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 Potatoes and the Stability of the Universe

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

Perverse Cohomology of Potatoes 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 Potatoes 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 Potatoes 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

$$H^{237}_{pot}(U) = 10^{10^{10}} \text{ potatoes.}$$

Perverse Cohomology of Potatoes 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

$$H^{237}_{pot}(U) = 10^{10^{10}} \text{ potatoes.}$$

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

Perverse Cohomology of Potatoes 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

$$H^{237}_{pot}(U) = 10^{10^{10}} \text{ potatoes.}$$

This is too big, therefore, the universe is unstable. $\square$

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

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

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

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

We usually want to find the *maximum value* of f (subject to the constaints) and *find some/all values of x for which $f(x)$ is maximal* (an "argmax" problem).

We usually want to find the *maximum value* of f (subject to the constaints) and *find some/all values of x for which $f(x)$ is maximal* (an "argmax" problem).

Sometimes, one wants to *minimize* $f(x)$, but this is equivalent to *maximizing* $-f(x)$.

We usually want to find the *maximum value* of f (subject to the constaints) and *find some/all values of x for which $f(x)$ is maximal* (an "argmax" problem).

Sometimes, one wants to *minimize* $f(x)$, but this is equivalent to *maximizing* $-f(x)$.

Sometimes the domain of f is a *discrete* subset of $\mathbb{R}^n$, for example, $\{0, 1\}^n$.

We usually want to find the *maximum value* of f (subject to the constaints) and *find some/all values of x for which $f(x)$ is maximal* (an "argmax" problem).

Sometimes, one wants to *minimize* $f(x)$, but this is equivalent to *maximizing* $-f(x)$.

Sometimes the domain of f is a *discrete* subset of $\mathbb{R}^n$, for example, $\{0, 1\}^n$.

The complexity of optimization problems over discrete domains is often worse that than it is over continuous domains (NP-hard).

We usually want to find the *maximum value* of f (subject to the constaints) and *find some/all values of x for which $f(x)$ is maximal* (an "argmax" problem).

Sometimes, one wants to *minimize* $f(x)$, but this is equivalent to *maximizing* $-f(x)$.

Sometimes the domain of f is a *discrete* subset of $\mathbb{R}^n$, for example, $\{0, 1\}^n$.

The complexity of optimization problems over discrete domains is often worse that than it is over continuous domains (NP-hard).

When we don't know how to solve efficiently a discrete optimization problem, we can try solving a *relaxation* of the problem.

This means that we let x vary over $\mathbb{R}^n$ instead of a discrete domain.
If we are lucky, the relaxed problem can be solved efficiently.

This means that we let x vary over $\mathbb{R}^n$ instead of a discrete domain.
If we are lucky, the relaxed problem can be solved efficiently.

However, the solutions of the relaxation are *rarely* solutions of the original problem.

This means that we let x vary over $\mathbb{R}^n$ instead of a discrete domain.
If we are lucky, the relaxed problem can be solved efficiently.

However, the solutions of the relaxation are *rarely* solutions of the original problem.

Solutions of the relaxation have to be "rounded". This is not always possible.

This means that we let x vary over $\mathbb{R}^n$ instead of a discrete domain.
If we are lucky, the relaxed problem can be solved efficiently.

However, the solutions of the relaxation are *rarely* solutions of the original problem.

Solutions of the relaxation have to be "rounded". This is not always possible.

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

A Simple Example

For example, find the maximum of

$$f(x, y) = 5x^2 + 4xy + 2y^2$$

on the unit circle

$$x^2 + y^2 = 1.$$

A Simple Example

For example, find the maximum of

$$f(x, y) = 5x^2 + 4xy + 2y^2$$

on the unit circle

$$x^2 + y^2 = 1.$$

It turns out that the maximum of f on the unit circle is 6 and that it is achieved for

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}.$$

A Simple Example

For example, find the maximum of

$$f(x, y) = 5x^2 + 4xy + 2y^2$$

on the unit circle

$$x^2 + y^2 = 1.$$

It turns out that the maximum of f on the unit circle is 6 and that it is achieved for

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}.$$

How did I figure that out?

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

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

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

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

The matrix

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$$

is symmetric ($A = A^\top$), so it can be *diagonalized*.

This means that there are (unit) vectors, e_1, e_2, that form a *basis* of $\mathbb{R}^2$ and such that

$$Ae_i = \lambda_i e_i, \quad i = 1, 2,$$

where the scalars, λ_1, λ_2, are real.

This means that there are (unit) vectors, e_1, e_2, that form a *basis* of $\mathbb{R}^2$ and such that

$$Ae_i = \lambda_i e_i, \quad i = 1, 2,$$

where the scalars, λ_1, λ_2, are real.

The vectors, e_1, e_2, are *eigenvectors* and the numbers, λ_1, λ_2, are *eigenvalues*, of A.

We say that e_i is an *eigenvector associated with* λ_i.

This means that there are (unit) vectors, e_1, e_2, that form a *basis* of $\mathbb{R}^2$ and such that

$$Ae_i = \lambda_i e_i, \quad i = 1, 2,$$

where the scalars, λ_1, λ_2, are real.

The vectors, e_1, e_2, are *eigenvectors* and the numbers, λ_1, λ_2, are *eigenvalues*, of A.

We say that e_i is an *eigenvector associated with* λ_i.

The eigenvalues of A are the zeros of the *characteristic polynomial*,

$$\det(\lambda I - A) = \begin{vmatrix} 5 - \lambda & 2 \\ 2 & 2 - \lambda \end{vmatrix} = \lambda^2 - 7\lambda + 6.$$

Furthermore, e_1 and e_2 are *orthogonal*, which means that their inner product is zero: $e_1 \cdot e_2 = 0$.

Furthermore, e_1 and e_2 are *orthogonal*, which means that their inner product is zero: $e_1 \cdot e_2 = 0$.

It turns out that

$$\lambda_1 = 6, \qquad \lambda_2 = 1,$$

and

$$e_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}, \qquad e_2 = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$$

Furthermore, e_1 and e_2 are *orthogonal*, which means that their inner product is zero: $e_1 \cdot e_2 = 0$.

It turns out that

$$\lambda_1 = 6, \qquad \lambda_2 = 1,$$

and

$$e_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}, \qquad e_2 = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$$

so we can write

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{-\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$$

The matrix

$$P = \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$$

has the following properties:

$$PP^{\top} = P^{\top}P = I.$$

The matrix

$$P = \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$$

has the following properties:

$$PP^\top = P^\top P = I.$$

A matrix, P, such that

$$PP^\top = P^\top P = I.$$

is called an *orthogonal matrix*.

Observe that

$$
\begin{aligned}
f(x,y) &= (x,y)\begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\
&= (x,y)\begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{-\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\
&= (x,y)P\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}P^\top\begin{pmatrix} x \\ y \end{pmatrix}
\end{aligned}
$$

Observe that

$$
\begin{aligned}
f(x,y) &= (x,y)\begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\
&= (x,y)\begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{-\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\
&= (x,y)P\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}P^{\top}\begin{pmatrix} x \\ y \end{pmatrix}
\end{aligned}
$$

If we let

$$
\begin{pmatrix} u \\ v \end{pmatrix} = P^{\top}\begin{pmatrix} x \\ y \end{pmatrix},
$$

then

$$
\begin{aligned}
f(u,v) &= (u,v)\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} \\
&= 6u^2 + v^2.
\end{aligned}
$$

Furthermore, the constraint

$$x^2 + y^2 = 1$$

can be written as

$$(x, y)\begin{pmatrix} x \\ y \end{pmatrix} = 1.$$

Furthermore, the constraint

$$x^2 + y^2 = 1$$

can be written as

$$(x, y)\begin{pmatrix} x \\ y \end{pmatrix} = 1.$$

Using the fact that

$$\begin{pmatrix} u \\ v \end{pmatrix} = P^\top \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{so} \quad \begin{pmatrix} x \\ y \end{pmatrix} = P \begin{pmatrix} u \\ v \end{pmatrix},$$

we get

$$1 = (x, y)\begin{pmatrix} x \\ y \end{pmatrix} = (u, v)P^\top P \begin{pmatrix} u \\ v \end{pmatrix} = (u, v)\begin{pmatrix} u \\ v \end{pmatrix},$$

Furthermore, the constraint

$$x^2 + y^2 = 1$$

can be written as

$$(x, y)\begin{pmatrix} x \\ y \end{pmatrix} = 1.$$

Using the fact that

$$\begin{pmatrix} u \\ v \end{pmatrix} = P^\top \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{so} \quad \begin{pmatrix} x \\ y \end{pmatrix} = P \begin{pmatrix} u \\ v \end{pmatrix},$$

we get

$$1 = (x, y)\begin{pmatrix} x \\ y \end{pmatrix} = (u, v)P^\top P\begin{pmatrix} u \\ v \end{pmatrix} = (u, v)\begin{pmatrix} u \\ v \end{pmatrix},$$

so

$$u^2 + v^2 = 1.$$

Therefore, we have to find the maximum of

$$f(u, v) = 6u^2 + v^2$$

where

$$u^2 + v^2 = 1.$$

Therefore, we have to find the maximum of

$$f(u,v) = 6u^2 + v^2$$

where

$$u^2 + v^2 = 1.$$

Note that on the circle, $u^2 + v^2 = 1$,

$$f(u,v) = 6u^2 + v^2 \leq 6(u^2 + v^2) \leq 6,$$

and

$$f(1,0) = 6.$$

Therefore, we have to find the maximum of

$$f(u, v) = 6u^2 + v^2$$

where

$$u^2 + v^2 = 1.$$

Note that on the circle, $u^2 + v^2 = 1$,

$$f(u, v) = 6u^2 + v^2 \leq 6(u^2 + v^2) \leq 6,$$

and

$$f(1, 0) = 6.$$

So, the *maximum* of f on the unit circle is indeed 6.

This maximum is achieved for $(u, v) = (1, 0)$, and since

$$\begin{pmatrix} x \\ y \end{pmatrix} = P \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

this yields

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}.$$

This maximum is achieved for $(u, v) = (1, 0)$, and since

$$\begin{pmatrix} x \\ y \end{pmatrix} = P \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

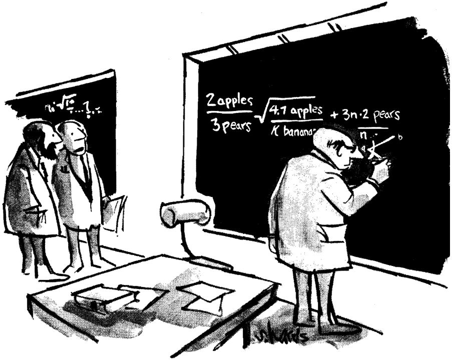
this yields

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}.$$

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.



"IF ONLY HE COULD THINK IN ABSTRACT TERMS."

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., $S(A)^\top = -S(A)$,

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., $S(A)^\top = -S(A)$, and

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

Indeed, is S is *skew symmetric*, as $f(x) = x^\top S x$ is a scalar, so

Indeed, is S is *skew symmetric*, as $f(x) = x^\top S x$ is a scalar, so

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

Indeed, is S is *skew symmetric*, as $f(x) = x^{\top}Sx$ is a scalar, so

$$\begin{aligned} f(x) &= f(x)^{\top} \\ &= (x^{\top}Sx)^{\top} \end{aligned}$$

Indeed, is S is *skew symmetric*, as $f(x) = x^\top S x$ is a scalar, so

$$\begin{aligned} f(x) &= f(x)^\top \\ &= (x^\top S x)^\top \\ &= x^\top S^\top x \end{aligned}$$

Indeed, is S is *skew symmetric*, as $f(x) = x^\top S x$ is a scalar, so

$$
\begin{aligned}
f(x) &= f(x)^\top \\
&= (x^\top S x)^\top \\
&= x^\top S^\top x \\
&= -x^\top S x
\end{aligned}
$$

Indeed, is S is *skew symmetric*, as $f(x) = x^\top Sx$ is a scalar, so

$$\begin{aligned} f(x) &= f(x)^\top \\ &= (x^\top Sx)^\top \\ &= x^\top S^\top x \\ &= -x^\top Sx \\ &= -f(x) \end{aligned}$$

Indeed, is S is *skew symmetric*, as $f(x) = x^{\top} S x$ is a scalar, so

$$\begin{aligned} f(x) &= f(x)^{\top} \\ &= (x^{\top} S x)^{\top} \\ &= x^{\top} S^{\top} x \\ &= -x^{\top} S x \\ &= -f(x) \end{aligned}$$

and we get $2f(x) = 0$, that is, $f(x) = 0$.

Indeed, is S is *skew symmetric*, as $f(x) = x^\top Sx$ is a scalar, so

$$\begin{aligned} f(x) = f(x)^\top \\ &= (x^\top Sx)^\top \\ &= x^\top S^\top x \\ &= -x^\top Sx \\ &= -f(x) \end{aligned}$$

and we get $2f(x) = 0$, that is, $f(x) = 0$.

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

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

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

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

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

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

Courant Fischer

Consequently, generalizing the proof given for $n = 2$, we have:

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

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

Courant Fischer

Consequently, generalizing the proof given for $n = 2$, we have:

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

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.

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.

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.

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.

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.

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, $T(k) = k/|S|$).

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

Very recently, Shi and Kennedy 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.

Very recently, Shi and Kennedy 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.

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.

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.

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 $C_e(r, \theta, \theta_{\max})$ 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 $0 < \delta_{\mathrm{min}} \leq \delta \leq \delta_{\mathrm{max}}$, we would like to solve the following optimization problem:

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

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:

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

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

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

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 $\delta_{\mathrm{min}} \leq \delta \leq \delta_{\mathrm{max}}$ 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.

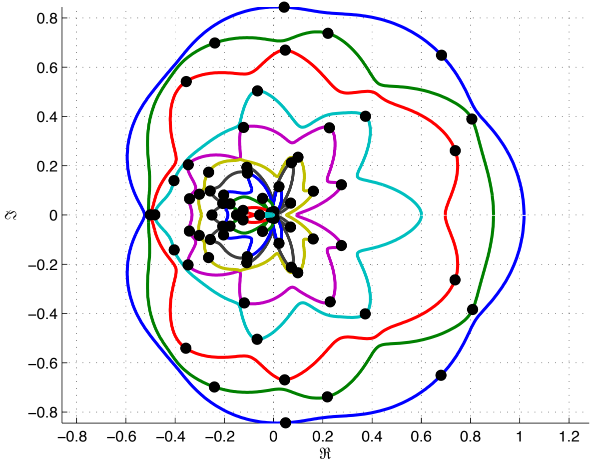


Figure: The eigenvalues of a matrix $H(\delta)$ which is not normal

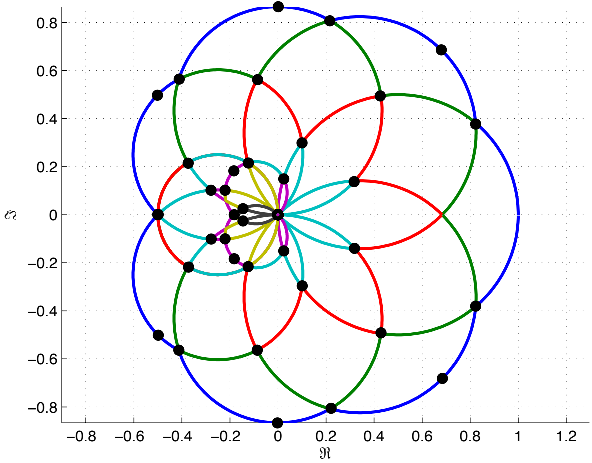


Figure: The eigenvalues of a normal matrix $H(\delta)$

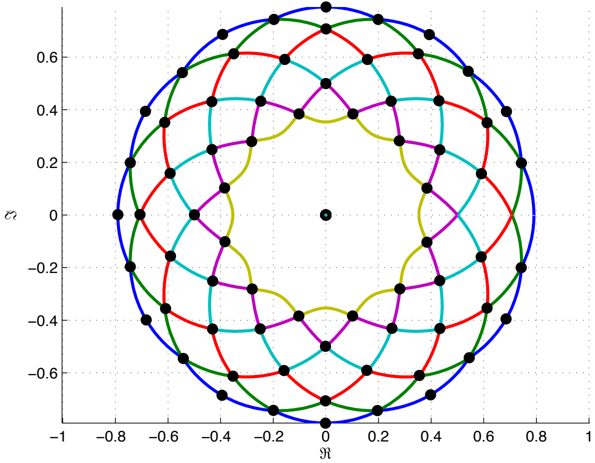


Figure: The eigenvalues of a matrix $H(\delta)$ which is near normal

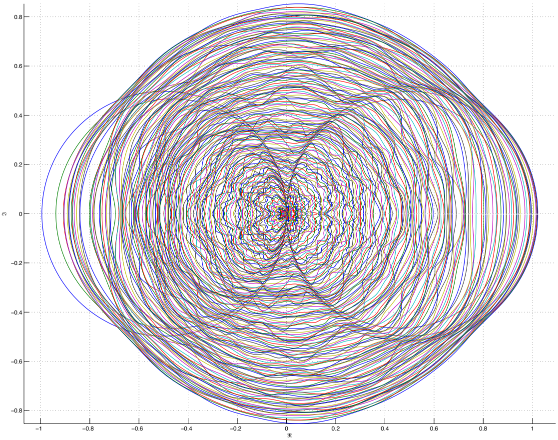


Figure: The eigenvalues of the matrix for an actual image

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

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

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.

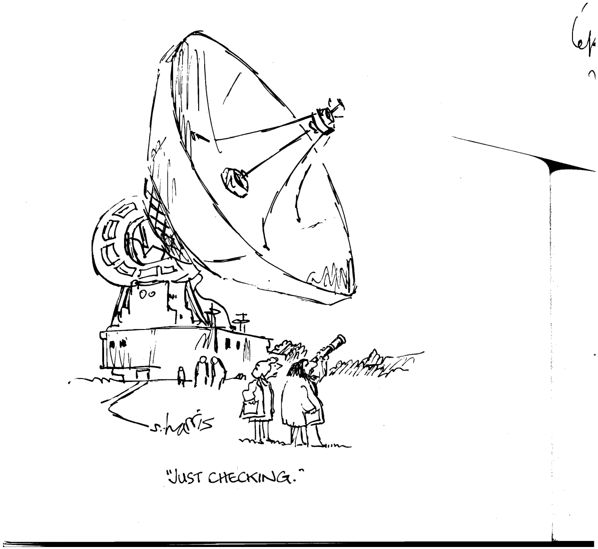


Figure: Just checking!

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

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

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 $x^* P x$ in polar form as

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

I proved that

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

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.

However, we can show that if $f(x, \delta) = x^* H x$ is a local maximum of f at (x, δ), then $\delta = \varphi$, the phase angle of $|x^* P x|$ and so, $x^* H x = |x^* P x|$.

Unfortunately, this doesn't not seem to help much in finding for which δ the function $f(x, \delta)$ has local maxima.

Unfortunately, this doesn't not seem to help much in finding for which δ the function $f(x, \delta)$ has local maxima.

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 $|x^* P x|$.

Unfortunately, this doesn't not seem to help much in finding for which δ the function $f(x, \delta)$ has local maxima.

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 $|x^* P x|$ (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.

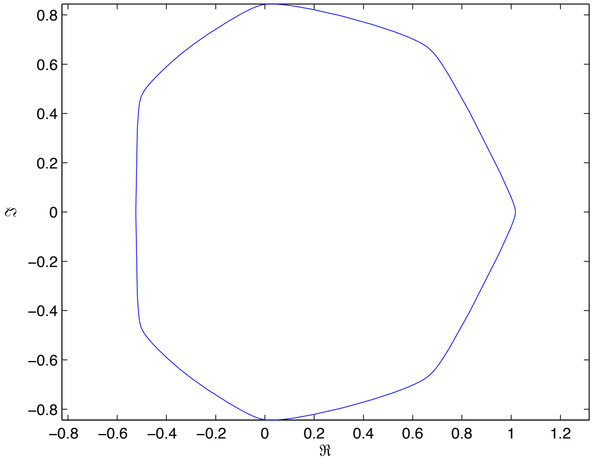


Figure: Numerical Range of a matrix which is not normal

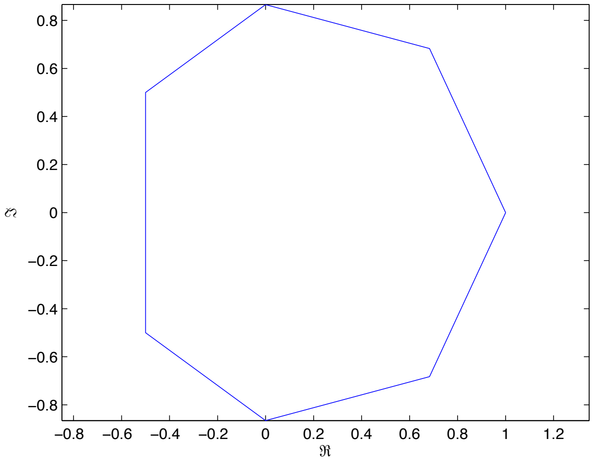


Figure: Numerical Range of a normal matrix

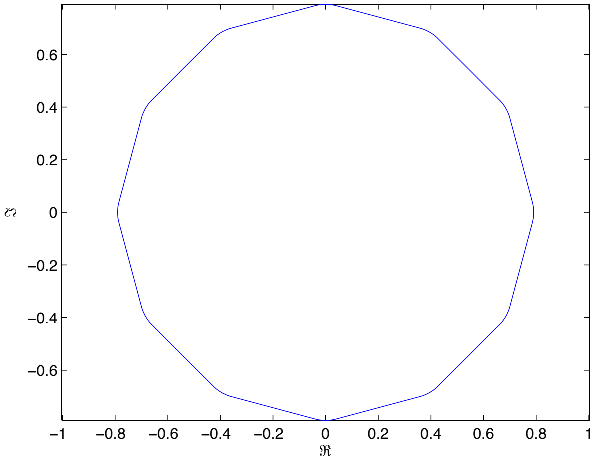


Figure: Numerical Range of a matrix which is near normal

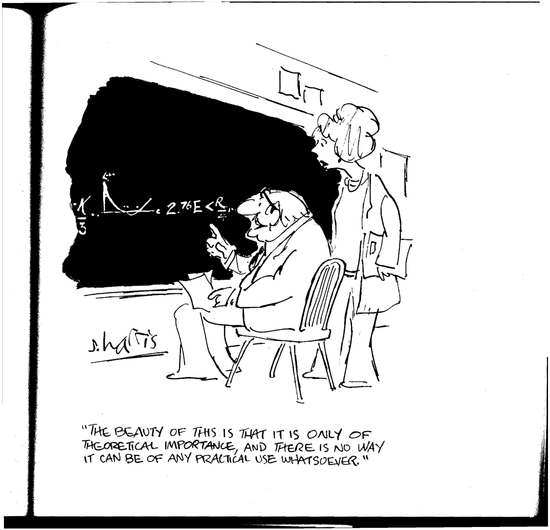


Figure: Beauty

The quantity

$$r(P) = \max\{|z| \mid z \in F(P)\}$$

is called the *numerical radius* of P.

It is obviously of interest to us since it corresponds to the maximum of $|x^* P x|$, over all unit vectors, x.

The quantity

$$r(P) = \max\{|z| \mid z \in F(P)\}$$

is called the *numerical radius* of P.

It is obviously of interest to us since it corresponds to the maximum of $|x^* P x|$, over all unit vectors, x.

It is easy to show that

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

and so,

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

The quantity

$$r(P) = \max\{|z| \mid z \in F(P)\}$$

is called the *numerical radius* of P.

It is obviously of interest to us since it corresponds to the maximum of $|x^* P x|$, over all unit vectors, x.

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, $F(P)$.

To show this, we use a proposition from Horn and Johnson whose proof is quite simple:

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

To show this, we use a proposition from Horn and Johnson whose proof is quite simple:

Theorem 3

For any $n \times n$ matrix, P, and any unit vector, $x \in \mathbb{C}^n$, the following properties are equivalent:

(1) $\mathrm{Re}(x^* P x) = \max\{\mathrm{Re}(z) \mid z \in F(P)\}$

(2) $x^* H(P) x = \max\{r \mid r \in F(H(P))\}$

(3) *The vector, x, is an eigenvector of $H(P)$ corresponding to the largest eigenvalue, λ_1, of $H(P)$.*

In fact, Theorem 3 immediately implies that

$$\max\{\mathrm{Re}(z) \mid z \in F(P)\} = \max\{r \mid r \in F(H(P))\} = \lambda_1.$$

In fact, Theorem 3 immediately implies that

$$\max\{\mathrm{Re}(z) \mid z \in F(P)\} = \max\{r \mid r \in F(H(P))\} = \lambda_1.$$

As a consequence, for every angle, $\delta \in [0, 2\pi)$, if we let λ_δ be the largest eigenvalue of the matrix $H(e^{-i\delta}P)$ and if x_δ is a corresponding unit eigenvector, then $z_\delta = x_\delta^* P x_\delta$ is on the boundary, $\partial F(P)$, of $F(P)$ and the line, L_δ, given by

$$\begin{aligned} L_\delta &= \{e^{i\delta}(\lambda_\delta + ti) \mid t \in \mathbb{R}\} \\ &= \{(x, y) \in \mathbb{R}^2 \mid \cos\delta\, x + \sin\delta\, y - \lambda_\delta = 0\}, \end{aligned}$$

is a supporting line of $F(P)$ at z_δ.

These properties can be exploited to find the local maxima of the objective function of our contour identification problem. Ryan Kennedy has done some work on this problem.

These properties can be exploited to find the local maxima of the objective function of our contour identification problem. Ryan Kennedy has done some work on this problem.

Much more work needs to be done, in particular

These properties can be exploited to find the local maxima of the objective function of our contour identification problem. Ryan Kennedy has done some work on this problem.

Much more work needs to be done, in particular

- Cope with the dimension of the matrix, P

These properties can be exploited to find the local maxima of the objective function of our contour identification problem. Ryan Kennedy has done some work on this problem.

Much more work needs to be done, in particular

- Cope with the dimension of the matrix, P
- Understand the role of various *normalizations* of P (stochastic, bi-stochastic).

These properties can be exploited to find the local maxima of the objective function of our contour identification problem. Ryan Kennedy has done some work on this problem.

Much more work needs to be done, in particular

- Cope with the dimension of the matrix, P
- Understand the role of various *normalizations* of P (stochastic, bi-stochastic).
- Shi and Kennedy have made recent progress on the issue of normalization.

Other Quadratic Optimization Problems

Some variations of the basic quadratic optimization problem have occurred in the contex of computer vision:

Other Quadratic Optimization Problems

Some variations of the basic quadratic optimization problem have occurred in the contex of computer vision:

- Adding linear constraints of the form

$$C^\top x = 0.$$

Other Quadratic Optimization Problems

Some variations of the basic quadratic optimization problem have occurred in the contex of computer vision:

- Adding linear constraints of the form

$$C^\top x = 0.$$

  It is relatively easy to eliminate these constraints using a QR-decomposition or a pseudo-inverse. This was shown by Golub (1973).

Other Quadratic Optimization Problems

Some variations of the basic quadratic optimization problem have occurred in the contex of computer vision:

- Adding linear constraints of the form

$$C^\top x = 0.$$

  It is relatively easy to eliminate these constraints using a QR-decomposition or a pseudo-inverse. This was shown by Golub (1973).

- Adding linear constraints of the form

$$C^\top x = t,$$

  where $t \neq 0$.

Other Quadratic Optimization Problems

Some variations of the basic quadratic optimization problem have occurred in the contex of computer vision:

- Adding linear constraints of the form

$$C^\top x = 0.$$

  It is relatively easy to eliminate these constraints using a QR-decomposition or a pseudo-inverse. This was shown by Golub (1973).

- Adding linear constraints of the form

$$C^\top x = t,$$

  where $t \neq 0$. This is a lot harder to deal with!

Gander, Golub and von Matt (1989) showed that these constraints can be eliminated *but* a linear term has to be added to the objective function.

Gander, Golub and von Matt (1989) showed that these constraints can be eliminated *but* a linear term has to be added to the objective function.

One needs to consider objective functions of the form

$$f(x) = x^\top A x + 2x^\top b,$$

with $b \neq 0$.

Gander, Golub and von Matt (1989) showed that these constraints can be eliminated *but* a linear term has to be added to the objective function.

One needs to consider objective functions of the form

$$f(x) = x^\top Ax + 2x^\top b,$$

with $b \neq 0$.

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

Gander, Golub and von Matt (1989) showed that these constraints can be eliminated *but* a linear term has to be added to the objective function.

One needs to consider objective functions of the form

$$f(x) = x^\top A x + 2x^\top b,$$

with $b \neq 0$.

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 $\mathbb{R}^n$, but this will have to wait for another talk!