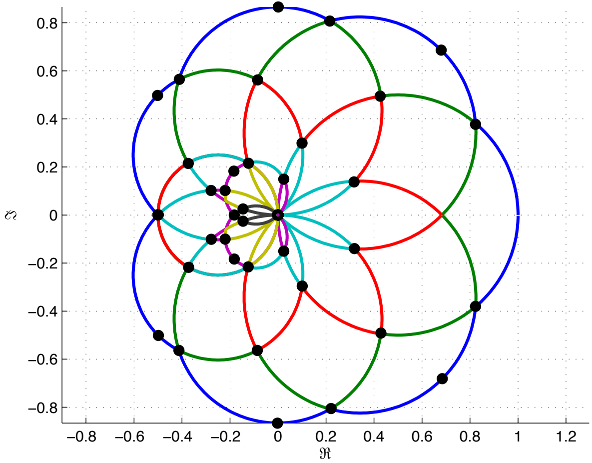

Figure: The eigenvalues of a normal matrix $H(\delta)$

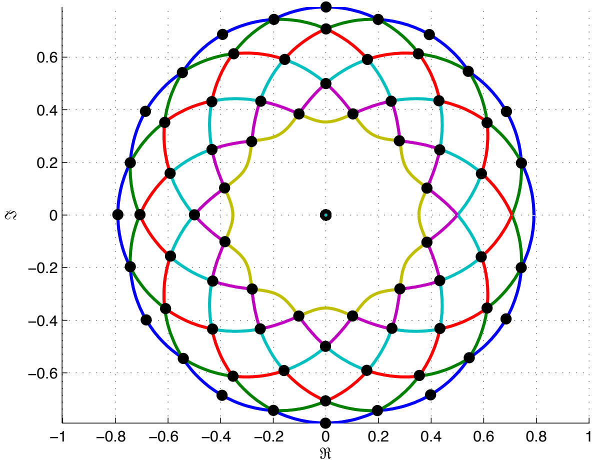

Figure: The eigenvalues of a matrix $H(\delta)$ which is near normal

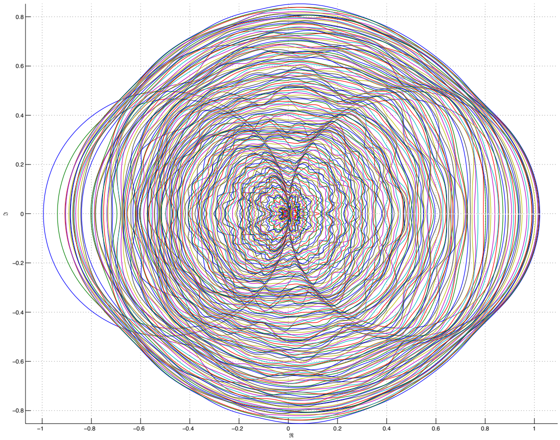

Figure: The eigenvalues of the matrix for an actual image

Derivatives of Eigenvectors and Eigenvalues

To solve our maximization problem, we need to study the variation of the largest eigenvalue, $\lambda_1(\delta)$, of $H(\delta)$.

Derivatives of Eigenvectors and Eigenvalues

To solve our maximization problem, we need to study the variation of the largest eigenvalue, $\lambda_1(\delta)$, of $H(\delta)$.

This problem has been studied before and it is possible to find explicit formulae for the derivative of a simple eigenvalue of $H(\delta)$ and for the derivative of a unit eigenvector of $H(\delta)$.

Shi and Cour obtained similar formulae in a different context.

Derivatives of Eigenvectors and Eigenvalues

To solve our maximization problem, we need to study the variation of the largest eigenvalue, $\lambda_1(\delta)$, of $H(\delta)$.

This problem has been studied before and it is possible to find explicit formulae for the derivative of a simple eigenvalue of $H(\delta)$ and for the derivative of a unit eigenvector of $H(\delta)$.

Shi and Cour obtained similar formulae in a different context.

It turns out that it is not easy to find clean and complete derivations of these formulae.

The best source is Peter Lax's linear algebra book (Chapter 9). A nice account is also found in a blog by Terence Tao.

Let $X(\delta)$ be a matrix function depending on the parameter δ.

It is proved in Lax (Chapter 9, Theorem 7 and Theorem 8) that if λ is a *simple* eigenvalue of $X(\delta)$, for $\delta = \delta_0$ and if u is a unit eigenvector associated with λ, then, in a small open interval around δ_0, the matrix $X(\delta)$ has a simple eigenvalue, $\lambda(\delta)$, that is differentiable (with $\lambda(\delta_0) = \lambda$) and that there is a choice of an eigenvector, $u(t)$, associated with $\lambda(t)$, so that $u(t)$ is also differentiable (with $u(\delta_0) = u$).

Let $X(\delta)$ be a matrix function depending on the parameter δ.

It is proved in Lax (Chapter 9, Theorem 7 and Theorem 8) that if λ is a *simple* eigenvalue of $X(\delta)$, for $\delta = \delta_0$ and if u is a unit eigenvector associated with λ, then, in a small open interval around δ_0, the matrix $X(\delta)$ has a simple eigenvalue, $\lambda(\delta)$, that is differentiable (with $\lambda(\delta_0) = \lambda$) and that there is a choice of an eigenvector, $u(t)$, associated with $\lambda(t)$, so that $u(t)$ is also differentiable (with $u(\delta_0) = u$).

In the case of an eigenvalue, the proof uses the implicit function theorem applied to the characteristic polynomial, $\det(\lambda I - X(\delta))$.

Let $X(\delta)$ be a matrix function depending on the parameter δ.

It is proved in Lax (Chapter 9, Theorem 7 and Theorem 8) that if λ is a *simple* eigenvalue of $X(\delta)$, for $\delta = \delta_0$ and if u is a unit eigenvector associated with λ, then, in a small open interval around δ_0, the matrix $X(\delta)$ has a simple eigenvalue, $\lambda(\delta)$, that is differentiable (with $\lambda(\delta_0) = \lambda$) and that there is a choice of an eigenvector, $u(t)$, associated with $\lambda(t)$, so that $u(t)$ is also differentiable (with $u(\delta_0) = u$).

In the case of an eigenvalue, the proof uses the implicit function theorem applied to the characteristic polynomial, $\det(\lambda I - X(\delta))$.

The proof of differentiability for an eigenvector is more involved and uses the non-vanishing of some principal minor of $\det(\lambda I - X(\delta))$.

The formula for the derivative of an eigenvector is simpler if we assume $X(\delta)$ to be normal. In this case, we get

The formula for the derivative of an eigenvector is simpler if we assume $X(\delta)$ to be normal. In this case, we get

Theorem 2

Let $X(\delta)$ be a normal matrix that depends differentiably on δ. If λ is any simple eigenvalue of X at δ_0 (it has algebraic multiplicity 1) and if u is the corresponding unit eigenvector, then the derivatives at $\delta = \delta_0$ of $\lambda(\delta)$ and $u(\delta)$ are given by

$$\lambda' = u^* X' u$$
$$u' = (\lambda I - X)^{\dagger} X' u,$$

where $(\lambda I - X)^{\dagger}$ is the pseudo-inverse of $\lambda I - X$, X' is the derivative of X at $\delta = \delta_0$ and u' is orthogonal to u.

Proof.

If X is a normal matrix, it is well known that $Xu = \lambda u$ iff $X^* u = \overline{\lambda} u$ and so, if $Xu = \lambda u$ then

$$u^* X = \lambda u^*.$$

Proof.

If X is a normal matrix, it is well known that $Xu = \lambda u$ iff $X^* u = \overline{\lambda} u$ and so, if $Xu = \lambda u$ then

$$u^* X = \lambda u^*.$$

Taking the derivative of $Xu = \lambda u$ and using the chain rule, we get

$$X' u + X u' = \lambda' u + \lambda u'.$$

Proof.

If X is a normal matrix, it is well known that $Xu = \lambda u$ iff $X^* u = \overline{\lambda} u$ and so, if $Xu = \lambda u$ then

$$u^* X = \lambda u^*.$$

Taking the derivative of $Xu = \lambda u$ and using the chain rule, we get

$$X' u + X u' = \lambda' u + \lambda u'.$$

By taking the inner product with u^*, we get

$$u^* X' u + u^* X u' = \lambda' u^* u + \lambda u^* u'.$$

Proof.

If X is a normal matrix, it is well known that $Xu = \lambda u$ iff $X^* u = \overline{\lambda} u$ and so, if $Xu = \lambda u$ then

$$u^* X = \lambda u^*.$$

Taking the derivative of $Xu = \lambda u$ and using the chain rule, we get

$$X' u + X u' = \lambda' u + \lambda u'.$$

By taking the inner product with u^*, we get

$$u^* X' u + u^* X u' = \lambda' u^* u + \lambda u^* u'.$$

However, $u^* X = \lambda u^*$, so $u^* X u' = \lambda u^* u'$, and as u is a unit vector, $u^* u = 1$,

Proof.

If X is a normal matrix, it is well known that $Xu = \lambda u$ iff $X^* u = \overline{\lambda} u$ and so, if $Xu = \lambda u$ then

$$u^* X = \lambda u^*.$$

Taking the derivative of $Xu = \lambda u$ and using the chain rule, we get

$$X' u + X u' = \lambda' u + \lambda u'.$$

By taking the inner product with u^*, we get

$$u^* X' u + u^* X u' = \lambda' u^* u + \lambda u^* u'.$$

However, $u^* X = \lambda u^*$, so $u^* X u' = \lambda u^* u'$, and as u is a unit vector, $u^* u = 1$, so

$$u^* X' u + \lambda u^* u' = \lambda' + \lambda u^* u',$$

that is, $\lambda' = u^* X' u$. □

Proof.

If X is a normal matrix, it is well known that $Xu = \lambda u$ iff $X^* u = \overline{\lambda} u$ and so, if $Xu = \lambda u$ then

$$u^* X = \lambda u^*.$$

Taking the derivative of $Xu = \lambda u$ and using the chain rule, we get

$$X' u + X u' = \lambda' u + \lambda u'.$$

By taking the inner product with u^*, we get

$$u^* X' u + u^* X u' = \lambda' u^* u + \lambda u^* u'.$$

However, $u^* X = \lambda u^*$, so $u^* X u' = \lambda u^* u'$, and as u is a unit vector, $u^* u = 1$, so

$$u^* X' u + \lambda u^* u' = \lambda' + \lambda u^* u',$$

that is, $\lambda' = u^* X' u$. □

Deriving the formula for the derivative of u is more involved.

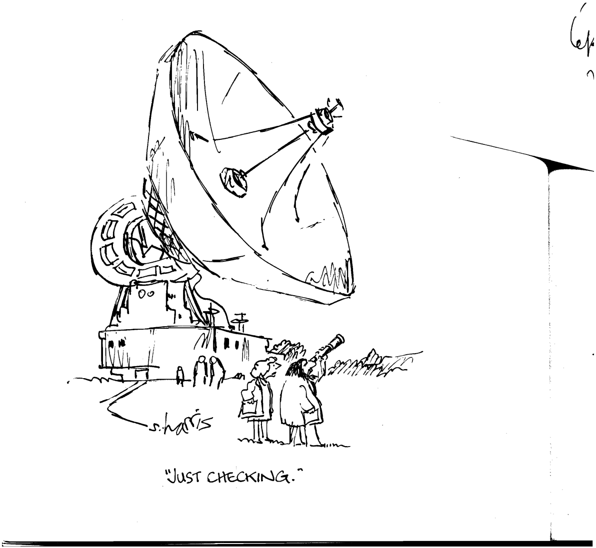

Figure: Just checking!

The Field of Values of P

It turns out that

$$x^* H(\delta) x \leq |x^* P x|$$

for all x and all δ, and this has some important implications regarding the local maxima of these two functions.

The Field of Values of P

It turns out that

$$x^* H(\delta)x \leq |x^* Px|$$

for all x and all δ, and this has some important implications regarding the local maxima of these two functions.

In fact, if we write $x^* Px$ in polar form as

$$x^* Px = |x^* Px|(\cos\varphi + i\sin\varphi),$$

I proved that

$$x^* H(\delta)x = |x^* Px| \cos(\delta - \varphi).$$

This implies that

$$x^* H(\delta) x \leq |x^* P x|$$

for all $x \in \mathbb{C}^n$ and all δ, $(0 \leq \delta \leq 2\pi)$, with equality iff

$$\delta = \varphi,$$

the argument (phase angle) of $x^* P x$.

This implies that

$$x^* H(\delta) x \leq |x^* P x|$$

for all $x \in \mathbb{C}^n$ and all δ, $(0 \leq \delta \leq 2\pi)$, with equality iff

$$\delta = \varphi,$$

the argument (phase angle) of $x^* P x$.

In particular, for x fixed, $f(x, \delta) = x^* H x$ has a local optimum when $\delta = \varphi$ and, in this case, $x^* H x = |x^* P x|$.

The inequality $x^* H x \leq |x^* P x|$ also implies that *if $|x^* P x|$ achieves a local maximum for some vector, x, then $f(x, \delta) = x^* H x$ achieves a local maximum equal to $|x^* P x|$ for $\delta = \varphi$ and for the same x* (where φ is the argument of $x^* P x$).

The inequality $x^* Hx \leq |x^* Px|$ also implies that *if $|x^* Px|$ achieves a local maximum for some vector, x, then $f(x, \delta) = x^* Hx$ achieves a local maximum equal to $|x^* Px|$ for $\delta = \varphi$ and for the same x* (where φ is the argument of $x^* Px$).

Furthermore, x must be an eigenvector of $H(\varphi)$.

The inequality $x^* Hx \leq |x^* Px|$ also implies that *if* $|x^* Px|$ *achieves a local maximum for some vector,* x*, then* $f(x, \delta) = x^* Hx$ *achieves a local maximum equal to* $|x^* Px|$ *for* $\delta = \varphi$ *and for the same* x (where φ is the argument of $x^* Px$).

Furthermore, x must be an eigenvector of $H(\varphi)$.

Generally, if $f(x, \delta) = x^* Hx$ is a local maximum of f at (x, δ), then $|x^* Px|$ is *not* necessarily a local maximum at x.

The inequality $x^* H x \leq |x^* P x|$ also implies that *if $|x^* P x|$ achieves a local maximum for some vector, x, then $f(x, \delta) = x^* H x$ achieves a local maximum equal to $|x^* P x|$ for $\delta = \varphi$ and for the same x* (where φ is the argument of $x^* P x$).

Furthermore, x must be an eigenvector of $H(\varphi)$.

Generally, if $f(x, \delta) = x^* H x$ is a local maximum of f at (x, δ), then $|x^* P x|$ is *not* necessarily a local maximum at x.

Still, since the maxima of $|x^* P x|$ dominate the maxima of $x^* H(\delta) x$, and are a subset of those maxima, it is useful to understand better how to find the local maxima of $|x^* P x|$.

The determination of the local extrema of $|x^*Px|$ (with $x^*x = 1$) is closely related to the structure of the set of complex numbers

$$F(P) = \{x^*Px \in \mathbb{C} \mid x \in \mathbb{C}^n,\, x^*x = 1\},$$

known as the *field of values* of P or the *numerical range* of P (the notation $W(P)$ is also commonly used, corresponding to the German terminology "Wertvorrat" or "Wertevorrat").

The determination of the local extrema of $|x^* P x|$ (with $x^* x = 1$) is closely related to the structure of the set of complex numbers

$$F(P) = \{x^* P x \in \mathbb{C} \mid x \in \mathbb{C}^n,\, x^* x = 1\},$$

known as the *field of values* of P or the *numerical range* of P (the notation $W(P)$ is also commonly used, corresponding to the German terminology "Wertvorrat" or "Wertevorrat").

This set was studied as early as 1918 by Toeplitz and Hausdorff who proved that $F(P)$ is *convex*.

Figure: Felix Hausdorff, 1868-1942 (left) and Otto Toeplitz, 1881-1940 (right)

The next three Figures were produced by Ryan Kennedy.

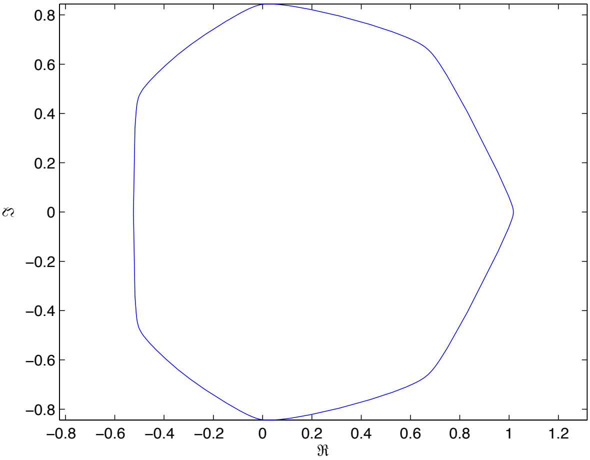

Figure: Numerical Range of a matrix which is not normal

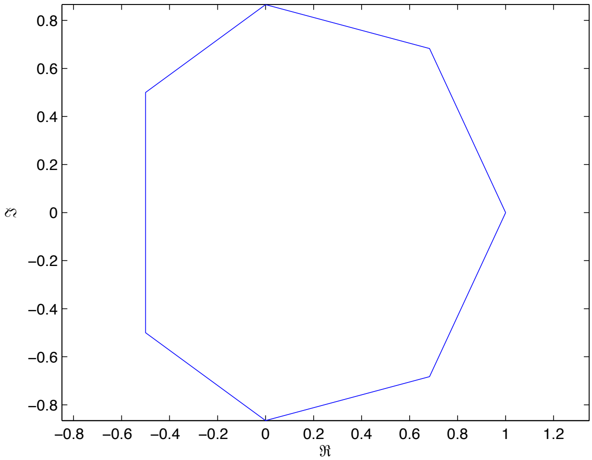

Figure: Numerical Range of a normal matrix

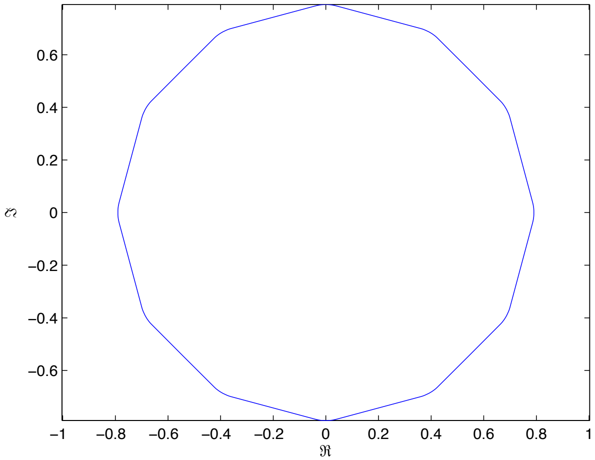

Figure: Numerical Range of a matrix which is near normal

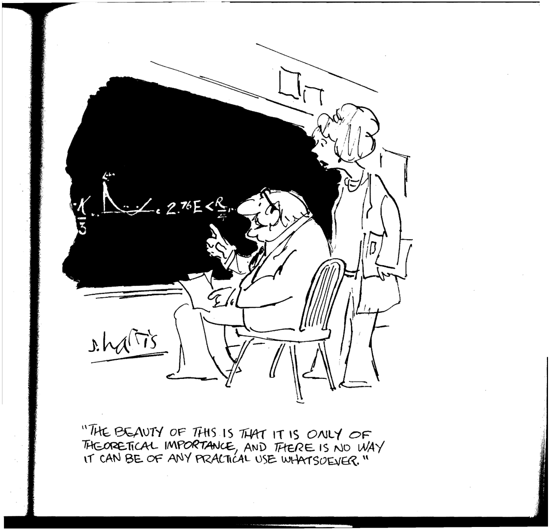

Figure: Beauty

It is easy to show that

$$F(e^{-i\delta}P) = e^{-i\delta}F(P)$$

and so,

$$F(P) = e^{i\delta}F(e^{-i\delta}P).$$

It is easy to show that

$$F(e^{-i\delta}P) = e^{-i\delta}F(P)$$

and so,

$$F(P) = e^{i\delta}F(e^{-i\delta}P).$$

Geometrically, this means that $F(P)$ is obtained from $F(e^{-i\delta}P)$ by rotating it by δ.

It is easy to show that

$$F(e^{-i\delta}P) = e^{-i\delta}F(P)$$

and so,

$$F(P) = e^{i\delta}F(e^{-i\delta}P).$$

Geometrically, this means that $F(P)$ is obtained from $F(e^{-i\delta}P)$ by rotating it by δ.

This fact yields a nice way of finding supporting lines for the convex set, $F(P)$.

Figure: Images with top 20 contours extracted

Adding Affine Constraints to the Basic Problem

Gander, Golub and von Matt (1989) considered the following problem: Given an $(n+m)\times(n+m)$ real symmetric matrix, A, (with $n > 0$), an $(n+m)\times m$ matrix, N, with full rank and a nonzero vector, $t \in \mathbb{R}^m$, with $\left\|(N^\top)^\dagger t\right\| < 1$

Adding Affine Constraints to the Basic Problem

Gander, Golub and von Matt (1989) considered the following problem: Given an $(n+m)\times(n+m)$ real symmetric matrix, A, (with $n > 0$), an $(n+m)\times m$ matrix, N, with full rank and a nonzero vector, $t \in \mathbb{R}^m$, with $\left\|(N^\top)^\dagger t\right\| < 1$

$$
\begin{aligned}
\text{minimize} \quad & x^\top A x \\
\text{subject to} \quad & x^\top x = 1, \quad x \in \mathbb{R}^{n+m} \\
& N^\top x = t.
\end{aligned}
$$

The condition $\left\|(N^\top)^\dagger t\right\| < 1$ ensures that the problem has a solution and is not trivial.

The condition $\left\|(N^\top)^\dagger t\right\| < 1$ ensures that the problem has a solution and is not trivial.

It can be shown that the affine constraints, $N^\top x = t$, can be eliminated, but *a linear term needs to be added to the objective function*.

The condition $\left\|(N^\top)^\dagger t\right\| < 1$ ensures that the problem has a solution and is not trivial.

It can be shown that the affine constraints, $N^\top x = t$, can be eliminated, but *a linear term needs to be added to the objective function*.

One way to do so is to use a QR decomposition of N.

If

$$N = P\begin{pmatrix} R \\ 0 \end{pmatrix}$$

where P is an orthogonal matrix and R is an $m \times m$ *invertible upper triangular matrix*,

If

$$N = P \begin{pmatrix} R \\ 0 \end{pmatrix}$$

where P is an orthogonal matrix and R is an $m \times m$ *invertible upper triangular matrix*, then we get the simplified problem

$$\begin{aligned} \text{minimize} \quad & z^\top C z + 2 z^\top b \\ \text{subject to} \quad & z^\top z = s^2, \quad z \in \mathbb{R}^m, \end{aligned}$$

where C is a block in the matrix $P^\top A P$.

If

$$N = P\begin{pmatrix} R \\ 0 \end{pmatrix}$$

where P is an orthogonal matrix and R is an $m \times m$ *invertible upper triangular matrix*, then we get the simplified problem

$$\begin{array}{ll} \text{minimize} & z^\top C z + 2 z^\top b \\ \text{subject to} & z^\top z = s^2, \quad z \in \mathbb{R}^m, \end{array}$$

where C is a block in the matrix $P^\top A P$.

This problem was studied by Gander, Golub and von Matt (1989).

If

$$N = P\begin{pmatrix} R \\ 0 \end{pmatrix}$$

where P is an orthogonal matrix and R is an $m \times m$ *invertible upper triangular matrix*, then we get the simplified problem

$$\begin{aligned} &\text{minimize} && z^\top C z + 2z^\top b \\ &\text{subject to} && z^\top z = s^2, \quad z \in \mathbb{R}^m, \end{aligned}$$

where C is a block in the matrix $P^\top A P$.

This problem was studied by Gander, Golub and von Matt (1989).

I also have a solution to this problem involving an *algebraic curve generalizing the hyperbola to* $\mathbb{R}^n$, which will be presented next.

Quadratic Optimization with an Affine Quadratic Function

We now focus on the following problem:

Quadratic Optimization with an Affine Quadratic Function

We now focus on the following problem:

If A is a real $n \times n$ symmetric matrix and $b \in \mathbb{R}$ is any vector,

$$\begin{aligned} &\text{maximize} && x^\top Ax + 2x^\top b \\ &\text{subject to} && x^\top x = 1, \quad x \in \mathbb{R}^n, \end{aligned}$$

where $b \neq 0$.

Quadratic Optimization with an Affine Quadratic Function

We now focus on the following problem:

If A is a real $n \times n$ symmetric matrix and $b \in \mathbb{R}$ is any vector,

$$\begin{array}{ll} \text{maximize} & x^\top A x + 2x^\top b \\ \text{subject to} & x^\top x = 1, \quad x \in \mathbb{R}^n, \end{array}$$

where $b \neq 0$.

This time, we can't proceed algebraically directly, so we will use a *Lagrangian*.

Finding Extrema Using Lagrangians and Lagrange Multipliers

We know from calculus that if $f\colon \Omega \to \mathbb{R}$ is a real-valued function defined on an *open* subset, Ω, of $\mathbb{R}^n$ and if f has an *extremum* at $x \in \Omega$ and f is differentiable at x, then

$$f'(x) = 0,$$

where f' is the *derivative* of f or, equivalently,

$$(\mathrm{grad}\, f)(x) = 0.$$

Finding Extrema Using Lagrangians and Lagrange Multipliers

We know from calculus that if $f\colon \Omega \to \mathbb{R}$ is a real-valued function defined on an *open* subset, Ω, of $\mathbb{R}^n$ and if f has an *extremum* at $x \in \Omega$ and f is differentiable at x, then

$$f'(x) = 0,$$

where f' is the *derivative* of f or, equivalently,

$$(\mathrm{grad}\, f)(x) = 0.$$

However, in our situation, the function f is defined on the sphere

$$S^{n-1} = \{x \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 = 1\},$$

which is *not open* and, in fact, is closed!

Thus, the vanishing of the derivative of f is *not* a necessary condition for an extremum!

Thus, the vanishing of the derivative of f is *not* a necessary condition for an extremum!

For example, it *fails* for the function $f(x) = x$ on $[0, 1]$.

Thus, the vanishing of the derivative of f is *not* a necessary condition for an extremum!

For example, it *fails* for the function $f(x) = x$ on $[0, 1]$.

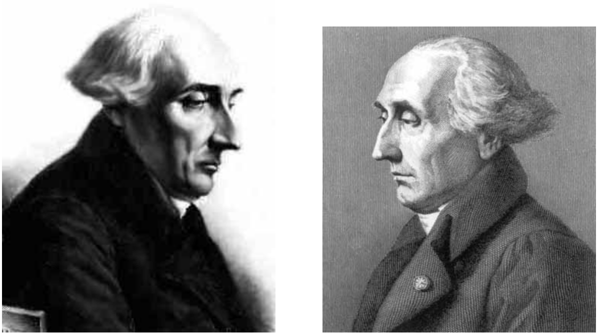

Figure: Joseph-Louis Lagrange, 1736-1813

Thus, the vanishing of the derivative of f is *not* a necessary condition for an extremum!

For example, it *fails* for the function $f(x) = x$ on $[0, 1]$.

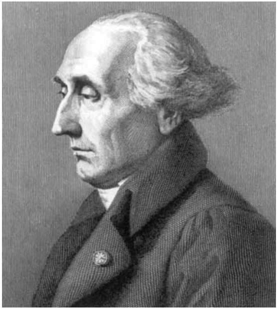

Figure: Joseph-Louis Lagrange, 1736-1813

Fortunately, Lagrange found a way to tackle this problem.

Lagrangians

The trick is to deal with the *constraints* defining the domain of f by forming the *Lagrangian* of the problem:

$$L(x, \lambda) = x^\top Ax - \lambda(x^\top x - 1),$$

where the scalar, $\lambda \in \mathbb{R}$, is called a *Lagrange multiplier*.

Lagrangians

The trick is to deal with the *constraints* defining the domain of f by forming the *Lagrangian* of the problem:

$$L(x, \lambda) = x^\top A x - \lambda(x^\top x - 1),$$

where the scalar, $\lambda \in \mathbb{R}$, is called a *Lagrange multiplier*.

Then, it can be shown that if $x \in S^{n-1}$ is an extremum of f, then for some $\lambda \in \mathbb{R}$,

$$(\mathrm{grad}\, L)(x, \lambda) = 0,$$

Lagrangians

The trick is to deal with the *constraints* defining the domain of f by forming the *Lagrangian* of the problem:

$$L(x, \lambda) = x^\top A x - \lambda(x^\top x - 1),$$

where the scalar, $\lambda \in \mathbb{R}$, is called a *Lagrange multiplier*.

Then, it can be shown that if $x \in S^{n-1}$ is an extremum of f, then for some $\lambda \in \mathbb{R}$,

$$(\mathrm{grad}\, L)(x, \lambda) = 0,$$

Thus, we must have

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial \lambda} = 0.$$

Now,

$$\frac{\partial L}{\partial x} = Ax - \lambda x, \quad \frac{\partial L}{\partial \lambda} = x^\top x - 1,$$

Now,

$$\frac{\partial L}{\partial x} = Ax - \lambda x, \quad \frac{\partial L}{\partial \lambda} = x^\top x - 1,$$

so we get the *necessary* conditions:

$$\begin{aligned} Ax &= \lambda x \\ x^\top x &= 1, \end{aligned}$$

Now,

$$\frac{\partial L}{\partial x} = Ax - \lambda x, \quad \frac{\partial L}{\partial \lambda} = x^\top x - 1,$$

so we get the *necessary* conditions:

$$\begin{aligned} Ax &= \lambda x \\ x^\top x &= 1, \end{aligned}$$

which are equivalent to finding some eigenvalue, λ, of A and a corresponding unit eigenvector, x, as before!

Now,

$$\frac{\partial L}{\partial x} = Ax - \lambda x, \quad \frac{\partial L}{\partial \lambda} = x^\top x - 1,$$

so we get the *necessary* conditions:

$$\begin{aligned} Ax &= \lambda x \\ x^\top x &= 1, \end{aligned}$$

which are equivalent to finding some eigenvalue, λ, of A and a corresponding unit eigenvector, x, as before!

Warning: The vanishing of the Lagrangian is only a *necessary* condition for an extremum.

To find out whether (x, λ) corresponds to a maximum, a *local study* of the behavior of f near x is *required*.

To find out whether (x, λ) corresponds to a maximum, a *local study* of the behavior of f near x is *required*.

In our earlier study, we showed that f has a global maximum for any unit eigenvector associated with the *largest* eigenvalue of A.

To find out whether (x, λ) corresponds to a maximum, a *local study* of the behavior of f near x is *required*.

In our earlier study, we showed that f has a global maximum for any unit eigenvector associated with the *largest* eigenvalue of A.

The Lagrangian also vanishes for the other eigenvalues of A, but if they are smaller, they do not yield a maximum!

To find out whether (x, λ) corresponds to a maximum, a *local study* of the behavior of f near x is *required*.

In our earlier study, we showed that f has a global maximum for any unit eigenvector associated with the *largest* eigenvalue of A.

The Lagrangian also vanishes for the other eigenvalues of A, but if they are smaller, they do not yield a maximum!

Let us now figure out the Lagrangian of our problem involving an affine objective function.

The Lagrangian, $L(x, \lambda)$, of our problem, is

The Lagrangian, $L(x, \lambda)$, of our problem, is

$$L(x, \lambda) = x^\top A x + 2x^\top b - \lambda(x^\top x - 1).$$

The Lagrangian, $L(x, \lambda)$, of our problem, is

$$L(x, \lambda) = x^\top Ax + 2x^\top b - \lambda(x^\top x - 1).$$

We know that a *necessary condition* for the function, $f(x) = x^\top Ax + 2x^\top b$, to have a *local extremum* on the unit sphere, is that $L(x, \lambda)$ has a *critical point*, which means that

The Lagrangian, $L(x, \lambda)$, of our problem, is

$$L(x, \lambda) = x^\top Ax + 2x^\top b - \lambda(x^\top x - 1).$$

We know that a *necessary condition* for the function, $f(x) = x^\top Ax + 2x^\top b$, to have a *local extremum* on the unit sphere, is that $L(x, \lambda)$ has a *critical point*, which means that

$$\frac{\partial L}{\partial x} = 0, \qquad \frac{\partial L}{\partial \lambda} = 0.$$

Since

$$\frac{\partial L}{\partial x} = 2Ax + 2b - 2\lambda x, \qquad \frac{\partial L}{\partial \lambda} = x^\top x - 1,$$

necessary conditions for f to have a local extremum are

$$\begin{aligned} (\lambda I - A)x &= b \\ x^\top x &= 1. \end{aligned}$$

Since

$$\frac{\partial L}{\partial x} = 2Ax + 2b - 2\lambda x, \qquad \frac{\partial L}{\partial \lambda} = x^\top x - 1,$$

necessary conditions for f to have a local extremum are

$$\begin{aligned} (\lambda I - A)x &= b \\ x^\top x &= 1. \end{aligned}$$

Recall that that $b \neq 0$. Since A is a symmetric matrix, it can be diagonalized and we can write

$$A = Q^\top \Sigma Q,$$

where Σ is a (real) diagonal matrix and Q is an orthogonal matrix.

Substituting the righthand side of A into our system, we get

$$\begin{aligned} Q^\top(\lambda I - \Sigma)Qx &= b \\ x^\top x &= 1, \end{aligned}$$

which yields

$$\begin{aligned} (\lambda I - \Sigma)Qx &= Qb \\ (Qx)^\top Qx &= 1. \end{aligned}$$

If we let $c = Qb$ and $y = Qx$, the above system becomes

$$\begin{aligned} (\lambda I - \Sigma)y &= c \\ y^\top y &= 1, \end{aligned}$$

where Σ is a *diagonal matrix*.

If we let $c = Qb$ and $y = Qx$, the above system becomes

$$\begin{aligned} (\lambda I - \Sigma)y &= c \\ y^\top y &= 1, \end{aligned}$$

where Σ is a *diagonal matrix*.

The solutions of the original system

$$\begin{aligned} (\lambda I - A)x &= b \\ x^\top x &= 1 \end{aligned}$$

are obtained using the equation $x = Q^\top y$.

Solution in the Generic Case

Let us first assume that the eigenvalues of A are *all distinct* and order them in decreasing order so that $\sigma_1 > \sigma_2 > \cdots > \sigma_n$.

Solution in the Generic Case

Let us first assume that the eigenvalues of A are *all distinct* and order them in decreasing order so that $\sigma_1 > \sigma_2 > \cdots > \sigma_n$.

The system

$$(\lambda I - \Sigma)y = c$$

defines a *parametric curve*, $C(\Sigma, c)$, in $\mathbb{R}^n$, for all $\lambda \neq \sigma_i$, $1 \leq i \leq n$, where the ith coordinate of a point on the curve is given by

$$y_i(\lambda) = \frac{c_i}{\lambda - \sigma_i}.$$

Solution in the Generic Case

Let us first assume that the eigenvalues of A are *all distinct* and order them in decreasing order so that $\sigma_1 > \sigma_2 > \cdots > \sigma_n$.

The system

$$(\lambda I - \Sigma)y = c$$

defines a *parametric curve*, $C(\Sigma, c)$, in $\mathbb{R}^n$, for all $\lambda \neq \sigma_i$, $1 \leq i \leq n$, where the ith coordinate of a point on the curve is given by

$$y_i(\lambda) = \frac{c_i}{\lambda - \sigma_i}.$$

If $c_i \neq 0$, for $i = 1, \ldots, n$, then $y_i(\lambda) \longrightarrow \pm\infty$ when $\lambda \longrightarrow \sigma_i$ and note that $y \longrightarrow 0$ when $\lambda \longrightarrow \pm\infty$.

In this case, the *solutions* of the system

$$\begin{aligned}(\lambda I - \Sigma)y &= c \\ y^\top y &= 1\end{aligned}$$

are the *points of intersection* of the curve, $C(\Sigma, c)$, with the unit sphere, $y^\top y = 1$.

In this case, the *solutions* of the system

$$\begin{aligned}(\lambda I - \Sigma)y &= c \\ y^\top y &= 1\end{aligned}$$

are the *points of intersection* of the curve, $C(\Sigma, c)$, with the unit sphere, $y^\top y = 1$.

The (connected) branch of the curve, $C(\Sigma, c)$, for which $\lambda \in (-\infty, \sigma_n) \cup (\sigma_1, +\infty)$ always intersects the unit sphere, since it passes through the origin for $\lambda = \pm\infty$.

When $\lambda \longrightarrow \sigma_n$ from $-\infty$, the line parallel to the y_n-axis for which

$$y_1 = \frac{c_1}{\sigma_n - \sigma_1}, \ldots, y_{n-1} = \frac{c_n}{\sigma_n - \sigma_{n-1}}$$

is an asymptote and

When $\lambda \longrightarrow \sigma_n$ from $-\infty$, the line parallel to the y_n-axis for which

$$y_1 = \frac{c_1}{\sigma_n - \sigma_1}, \ldots, y_{n-1} = \frac{c_n}{\sigma_n - \sigma_{n-1}}$$

is an asymptote and when $\lambda \longrightarrow \sigma_1$ from $+\infty$, the line parallel to the y_1-axis for which

$$y_2 = \frac{c_2}{\sigma_1 - \sigma_2}, \ldots, y_n = \frac{c_{n-1}}{\sigma_1 - \sigma_n}$$

is another asymptote.

When $\lambda \longrightarrow \sigma_n$ from $-\infty$, the line parallel to the y_n-axis for which

$$y_1 = \frac{c_1}{\sigma_n - \sigma_1}, \ldots, y_{n-1} = \frac{c_n}{\sigma_n - \sigma_{n-1}}$$

is an asymptote and when $\lambda \longrightarrow \sigma_1$ from $+\infty$, the line parallel to the y_1-axis for which

$$y_2 = \frac{c_2}{\sigma_1 - \sigma_2}, \ldots, y_n = \frac{c_{n-1}}{\sigma_1 - \sigma_n}$$

is another asymptote.

The curve, $C(\Sigma, c)$, has $n-1$ other connected branches, one for each interval (σ_i, σ_{i-1}), where $i = n, \ldots, 2$, and these branches also have asymptotes.

If $c_i = 0$ for some i, the situation is more sutble.

If $c_i = 0$ for some i, the situation is more sutble.

Let us consider the case $n = 2$.

If $c_i = 0$ for some i, the situation is more sutble.

Let us consider the case $n = 2$.

When $n = 2$, we have the system of equations

$$\begin{aligned}(\lambda - \sigma_1)y_1 &= c_1\\ (\lambda - \sigma_2)y_2 &= c_2\\ y_1^2 + y_2^2 &= 1.\end{aligned}$$

If $c_1 = 0$, then, as $c_2 \neq 0$, the two linear equations have a solution iff $\lambda \neq \sigma_2$.

If $c_1 = 0$, then, as $c_2 \neq 0$, the two linear equations have a solution iff $\lambda \neq \sigma_2$.

Case 1. If (y_1, y_2) is a solution of the system

$$
\begin{aligned}
(\lambda - \sigma_1)y_1 &= 0 \\
(\lambda - \sigma_2)y_2 &= c_2
\end{aligned}
$$

with $y_1 = 0$, then, this system defines the line of equation $y_1 = 0$.

If $c_1 = 0$, then, as $c_2 \neq 0$, the two linear equations have a solution iff $\lambda \neq \sigma_2$.

Case 1. If (y_1, y_2) is a solution of the system

$$\begin{aligned} (\lambda - \sigma_1) y_1 &= 0 \\ (\lambda - \sigma_2) y_2 &= c_2 \end{aligned}$$

with $y_1 = 0$, then, this system defines the line of equation $y_1 = 0$.

This line intersects the unit circle $y_1^2 + y_2^2 = 1$ for $y_2 = \pm 1$.

If $c_1 = 0$, then, as $c_2 \neq 0$, the two linear equations have a solution iff $\lambda \neq \sigma_2$.

Case 1. If (y_1, y_2) is a solution of the system

$$\begin{aligned} (\lambda - \sigma_1) y_1 &= 0 \\ (\lambda - \sigma_2) y_2 &= c_2 \end{aligned}$$

with $y_1 = 0$, then, this system defines the line of equation $y_1 = 0$.

This line intersects the unit circle $y_1^2 + y_2^2 = 1$ for $y_2 = \pm 1$.

Since $c_2 \neq 0$, our system has the two solutions $(y_1, y_2) = (0, \pm 1)$ for $\lambda = \sigma_2 \pm c_2$.

Case 2. If (y_1, y_2) *with* $y_1 \neq 0$ is a solution of the system

$$\begin{aligned}(\lambda - \sigma_1)y_1 &= 0\\(\lambda - \sigma_2)y_2 &= c_2\end{aligned}$$

then we must have $\lambda = \sigma_1$.

Case 2. If (y_1, y_2) *with* $y_1 \neq 0$ is a solution of the system

$$\begin{aligned} (\lambda - \sigma_1) y_1 &= 0 \\ (\lambda - \sigma_2) y_2 &= c_2 \end{aligned}$$

then we must have $\lambda = \sigma_1$.

In this case, the above system reduces to the single equation

$$(\sigma_1 - \sigma_2) y_2 = c_2$$

which defines the line of equation

$$y_2 = \frac{c_2}{\sigma_1 - \sigma_2}.$$

This line intersects the unit circle $y_1^2 + y_2^2 = 1$ iff

$$y_1^2 = 1 - \frac{c_2^2}{(\sigma_1 - \sigma_2)^2}.$$

This line intersects the unit circle $y_1^2 + y_2^2 = 1$ iff

$$y_1^2 = 1 - \frac{c_2^2}{(\sigma_1 - \sigma_2)^2}.$$

This equation has real nonzero solutions iff

$$c_2^2 < (\sigma_1 - \sigma_2)^2$$

and if so, the solutions to our system are

$$y_1 = \pm\sqrt{1 - y_2^2}, \qquad y_2 = \frac{c_2}{\sigma_1 - \sigma_2}.$$

This line intersects the unit circle $y_1^2 + y_2^2 = 1$ iff

$$y_1^2 = 1 - \frac{c_2^2}{(\sigma_1 - \sigma_2)^2}.$$

This equation has real nonzero solutions iff

$$c_2^2 < (\sigma_1 - \sigma_2)^2$$

and if so, the solutions to our system are

$$y_1 = \pm\sqrt{1 - y_2^2}, \qquad y_2 = \frac{c_2}{\sigma_1 - \sigma_2}.$$

In summary, when $c_1 = 0$, $(y_1, y_2) = (0, \pm 1)$ are solutions and there are possibly two extra solutions if $\lambda = \sigma_1$ and $c_2^2 < (\sigma_1 - \sigma_2)^2$.

The case where $c_2 = 0$ is similar. We find that $(y_1, y_2) = (\pm 1, 0)$ are solutions and there are possibly two extra solutions if $\lambda = \sigma_2$ and $c_1^2 < (\sigma_2 - \sigma_1)^2$.

The case where $c_2 = 0$ is similar. We find that $(y_1, y_2) = (\pm 1, 0)$ are solutions and there are possibly two extra solutions if $\lambda = \sigma_2$ and $c_1^2 < (\sigma_2 - \sigma_1)^2$.

Case 3. If $c_1 \neq 0$ and $c_2 \neq 0$, by solving for λ in terms of y_1, we get

$$\lambda = \frac{c_1}{y_1} + \sigma_1$$

and by substituting in the second equation we get

$$y_2 = \frac{c_2 y_1}{c_1 + (\sigma_1 - \sigma_2) y_1}.$$

This is the equation of a *hyperbola* passing through the origin and with two asymptotes parallel to the y_1 and the y_2 axes, namely,

$$y_1 = -\frac{c_1}{\sigma_1 - \sigma_2}$$

and

$$y_2 = \frac{c_2}{\sigma_1 - \sigma_2}.$$

This is the equation of a *hyperbola* passing through the origin and with two asymptotes parallel to the y_1 and the y_2 axes, namely,

$$y_1 = -\frac{c_1}{\sigma_1 - \sigma_2}$$

and

$$y_2 = \frac{c_2}{\sigma_1 - \sigma_2}.$$

The branch of the hyperbola passing through the origin intersects the unit circle, $y_1^2 + y_2^2 = 1$, in two points and, in general, the other branch of the hyperbola also intersects the unit circle in two points as illustrated in the next Figure.

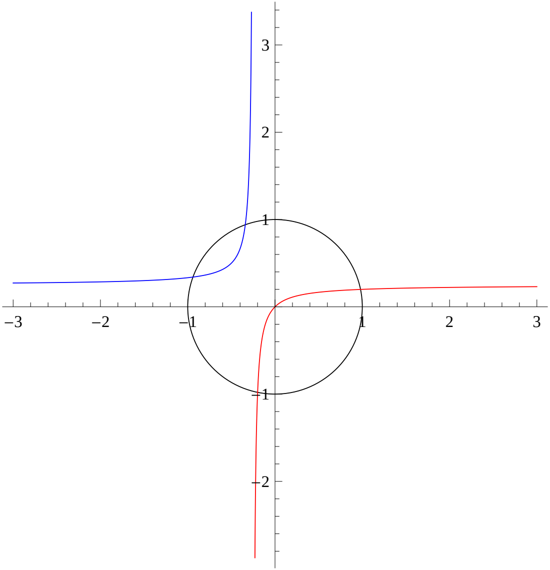

Figure: Intersections of $C(\Sigma, c)$ (a hyperbola) with the unit circle

Therefore, in general, the hyperbola intersects the unit circle in four points and always in at least two points. The corresponding values of λ are given by the equation

$$\frac{c_1^2}{(\lambda - \sigma_1)^2} + \frac{c_2^2}{(\lambda - \sigma_2)^2} = 1,$$

which yields a polynomial equation of degree 4.

Therefore, in general, the hyperbola intersects the unit circle in four points and always in at least two points. The corresponding values of λ are given by the equation

$$\frac{c_1^2}{(\lambda - \sigma_1)^2} + \frac{c_2^2}{(\lambda - \sigma_2)^2} = 1,$$

which yields a polynomial equation of degree 4.

In the general case, $n \geq 2$, we have the following theorem:

Theorem 3

If the eigenvalues of the $n \times n$ symmetric matrix, A, are all distinct, then there are $2m$ values of λ, say $\lambda_1 > \lambda_2 \geq \lambda_3 > \cdots > \lambda_{2m-2} \geq \lambda_{2m-1} > \lambda_{2m}$, with $1 \leq m \leq n$, such that the system

$$\begin{aligned}(\lambda I - A)x &= b \\ x^\top x &= 1\end{aligned}$$

(with $b \neq 0$) has a solution, (λ, x). As a consequence, the Lagrangian,

$$L(x, \lambda) = x^\top A x + 2x^\top b - \lambda(x^\top x - 1),$$

has at least two and at most $2n$ critical point, (x, λ).

Theorem (continued)

Furthermore, the eigenvalues,
$\sigma_1 > \sigma_2 > \cdots > \sigma_n$*, of* A *separate the* λ*'s, which means that*

1. $\lambda_1 \geq \sigma_1$
2. $\lambda_{2m} \leq \sigma_n$
3. *For every* λ_i*, with* $2 \leq i \leq 2m-1$*, either* $\lambda_i = \sigma_j$ *for some* j *with* $1 \leq j \leq n$*, or there is some* j*, with* $1 \leq j \leq n-1$*, so that* $\sigma_j > \lambda_i > \sigma_{j+1}$*.*

If $c_i \neq 0$ for $i = 1, \ldots, n$, then the curve, $C(\Sigma, c)$, is a kind of generalized hyperbola in $\mathbb{R}^n$, with n asymptotes corresponding the the values $\lambda = \sigma_i$.

If $c_i \neq 0$ for $i = 1, \ldots, n$, then the curve, $C(\Sigma, c)$, is a kind of generalized hyperbola in $\mathbb{R}^n$, with n asymptotes corresponding the the values $\lambda = \sigma_i$.

An example of this curve in shown for $n = 3$ in the next Figure.

If $c_i \neq 0$ for $i = 1, \ldots, n$, then the curve, $C(\Sigma, c)$, is a kind of generalized hyperbola in $\mathbb{R}^n$, with n asymptotes corresponding the the values $\lambda = \sigma_i$.

An example of this curve in shown for $n = 3$ in the next Figure.

In order for some, y, on the curve $C(\Sigma, c)$ to belong to the unit sphere, the equation

$$\sum_{i=1}^{n} \frac{c_i^2}{(\lambda - \sigma_i)^2} = 1,$$

must hold, which yields a polynomial equation of degree $2n$.

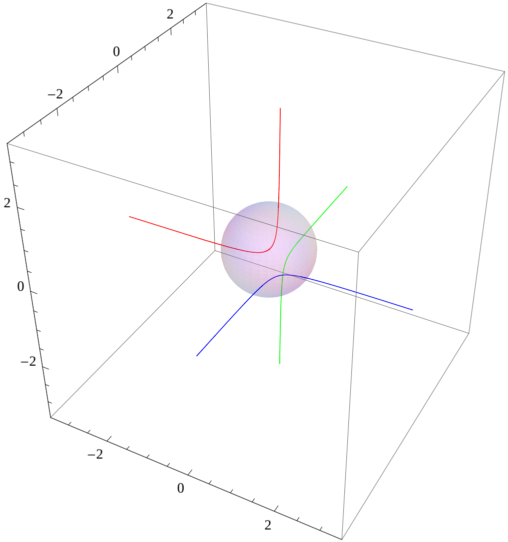

Figure: Intersections of $C(\Sigma, c)$ with the unit sphere ($n = 3$)

Solution in the General Case (Multiple Eigenvalues)

Fortunately, when the matrix, A, has multiple eigenvalues, Theorem 4 can still be proved pretty much as before except for some notational complications.

Solution in the General Case (Multiple Eigenvalues)

Fortunately, when the matrix, A, has multiple eigenvalues, Theorem 4 can still be proved pretty much as before except for some notational complications.

Theorem 4

If the $n \times n$ symmetric matrix, A, has p distinct eigenvalues, $\sigma_1 > \sigma_2 > \cdots > \sigma_p$, each with multiplicity $k_i \geq 1$, with $k_1 + \cdots + k_p = n$, then there are $2m$ values of λ, say $\lambda_1 > \lambda_2 \geq \lambda_3 > \cdots > \lambda_{2m-2} \geq \lambda_{2m-1} > \lambda_{2m}$, with $1 \leq m \leq p$, such that the system

$$\begin{aligned} (\lambda I - A)x &= b \\ x^\top x &= 1 \end{aligned}$$

(with $b \neq 0$) has a solution, (λ, x).

Theorem (continued)

As a consequence, there are at least two and at most $2p$ values of λ for which the Lagrangian,

$$L(x, \lambda) = x^\top Ax + 2x^\top b - \lambda(x^\top x - 1),$$

has a critical point, (x, λ), but there may be infinitely many x for which (x, λ) is a critical point. Furthermore, the distinct eigenvalues, $\sigma_1 > \sigma_2 > \cdots > \sigma_p$, of A separate the λ's, which means that

1. $\lambda_1 \geq \sigma_1$
2. $\lambda_{2m} \leq \sigma_p$
3. *For every λ_i, with $2 \leq i \leq 2m-1$, either $\lambda_i = \sigma_j$ for some j with $1 \leq j \leq p$, or there is some j, with $1 \leq j \leq p-1$, so that $\sigma_j > \lambda_i > \sigma_{j+1}$.*