Optimization of Polynomial Roots, Eigenvalues and Pseudospectra

Michael L. Overton

Courant Institute of Mathematical Sciences
NYU

FoCM 2011
Budapest

Part I
Globally Optimizing the Roots of a Monic Polynomial subject to One Affine Constraint with

V. Blondel (Louvain)
M. Gürbüzbalaban (NYU)
A. Megretski (MIT)

The Root Radius and the Root Abscissa

Let ρ denote the *root radius* of a polynomial:

$$\rho(p) = \max\{|z| \ : \ p(z) = 0, \ z \in \mathbf{C}\}.$$

The Root Radius and the Root Abscissa

Let ρ denote the *root radius* of a polynomial:

$$\rho(p) = \max \{|z| \ : \ p(z) = 0, \ z \in \mathbf{C}\} .$$

In discrete-time linear dynamical systems, $\rho(p) < 1$ is the *stability* condition.

The Root Radius and the Root Abscissa

Let ρ denote the *root radius* of a polynomial:

$$\rho(p) = \max\{|z| \ : \ p(z) = 0, \ z \in \mathbf{C}\}.$$

In discrete-time linear dynamical systems, $\rho(p) < 1$ is the *stability* condition.

Let α denote the *root abscissa*:

$$\alpha(p) = \max\{\mathrm{Re}(z) \ : \ p(z) = 0, \ z \in \mathbf{C}\}.$$

The Root Radius and the Root Abscissa

Let ρ denote the *root radius* of a polynomial:

$$\rho(p) = \max\{|z| \ : \ p(z) = 0, \ z \in \mathbf{C}\}.$$

In discrete-time linear dynamical systems, $\rho(p) < 1$ is the *stability* condition.

Let α denote the *root abscissa*:

$$\alpha(p) = \max\{\mathrm{Re}(z) \ : \ p(z) = 0, \ z \in \mathbf{C}\}.$$

In continuous-time linear dynamical systems, $\alpha(p) < 0$ is the stability condition.

Stability Optimization over a Polynomial Family

As functions of the polynomial coefficients, the radius ρ and abscissa α are

Stability Optimization over a Polynomial Family

As functions of the polynomial coefficients, the radius ρ and abscissa α are

- not convex

Stability Optimization over a Polynomial Family

As functions of the polynomial coefficients, the radius ρ and abscissa α are

- not convex
- not Lipschitz near polynomials with a multiple root

Stability Optimization over a Polynomial Family

As functions of the polynomial coefficients, the radius ρ and abscissa α are

- not convex
- not Lipschitz near polynomials with a multiple root

So, in general, global minimization of the radius or abscissa over an affine family of monic polynomials, pushing the roots as far as possible towards the origin or left in the complex plane, seems hard.

Stability Optimization over a Polynomial Family

As functions of the polynomial coefficients, the radius ρ and abscissa α are

- not convex
- not Lipschitz near polynomials with a multiple root

So, in general, global minimization of the radius or abscissa over an affine family of monic polynomials, pushing the roots as far as possible towards the origin or left in the complex plane, seems hard.

Indeed, variations on the question of whether a polynomial family contains one that is stable (has roots inside the unit circle or in the left-half plane) have been studied for decades.

Stability Optimization over a Polynomial Family

As functions of the polynomial coefficients, the radius ρ and abscissa α are

- not convex
- not Lipschitz near polynomials with a multiple root

So, in general, global minimization of the radius or abscissa over an affine family of monic polynomials, pushing the roots as far as possible towards the origin or left in the complex plane, seems hard.

Indeed, variations on the question of whether a polynomial family contains one that is stable (has roots inside the unit circle or in the left-half plane) have been studied for decades.

But if an affine family of monic polynomials of degree n has $n-1$ free parameters, this question can be answered efficiently.

Stability Optimization over a Polynomial Family

As functions of the polynomial coefficients, the radius ρ and abscissa α are

- not convex
- not Lipschitz near polynomials with a multiple root

So, in general, global minimization of the radius or abscissa over an affine family of monic polynomials, pushing the roots as far as possible towards the origin or left in the complex plane, seems hard.

Indeed, variations on the question of whether a polynomial family contains one that is stable (has roots inside the unit circle or in the left-half plane) have been studied for decades.

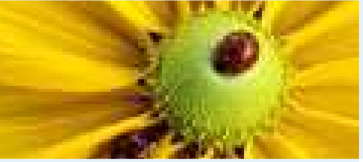
But if an affine family of monic polynomials of degree n has $n-1$ free parameters, this question can be answered efficiently.

Equivalently, there is *just one affine constraint* on the coefficients.

Optimizing the Root Radius, Real Case

Theorem RRR. *Let $B_0, B_1, \ldots, B_n$ be real scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \;:\; B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{R}\}.$$



Optimizing the Root Radius, Real Case

Theorem RRR. *Let $B_0, B_1, \ldots, B_n$ be real scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \;:\; B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{R}\}.$$

The optimization problem

$$\rho^* := \inf_{p \in P} \rho(p)$$

has a globally optimal solution of the form

$$p^*(z) = (z - \gamma)^{n-k}(z + \gamma)^k \in P$$

for some integer k with $0 \le k \le n$, where $\gamma = \rho^*$.

Optimizing the Root Radius, Real Case

Theorem RRR. *Let $B_0, B_1, \ldots, B_n$ be real scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \ : \ B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{R}\}.$$

The optimization problem

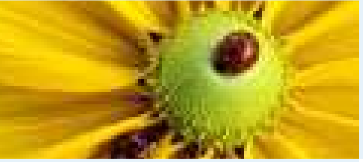
$$\rho^* := \inf_{p \in P} \rho(p)$$

has a globally optimal solution of the form

$$p^*(z) = (z-\gamma)^{n-k}(z+\gamma)^k \in P$$

for some integer k with $0 \le k \le n$, where $\gamma = \rho^*$.

Proof: uses implicit function theorem.



Algorithm to Find γ and k

Corollary RRR. *Let γ be the globally optimal value whose existence is asserted in Theorem RRR, and consider the set*

$$\Xi = \{r \in \mathbf{R} \ : \ g_k(r) = 0 \text{ for some } k \in \{0, 1, \ldots, n\}\}$$

where

$$g_k(z) = B_0 v_0 + B_1 v_1 z + \ldots + B_{n-1} v_{n-1} z^{n-1} + B_n v_n z^n$$

and $(v_0, \ldots, v_n)$ *is the convolution of the vectors*

$$\left(\binom{n-k}{0}, \binom{n-k}{1}, \ldots \binom{n-k}{n-k}\right) \text{ and } \left(\binom{k}{0}, -\binom{k}{1}, \ldots (-1)^k \binom{k}{k}\right)$$

for $k = 0, \ldots, n$.

Algorithm to Find γ and k

Corollary RRR. *Let γ be the globally optimal value whose existence is asserted in Theorem RRR, and consider the set*

$$\Xi = \{r \in \mathbf{R} \;:\; g_k(r) = 0 \text{ for some } k \in \{0, 1, \ldots, n\}\}$$

where

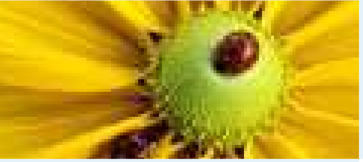
$$g_k(z) = B_0 v_0 + B_1 v_1 z + \ldots + B_{n-1} v_{n-1} z^{n-1} + B_n v_n z^n$$

and $(v_0, \ldots, v_n)$ is the convolution of the vectors

$$\left(\binom{n-k}{0}, \binom{n-k}{1}, \ldots \binom{n-k}{n-k}\right) \text{ and } \left(\binom{k}{0}, -\binom{k}{1}, \ldots (-1)^k \binom{k}{k}\right)$$

for $k = 0, \ldots, n$.

Then, $-\gamma$ is an element of Ξ with smallest magnitude.



Algorithm to Find γ and k

Corollary RRR. *Let γ be the globally optimal value whose existence is asserted in Theorem RRR, and consider the set*

$$\Xi = \{r \in \mathbf{R} \ : \ g_k(r) = 0 \text{ for some } k \in \{0, 1, \ldots, n\}\}$$

where

$$g_k(z) = B_0 v_0 + B_1 v_1 z + \ldots + B_{n-1} v_{n-1} z^{n-1} + B_n v_n z^n$$

and $(v_0, \ldots, v_n)$ is the convolution of the vectors

$$\left(\binom{n-k}{0}, \binom{n-k}{1}, \ldots \binom{n-k}{n-k}\right) \text{ and } \left(\binom{k}{0}, -\binom{k}{1}, \ldots (-1)^k \binom{k}{k}\right)$$

for $k = 0, \ldots, n$.

Then, $-\gamma$ is an element of Ξ with smallest magnitude.

Proof: immediate.

Optimizing the Root Radius: Complex Case

Theorem RRC. *Let $B_0, B_1, \ldots, B_n$ be complex scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \;:\; B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{C}\}.$$

Optimizing the Root Radius: Complex Case

Theorem RRC. *Let $B_0, B_1, \ldots, B_n$ be complex scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \ : \ B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{C}\}.$$

The optimization problem

$$\rho^* := \inf_{p \in P} \rho(p)$$

has an optimal solution of the form

$$p^*(z) = (z - \gamma)^n \in P$$

with $-\gamma$ given by a root of smallest magnitude of the polynomial

$$h(z) = B_n z^n + B_{n-1} \binom{n}{n-1} z^{n-1} + \ldots + B_1 \binom{n}{1} z + B_0.$$

Optimizing the Root Radius: Complex Case

Theorem RRC. *Let $B_0, B_1, \ldots, B_n$ be complex scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \ : \ B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{C}\}.$$

The optimization problem

$$\rho^* := \inf_{p \in P} \rho(p)$$

has an optimal solution of the form

$$p^*(z) = (z-\gamma)^n \in P$$

with $-\gamma$ given by a root of smallest magnitude of the polynomial

$$h(z) = B_n z^n + B_{n-1}\binom{n}{n-1} z^{n-1} + \ldots + B_1 \binom{n}{1} z + B_0.$$

Proof: Surprisingly, much more complicated than the real case.

Optimizing the Root Abscissa: Real Case

Theorem RAR. *Let $B_0, B_1, \ldots, B_n$ be real scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \ : \ B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{R}\}.$$

Theorem RAR. *Let $B_0, B_1, \ldots, B_n$ be real scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \;:\; B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{R}\}.$$

Let $k = \max\{j : B_j \neq 0\}$. and define the polynomial of degree k

$$h(z) = B_n z^n + B_{n-1} \binom{n}{n-1} z^{n-1} + \ldots + B_1 \binom{n}{1} z + B_0.$$

Optimizing the Root Abscissa: Real Case

Theorem RAR. *Let $B_0, B_1, \ldots, B_n$ be real scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \ : \ B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{R}\}.$$

Let $k = \max\{j : B_j \neq 0\}$. and define the polynomial of degree k

$$h(z) = B_n z^n + B_{n-1}\binom{n}{n-1} z^{n-1} + \ldots + B_1 \binom{n}{1} z + B_0.$$

The optimization problem

$$\alpha^* := \inf_{p \in P} \alpha(p).$$

has the infimal value

$$\alpha^* = \min\left\{\beta \in \mathbf{R} \ : \ h^{(i)}(-\beta) = 0 \text{ for some } i \in \{0, \ldots, k-1\}\right\},$$

where $h^{(i)}$ is the i-th derivative of h.

Continuation

Furthermore, the optimal value α^* is attained by a minimizing polynomial p^* if and only if $-\alpha^*$ is a root of h (as opposed to one of its derivatives), and in this case we can take

$$p^*(z) = (z - \gamma)^n \in P$$

with $\gamma = \alpha^*$.

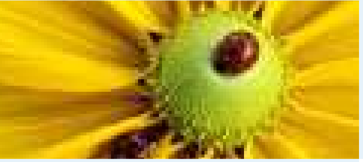
Continuation

Furthermore, the optimal value α^* is attained by a minimizing polynomial p^* if and only if $-\alpha^*$ is a root of h (as opposed to one of its derivatives), and in this case we can take

$$p^*(z) = (z - \gamma)^n \in P$$

with $\gamma = \alpha^*$.

Proof: substantially more complicated than radius case, because optimal value may not be attained.



When the Optimal Abscissa Value is not Attained

Theorem RAR2. Assume that $-\alpha^*$ is not a root of h. Let ℓ be the smallest integer $i \in \{1, \ldots, k-1\}$ for which $-\alpha^*$ is a root of $h^{(i)}$. Then, for all sufficiently small $\epsilon > 0$ there exists a real scalar M_ϵ for which

$$p_\epsilon(z) := (z - M_\epsilon)^m (z - (\alpha^* + \epsilon))^{n-m} \in P$$

where $m = \ell$ or $\ell + 1$, and $M_\epsilon \to -\infty$ as $\epsilon \to 0$.

When the Optimal Abscissa Value is not Attained

Theorem RAR2. Assume that $-\alpha^*$ is not a root of h. Let ℓ be the smallest integer $i \in \{1, \ldots, k-1\}$ for which $-\alpha^*$ is a root of $h^{(i)}$. Then, for all sufficiently small $\epsilon > 0$ there exists a real scalar M_ϵ for which

$$p_\epsilon(z) := (z - M_\epsilon)^m (z - (\alpha^* + \epsilon))^{n-m} \in P$$

where $m = \ell$ or $\ell + 1$, and $M_\epsilon \to -\infty$ as $\epsilon \to 0$.

In other words, as in the real radius case, two roots play a role, but only one is finite.

When the Optimal Abscissa Value is not Attained

Theorem RAR2. Assume that $-\alpha^*$ is not a root of h. Let ℓ be the smallest integer $i \in \{1, \ldots, k-1\}$ for which $-\alpha^*$ is a root of $h^{(i)}$. Then, for all sufficiently small $\epsilon > 0$ there exists a real scalar M_ϵ for which

$$p_\epsilon(z) := (z - M_\epsilon)^m (z - (\alpha^* + \epsilon))^{n-m} \in P$$

where $m = \ell$ or $\ell + 1$, and $M_\epsilon \to -\infty$ as $\epsilon \to 0$.

In other words, as in the real radius case, two roots play a role, but only one is finite.

There is a formula for M_ϵ, but it is a little complicated and depends on whether $m = \ell$ or $m = \ell + 1$.

When the Optimal Abscissa Value is not Attained

Theorem RAR2. Assume that $-\alpha^*$ is not a root of h. Let ℓ be the smallest integer $i \in \{1, \ldots, k-1\}$ for which $-\alpha^*$ is a root of $h^{(i)}$. Then, for all sufficiently small $\epsilon > 0$ there exists a real scalar M_ϵ for which

$$p_\epsilon(z) := (z - M_\epsilon)^m (z - (\alpha^* + \epsilon))^{n-m} \in P$$

where $m = \ell$ or $\ell + 1$, and $M_\epsilon \to -\infty$ as $\epsilon \to 0$.

In other words, as in the real radius case, two roots play a role, but only one is finite.

There is a formula for M_ϵ, but it is a little complicated and depends on whether $m = \ell$ or $m = \ell + 1$.

In practice, bad idea to make ϵ too small: then $|M_\epsilon|$ becomes large.

Optimizing the Abscissa: Real vs. Complex Case

We observed that, in the real case, the optimal value is not attained when one of the *derivatives of* h has a real root to the right of the *rightmost real root* of h.

Optimizing the Abscissa: Real vs. Complex Case

We observed that, in the real case, the optimal value is not attained when one of the *derivatives of* h has a real root to the right of the *rightmost real root* of h.

However, it is not possible that a derivative of h has a complex root to the right of the *rightmost complex root* of h. This follows immediately from the Gauss-Lucas theorem.

Optimizing the Abscissa: Real vs. Complex Case

We observed that, in the real case, the optimal value is not attained when one of the *derivatives of* h has a real root to the right of the *rightmost real root* of h.

However, it is not possible that a derivative of h has a complex root to the right of the *rightmost complex root* of h. This follows immediately from the Gauss-Lucas theorem.

This suggests the optimal abscissa value might always be attained in the complex case.

Optimizing the Abscissa: Real vs. Complex Case

We observed that, in the real case, the optimal value is not attained when one of the *derivatives of* h has a real root to the right of the *rightmost real root* of h.

However, it is not possible that a derivative of h has a complex root to the right of the *rightmost complex root* of h. This follows immediately from the Gauss-Lucas theorem.

This suggests the optimal abscissa value might always be attained in the complex case.

Indeed, this is the case...

Optimizing the Root Abscissa: Complex Case

Theorem RAC. *Let* $B_0, B_1, \ldots, B_n$ *be complex scalars (with* $B_1, \ldots, B_n$ *not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \ : \ B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{C}\}.$$

Optimizing the Root Abscissa: Complex Case

Theorem RAC. *Let $B_0, B_1, \ldots, B_n$ be complex scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \; : \; B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{C}\}.$$

The optimization problem

$$\alpha^* := \inf_{p \in P} \alpha(p)$$

has an optimal solution of the form

$$p^*(z) = (z - \gamma)^n \in P$$

with $-\gamma$ given by a root with largest real part of the polynomial

$$h(z) = B_n z^n + B_{n-1} \binom{n}{n-1} z^{n-1} + \ldots + B_1 \binom{n}{1} z + B_0.$$

Optimizing the Root Abscissa: Complex Case

Theorem RAC. *Let $B_0, B_1, \ldots, B_n$ be complex scalars (with $B_1, \ldots, B_n$ not all zero) and consider the affine family*

$$P = \{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n \ : \ B_0 + \sum_{j=1}^{n} B_j a_j = 0, a_i \in \mathbf{C}\}.$$

The optimization problem

$$\alpha^* := \inf_{p \in P} \alpha(p)$$

has an optimal solution of the form

$$p^*(z) = (z - \gamma)^n \in P$$

with $-\gamma$ given by a root with largest real part of the polynomial

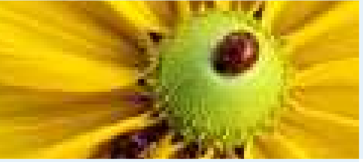
$$h(z) = B_n z^n + B_{n-1} \binom{n}{n-1} z^{n-1} + \ldots + B_1 \binom{n}{1} z + B_0.$$

Proof: follows (with some work) from Theorem RRC.

A Simple Example

Minimize the root abscissa over the set of polynomials

$$P = \{z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_{n-1} z + a_n \ : \ a_1 + a_2 = 0\}.$$



A Simple Example

Minimize the root abscissa over the set of polynomials

$$P = \{z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_{n-1} z + a_n \ : \ a_1 + a_2 = 0\}.$$

We have

$$h(z) = \binom{n}{2} z^2 + nz = \binom{n}{2} z(z + \frac{2}{n-1}).$$

A Simple Example

Minimize the root abscissa over the set of polynomials

$$P = \{z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_{n-1} z + a_n \ : \ a_1 + a_2 = 0\}.$$

We have

$$h(z) = \binom{n}{2} z^2 + nz = \binom{n}{2} z(z + \frac{2}{n-1}).$$

The roots of h are 0 and $\frac{-2}{n-1}$ and the only root of h' is $\frac{-1}{n-1}$ so $\alpha^* = 0$ and z^n is a global optimizer.

A Simple Example

Minimize the root abscissa over the set of polynomials

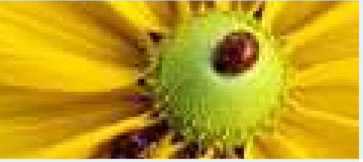
$$P = \{z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_{n-1} z + a_n \ : \ a_1 + a_2 = 0\}.$$

We have

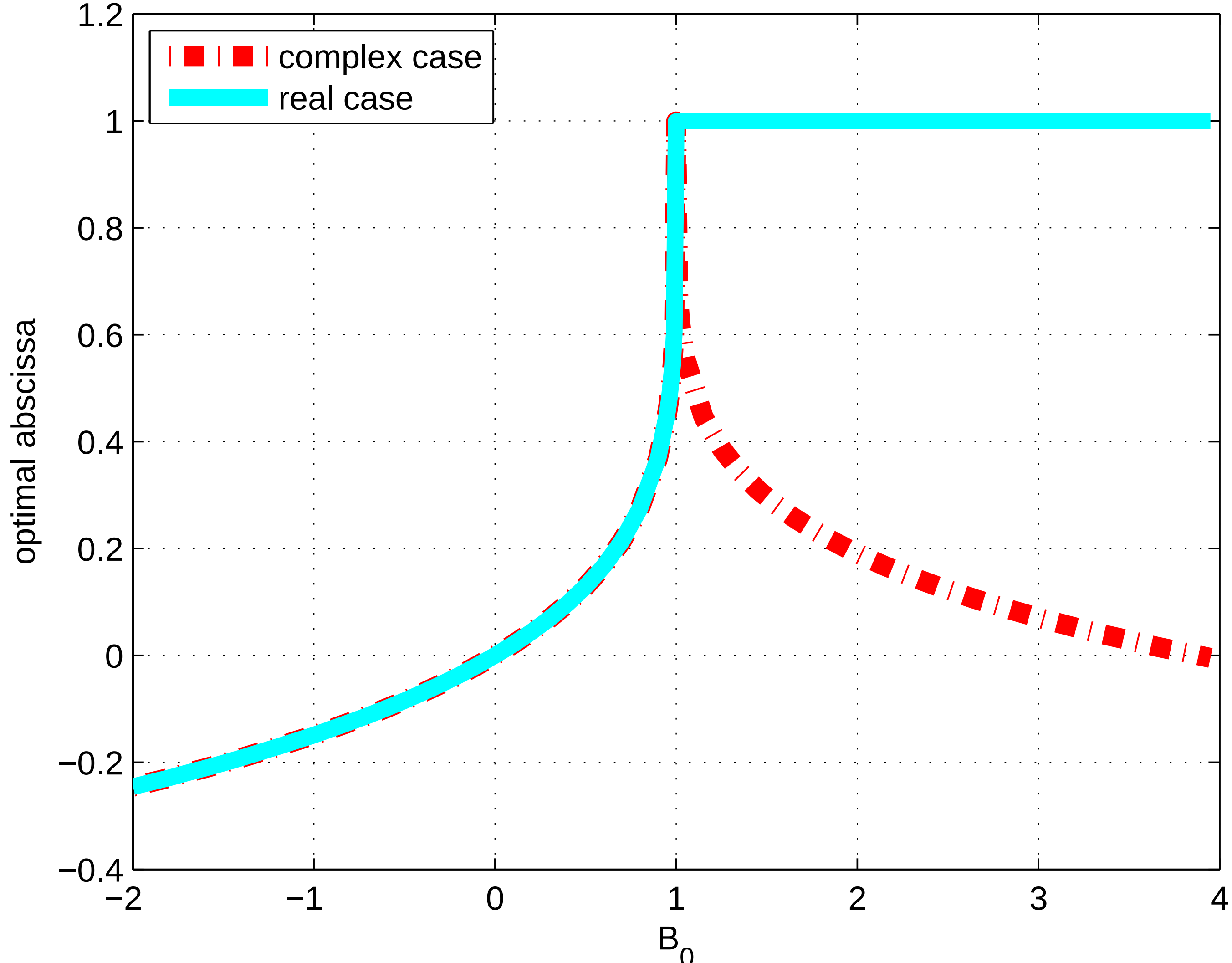
$$h(z) = \binom{n}{2} z^2 + nz = \binom{n}{2} z(z + \frac{2}{n-1}).$$

The roots of h are 0 and $\frac{-2}{n-1}$ and the only root of h' is $\frac{-1}{n-1}$ so $\alpha^* = 0$ and z^n is a global optimizer.

Theorem RAR proves global optimality over $a_i \in \mathbf{R}$ and Theorem RAC proves global optimality over $a_i \in \mathbf{C}$.

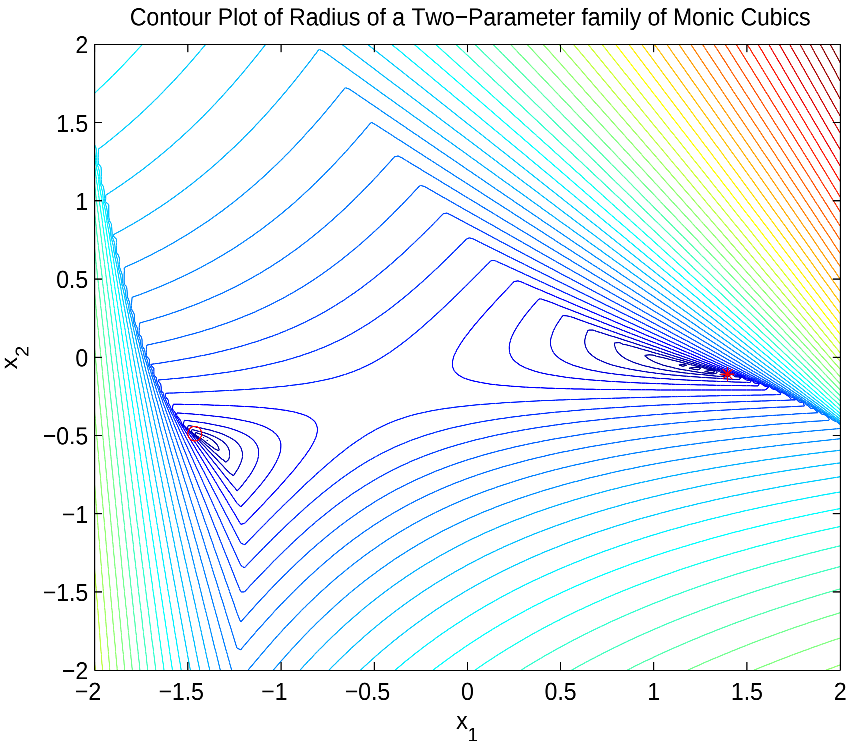


Optimal Root Abscissa for a Quintic Family

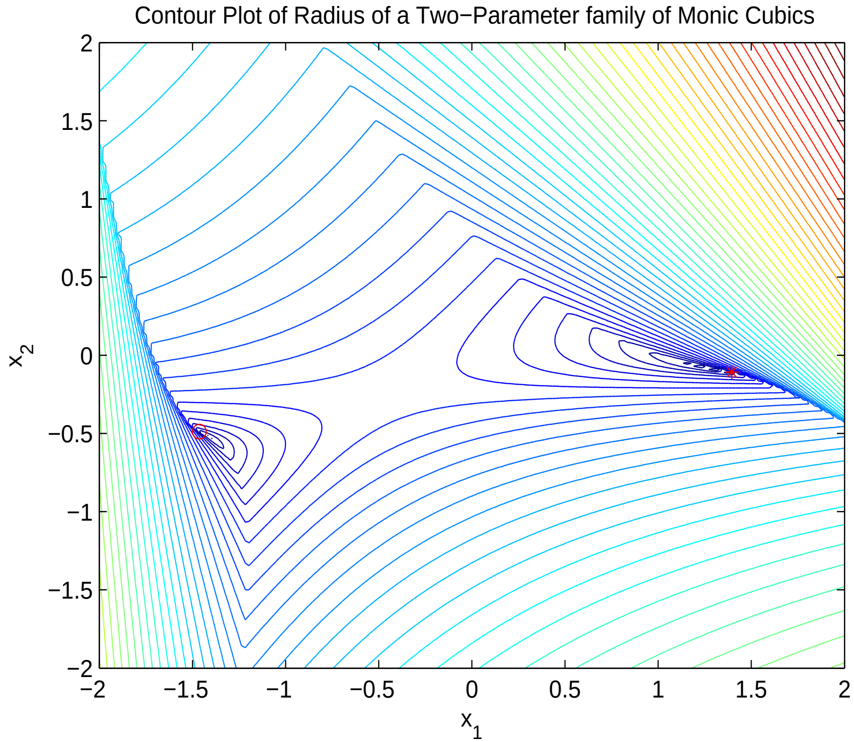


$z^5 + a_1 z^4 + \cdots + a_5$ with constraint $B_0 + \sum_{i=1}^5 a_i = 0$

Optimal Root Radius for a Cubic Family

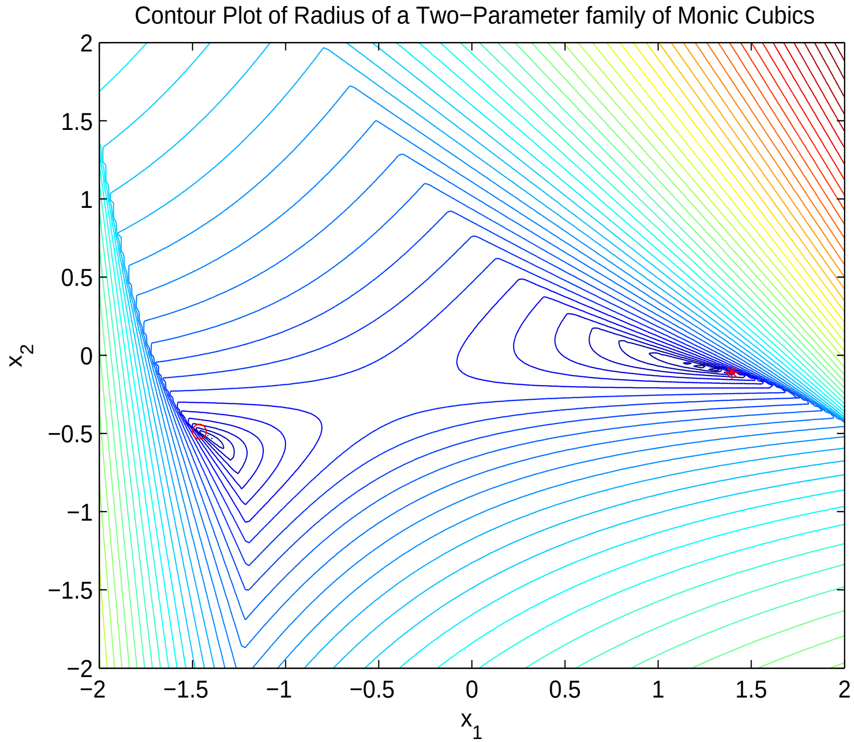


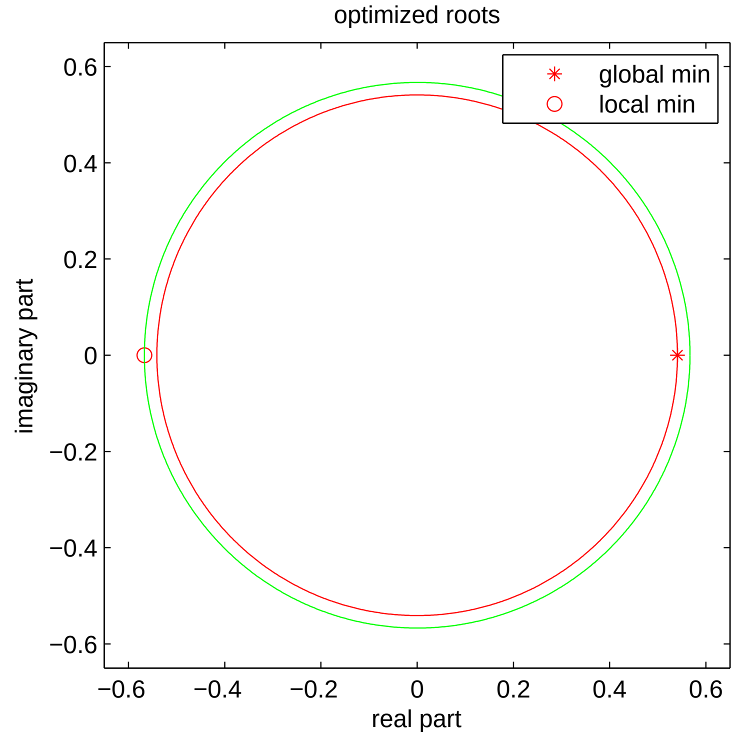
Optimal Root Radius for a Cubic Family



Randomly generated family with coefficients depending affinely on two real parameters x_1, x_2. Above left: contour plot in parameter space. Note the steep contours! '*' marks a global minimizer $(z - \gamma_1)^3$ and 'o' a local minimizer marks a local minimizer $(z - \gamma_2)^3$ where $-\gamma_1 \approx -0.541$ and $-\gamma_2 \approx 0.567$ are *both* roots of g_0 (defined in Corllary RRR).

Optimal Root Radius for a Cubic Family





Randomly generated family with coefficients depending affinely on two real parameters x_1, x_2. Above left: contour plot in parameter space. Note the steep contours! '*' marks a global minimizer $(z-\gamma_1)^3$ and 'o' a local minimizer marks a local minimizer $(z-\gamma_2)^3$ where $-\gamma_1 \approx -0.541$ and $-\gamma_2 \approx 0.567$ are *both* roots of g_0 (defined in Corllary RRR).

Right: complex plane plot of optimal roots. '*' is γ_1 and 'o' is γ_2.

Static Output Feedback Stabilization

Given the dynamical system

$$\dot{x} = Fx + Gu, \quad y = Hx$$

where $F \in \mathbf{R}^{n\times n}, G \in \mathbf{R}^{n\times p}$, $H \in \mathbf{R}^{m\times n}$.

Static Output Feedback Stabilization

Given the dynamical system

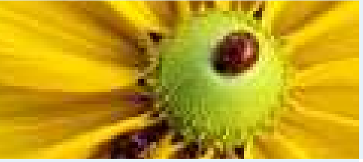
$$\dot{x} = Fx + Gu, \quad y = Hx$$

where $F \in \mathbf{R}^{n \times n}, G \in \mathbf{R}^{n \times p}$, $H \in \mathbf{R}^{m \times n}$.

SOF: find a controller $K \in \mathbf{R}^{p \times m}$ so that, setting $u = Ky$

$$\dot{x} = (F + GKH)x$$

is stable, that is all eigenvalues of $F + GKH$ are in the left half-plane, or prove that this is not possible.



Static Output Feedback Stabilization

Given the dynamical system

$$\dot{x} = Fx + Gu, \quad y = Hx$$

where $F \in \mathbf{R}^{n\times n}, G \in \mathbf{R}^{n\times p}$, $H \in \mathbf{R}^{m\times n}$.

SOF: find a controller $K \in \mathbf{R}^{p\times m}$ so that, setting $u = Ky$

$$\dot{x} = (F + GKH)x$$

is stable, that is all eigenvalues of $F + GKH$ are in the left half-plane, or prove that this is not possible.

A major open problem in control.

Static Output Feedback Stabilization

Given the dynamical system

$$\dot{x} = Fx + Gu, \quad y = Hx$$

where $F \in \mathbf{R}^{n\times n}, G \in \mathbf{R}^{n\times p}$, $H \in \mathbf{R}^{m\times n}$.

SOF: find a controller $K \in \mathbf{R}^{p\times m}$ so that, setting $u = Ky$

$$\dot{x} = (F + GKH)x$$

is stable, that is all eigenvalues of $F + GKH$ are in the left half-plane, or prove that this is not possible.

A major open problem in control.

But, if $p = 1$ and $m = n - 1$ (one input and $n - 1$ outputs)

$$\det(\lambda I - F - GKH) = \det(\lambda I - F) + KH\mathrm{adj}(\lambda I - F)G.$$

This is a monic polynomial with affine dependence on the $n - 1$ entries of $K \in \mathbf{R}^{1\times(n-1)}$ so the SOF problem can be solved explicitly using Theorem RAR.

Frequency Domain Stabilization

Another set of classical problems in control that, in a certain case, can be solved using the theorems given above.

Frequency Domain Stabilization

Another set of classical problems in control that, in a certain case, can be solved using the theorems given above.

An example: stabilizing the two-mass-spring dynamical system by a second-order controller.

Frequency Domain Stabilization

Another set of classical problems in control that, in a certain case, can be solved using the theorems given above.

An example: stabilizing the two-mass-spring dynamical system by a second-order controller.

Then, maximizing the closed-loop asymptotic decay rate is equivalent to solving the optimization problem

$$\min_{p \in P} \max_{z \in \mathbf{C}} \{\mathrm{Re}\, z \ : \ p(z) = 0\}$$

where

$$P = \{(z^4 + 2z^2)(x_0 + x_1 z + z^2) + y_0 + y_1 z + y_2 z^2 \ : x_0, x_1, y_0, y_1, y_2 \in \mathbf{R}\}$$

Frequency Domain Stabilization

Another set of classical problems in control that, in a certain case, can be solved using the theorems given above.

An example: stabilizing the two-mass-spring dynamical system by a second-order controller.

Then, maximizing the closed-loop asymptotic decay rate is equivalent to solving the optimization problem

$$\min_{p \in P} \max_{z \in \mathbf{C}} \{\mathrm{Re}\ z \ : \ p(z) = 0\}$$

where

$$P = \{(z^4+2z^2)(x_0+x_1z+z^2)+y_0+y_1z+y_2z^2 : x_0, x_1, y_0, y_1, y_2 \in \mathbf{R}\}$$

We can minimize the root abscissa explicitly using Theorem RAR as P is a set of monic polynomials with degree 6 whose coefficients depend affinely on 5 real parameters.

Caveats

Multiple roots are very sensitive to perturbation!

Caveats

Multiple roots are very sensitive to perturbation!

A random perturbation of size ϵ to the coefficients of a polynomial with a root that has multiplicity k moves the roots by $O(\epsilon^{1/k})$.

Caveats

Multiple roots are very sensitive to perturbation!

A random perturbation of size ϵ to the coefficients of a polynomial with a root that has multiplicity k moves the roots by $O(\epsilon^{1/k})$.

In practice, might want to locally optimize a more robust measure of stability.

Caveats

Multiple roots are very sensitive to perturbation!

A random perturbation of size ϵ to the coefficients of a polynomial with a root that has multiplicity k moves the roots by $O(\epsilon^{1/k})$.

In practice, might want to locally optimize a more robust measure of stability.

Independently of this, the monomial basis is a poor choice numerically unless the polynomial has very small degree.

Caveats

Multiple roots are very sensitive to perturbation!

A random perturbation of size ϵ to the coefficients of a polynomial with a root that has multiplicity k moves the roots by $O(\epsilon^{1/k})$.

In practice, might want to locally optimize a more robust measure of stability.

Independently of this, the monomial basis is a poor choice numerically unless the polynomial has very small degree.

Nonetheless, the optimal *value* can be computed accurately even if n is fairly large.

Extensions?

Let's define *active roots* as those whose modulus or real part equals ρ^* or α^* respectively: the others are *inactive*.

Extensions?

Let's define *active roots* as those whose modulus or real part equals ρ^* or α^* respectively: the others are *inactive*.

Consider minimizing the radius subject to κ constraints. In this case, we conjecture that there always exists an optimal polynomial with at most $\kappa - 1$ *inactive* roots.

Extensions?

Let's define *active roots* as those whose modulus or real part equals ρ^* or α^* respectively: the others are *inactive*.

Consider minimizing the radius subject to κ constraints. In this case, we conjecture that there always exists an optimal polynomial with at most $\kappa - 1$ *inactive* roots.

However, there does not seem to be a useful bound on the number of possible *distinct* active roots. Thus, computing global optimizers appears to be difficult.

Extensions?

Let's define *active roots* as those whose modulus or real part equals ρ^* or α^* respectively: the others are *inactive*.

Consider minimizing the radius subject to κ constraints. In this case, we conjecture that there always exists an optimal polynomial with at most $\kappa - 1$ *inactive* roots.

However, there does not seem to be a useful bound on the number of possible *distinct* active roots. Thus, computing global optimizers appears to be difficult.

Let's look at some randomly generated radius optimization examples with $\kappa = n - 2$, hence parametrized by two real parameters: the conjecture says that at least $n - \kappa + 1 = 3$ roots are active.

Extensions?

Let's define *active roots* as those whose modulus or real part equals ρ^* or α^* respectively: the others are *inactive*.

Consider minimizing the radius subject to κ constraints. In this case, we conjecture that there always exists an optimal polynomial with at most $\kappa - 1$ *inactive* roots.

However, there does not seem to be a useful bound on the number of possible *distinct* active roots. Thus, computing global optimizers appears to be difficult.

Let's look at some randomly generated radius optimization examples with $\kappa = n - 2$, hence parametrized by two real parameters: the conjecture says that at least $n - \kappa + 1 = 3$ roots are active.

Previous picture was $n = 3, \kappa = 1$, triple root.

Extensions?

Let's define *active roots* as those whose modulus or real part equals ρ^* or α^* respectively: the others are *inactive*.

Consider minimizing the radius subject to κ constraints. In this case, we conjecture that there always exists an optimal polynomial with at most $\kappa - 1$ *inactive* roots.

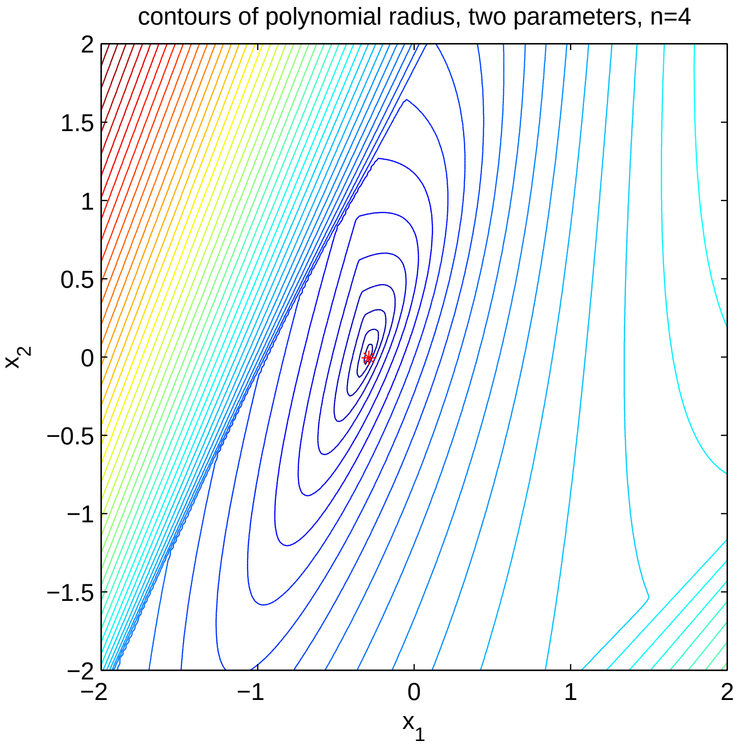
However, there does not seem to be a useful bound on the number of possible *distinct* active roots. Thus, computing global optimizers appears to be difficult.

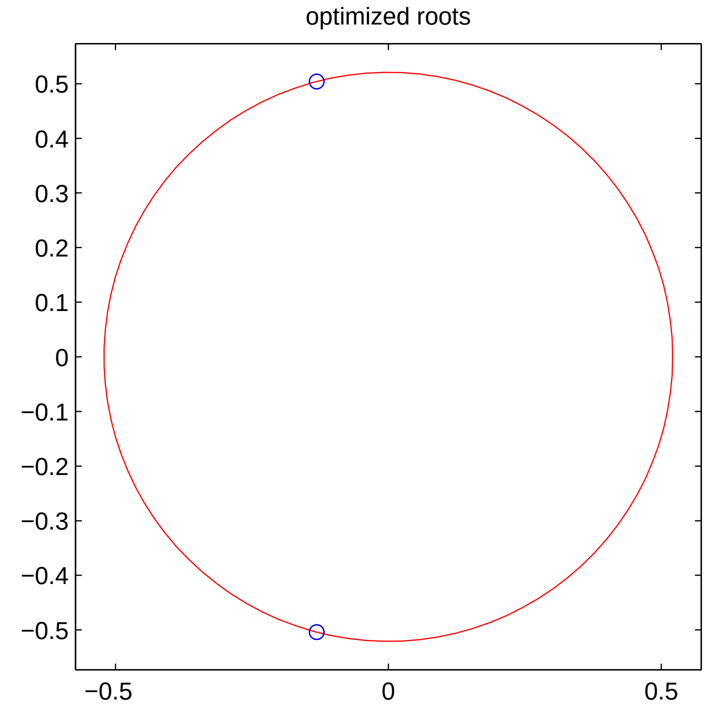
Let's look at some randomly generated radius optimization examples with $\kappa = n - 2$, hence parametrized by two real parameters: the conjecture says that at least $n - \kappa + 1 = 3$ roots are active.

Previous picture was $n = 3, \kappa = 1$, triple root.

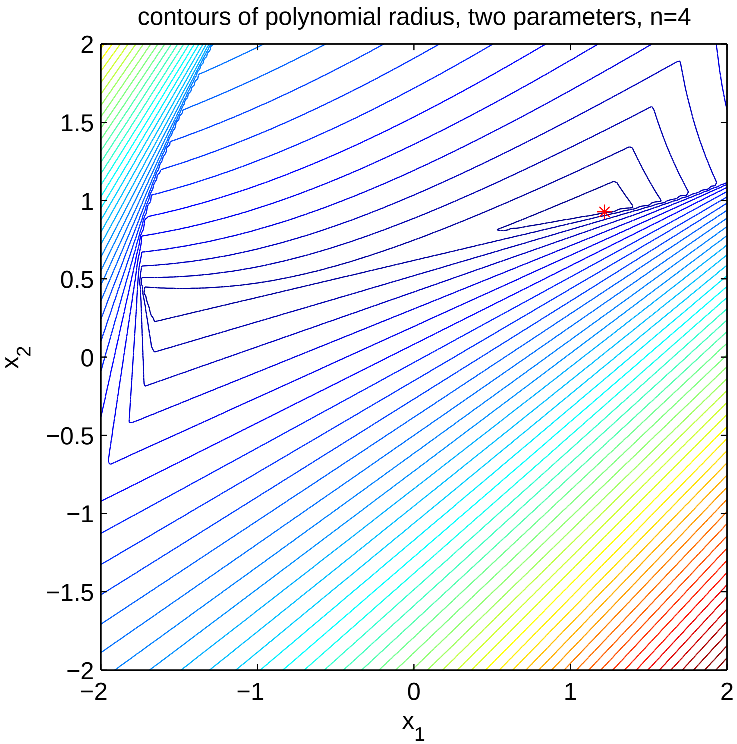
Since there are only two variables, use grid search to approximate global minimizer and then local optimization to refine this.

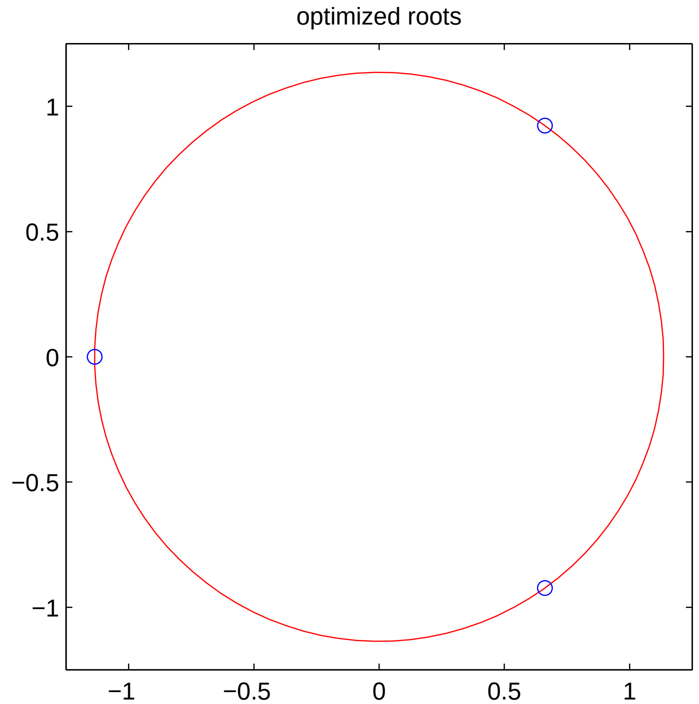
Degree $n = 4$: one double conjugate pair is active



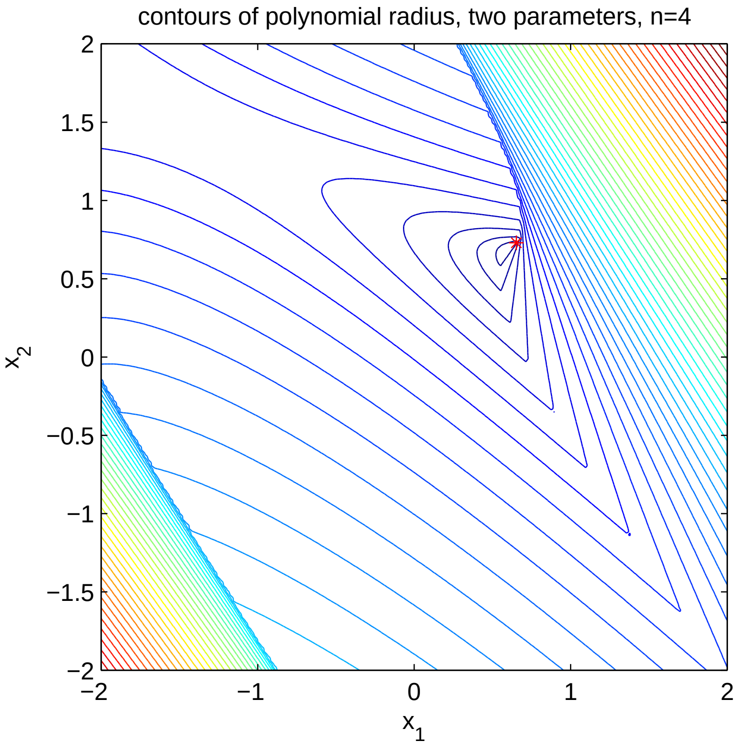


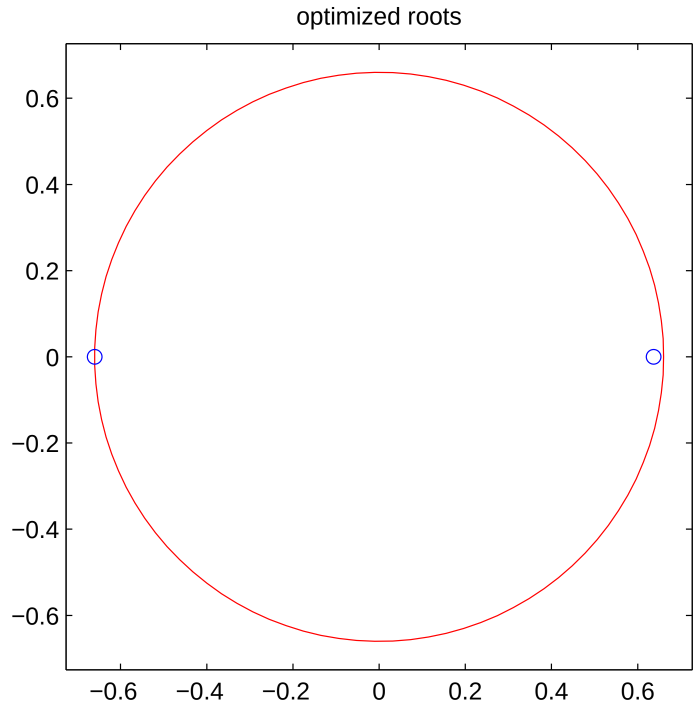
Degree $n = 4$: one double real and one simple conjugate pair are active



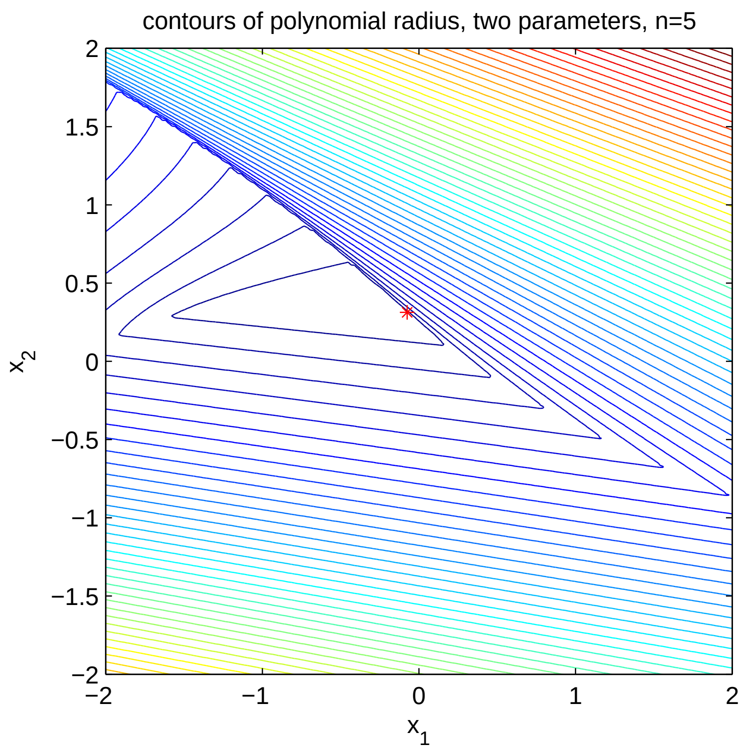


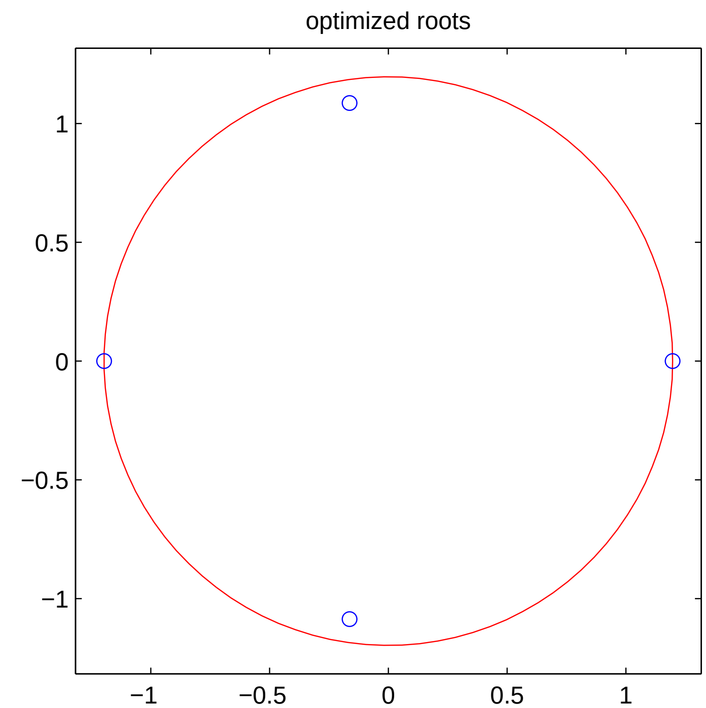
Degree $n = 4$: one triple real root is active



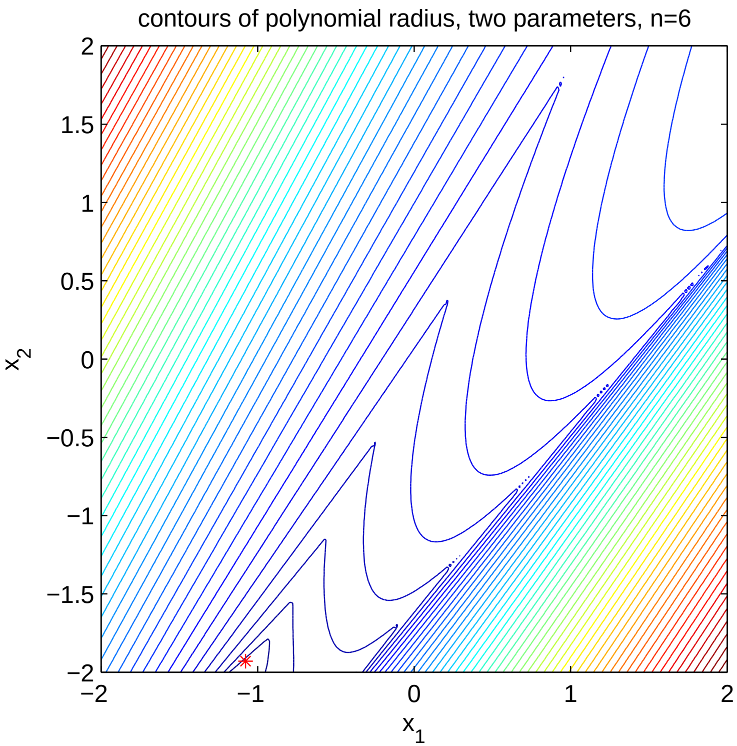


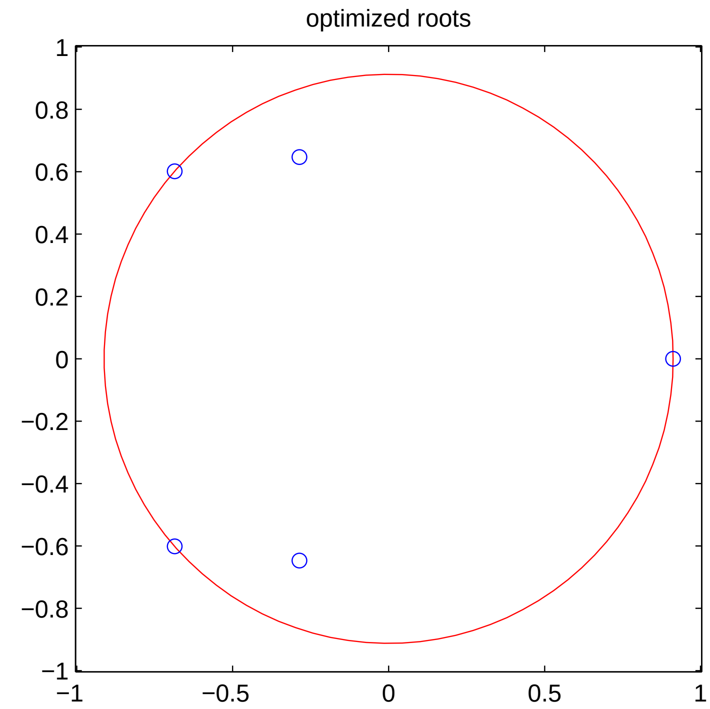
Degree $n = 5$: one double real and one simple real root are active





Degree $n = 6$: one double real and one simple conjugate pair are active





References for Part I

Explicit Solutions for Root Optimization of a Polynomial Family with One Affine Constraint
V.D. Blondel, M. Gürbüzbalaban, A. Megretski, M.L. Overton
In revision for IEEE Trans. Auto. Control.

References for Part I

Explicit Solutions for Root Optimization of a Polynomial Family with One Affine Constraint
V.D. Blondel, M. Gürbüzbalaban, A. Megretski, M.L. Overton
In revision for IEEE Trans. Auto. Control.

Based in part on a remarkable 1979 Ph.D. thesis by Raymond Chen, University of Florida. Chen obtained the infimal result for the abscissa in the real case (first part of Theorem RAR) but did not treat the case where the optimal value is not attained. He also obtained related results for the radius in the real case.

References for Part I

Explicit Solutions for Root Optimization of a Polynomial Family with One Affine Constraint
V.D. Blondel, M. Gürbüzbalaban, A. Megretski, M.L. Overton
In revision for IEEE Trans. Auto. Control.

Based in part on a remarkable 1979 Ph.D. thesis by Raymond Chen, University of Florida. Chen obtained the infimal result for the abscissa in the real case (first part of Theorem RAR) but did not treat the case where the optimal value is not attained. He also obtained related results for the radius in the real case.

The control applications are due to Chen, Rantzer and Henrion.

References for Part I

Explicit Solutions for Root Optimization of a Polynomial Family with One Affine Constraint
V.D. Blondel, M. Gürbüzbalaban, A. Megretski, M.L. Overton
In revision for IEEE Trans. Auto. Control.

Based in part on a remarkable 1979 Ph.D. thesis by Raymond Chen, University of Florida. Chen obtained the infimal result for the abscissa in the real case (first part of Theorem RAR) but did not treat the case where the optimal value is not attained. He also obtained related results for the radius in the real case.

The control applications are due to Chen, Rantzer and Henrion.

A publicly available MATLAB code implementing the constructive algorithms implicit in the theorems:

www.cs.nyu.edu/overton/software/affpoly/

Part II
Optimization of Eigenvalues
with
J.V. Burke (Wash.)
K.K. Gade (NYU)
S. Grundel (NYU)
A.S. Lewis (Cornell)

The Spectral Radius and the Spectral Abscissa

Now let $\rho : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral radius*:

$$\rho(X) = \max\left\{|z| \;:\; \det(X - zI) = 0,\; z \in \mathbf{C}\right\}.$$

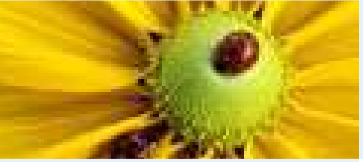
The Spectral Radius and the Spectral Abscissa

Now let $\rho : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral radius*:

$$\rho(X) = \max\{|z| \ : \ \det(X - zI) = 0, \ z \in \mathbf{C}\}.$$

Let $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral abscissa*:

$$\alpha(X) = \max\{\mathrm{Re}(z) \ : \ \det(X - zI) = 0, \ z \in \mathbf{C}\}.$$



The Spectral Radius and the Spectral Abscissa

Now let $\rho : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral radius*:

$$\rho(X) = \max\{|z| \;:\; \det(X - zI) = 0,\; z \in \mathbf{C}\}.$$

Let $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral abscissa*:

$$\alpha(X) = \max\{\mathrm{Re}(z) \;:\; \det(X - zI) = 0,\; z \in \mathbf{C}\}.$$

An eigenvalue is *active* if its modulus (real part) equals the spectral radius (abscissa).

The Spectral Radius and the Spectral Abscissa

Now let $\rho : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral radius*:

$$\rho(X) = \max\{|z| \ : \ \det(X - zI) = 0, \ z \in \mathbf{C}\}.$$

Let $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral abscissa*:

$$\alpha(X) = \max\{\mathrm{Re}(z) \ : \ \det(X - zI) = 0, \ z \in \mathbf{C}\}.$$

An eigenvalue is *active* if its modulus (real part) equals the spectral radius (abscissa).

The spectral functions ρ and α are not convex and are not Lipschitz near a matrix with an active multiple eigenvalue.

The Spectral Radius and the Spectral Abscissa

Now let $\rho : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral radius*:

$$\rho(X) = \max\{|z| \;:\; \det(X - zI) = 0,\; z \in \mathbf{C}\}.$$

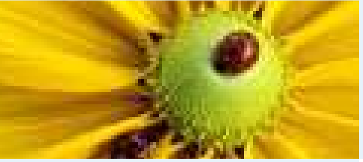
Let $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral abscissa*:

$$\alpha(X) = \max\{\mathrm{Re}(z) \;:\; \det(X - zI) = 0,\; z \in \mathbf{C}\}.$$

An eigenvalue is *active* if its modulus (real part) equals the spectral radius (abscissa).

The spectral functions ρ and α are not convex and are not Lipschitz near a matrix with an active multiple eigenvalue.

An eigenvalue is *nonderogatory* if it is associated with only one right eigenvector, that is its *geometric multiplicity* is 1.



The Spectral Radius and the Spectral Abscissa

Now let $\rho : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral radius*:

$$\rho(X) = \max\{|z| \;:\; \det(X - zI) = 0,\; z \in \mathbf{C}\}.$$

Let $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral abscissa*:

$$\alpha(X) = \max\{\mathrm{Re}(z) \;:\; \det(X - zI) = 0,\; z \in \mathbf{C}\}.$$

An eigenvalue is *active* if its modulus (real part) equals the spectral radius (abscissa).

The spectral functions ρ and α are not convex and are not Lipschitz near a matrix with an active multiple eigenvalue.

An eigenvalue is *nonderogatory* if it is associated with only one right eigenvector, that is its *geometric multiplicity* is 1.

Equivalently, it is associated with a single block in the Jordan form, whose dimension is the *(algebraic) multiplicity*.

The Spectral Radius and the Spectral Abscissa

Now let $\rho : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral radius*:

$$\rho(X) = \max\{|z| \ : \ \det(X - zI) = 0, \ z \in \mathbf{C}\}.$$

Let $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral abscissa*:

$$\alpha(X) = \max\{\mathrm{Re}(z) \ : \ \det(X - zI) = 0, \ z \in \mathbf{C}\}.$$

An eigenvalue is *active* if its modulus (real part) equals the spectral radius (abscissa).

The spectral functions ρ and α are not convex and are not Lipschitz near a matrix with an active multiple eigenvalue.

An eigenvalue is *nonderogatory* if it is associated with only one right eigenvector, that is its *geometric multiplicity* is 1.

Equivalently, it is associated with a single block in the Jordan form, whose dimension is the *(algebraic) multiplicity*.

Hermitian matrices: completely different story!

No Extension of Part I

The spectral radius and abscissa are the radius and abscissa of the characteristic polynomial of a matrix, but the results of Part I do not extend to the more general case of an affine family of $n \times n$ matrices depending on $n - 1$ parameters.

No Extension of Part I

The spectral radius and abscissa are the radius and abscissa of the characteristic polynomial of a matrix, but the results of Part I do not extend to the more general case of an affine family of $n \times n$ matrices depending on $n-1$ parameters.

For example, consider the matrix family

$$A(\xi) = \begin{bmatrix} \xi & 1 \\ -1 & \xi \end{bmatrix}.$$

This matrix depends affinely on a single parameter ξ, but its characteristic polynomial, a monic polynomial of degree 2, does not, so the results of Part I do not apply.

No Extension of Part I

The spectral radius and abscissa are the radius and abscissa of the characteristic polynomial of a matrix, but the results of Part I do not extend to the more general case of an affine family of $n \times n$ matrices depending on $n-1$ parameters.

For example, consider the matrix family

$$A(\xi) = \begin{bmatrix} \xi & 1 \\ -1 & \xi \end{bmatrix}.$$

This matrix depends affinely on a single parameter ξ, but its characteristic polynomial, a monic polynomial of degree 2, does not, so the results of Part I do not apply.

The minimal spectral radius of $A(\xi)$ is attained by $\xi = 0$, for which the eigenvalues are $\pm\mathbf{i}$.

The Diaconis-Holmes-Neal Sampler

A nonreversable Markov chain for Monte Carlo simulation. For $x \in (0,1)$, the transition matrix is $A(x) \in \mathbf{R}^{2n\times 2n}$ is

$$
\begin{bmatrix}
0 & 1-x & & & & & & & x & 0 \\
 & & 1-x & & & & & x & & \\
 & & & \ddots & & & \iddots & & & \\
 & & & & 1-x & x & & & & \\
 & & & & x & 1-x & & & & \\
 & & & \iddots & & & \ddots & & & \\
 & & x & & & & & 1-x & & \\
 & x & & & & & & & 1-x & \\
x & & & & & & & & & 1-x \\
1-x & & & & & & & & & x
\end{bmatrix}.
$$

The Diaconis-Holmes-Neal Sampler

A nonreversable Markov chain for Monte Carlo simulation. For $x \in (0,1)$, the transition matrix is $A(x) \in \mathbf{R}^{2n \times 2n}$ is

$$\begin{bmatrix}
0 & 1-x & & & & & & & x & 0 \\
 & & 1-x & & & & & x & & \\
 & & & \ddots & & & \iddots & & & \\
 & & & & 1-x & x & & & & \\
 & & & & x & 1-x & & & & \\
 & & & \iddots & & & \ddots & & & \\
 & & x & & & & & 1-x & & \\
 & x & & & & & & & 1-x & \\
x & & & & & & & & & 1-x \\
1-x & & & & & & & & & x
\end{bmatrix}.$$

Diaconis et. al. showed that for $x = 1/n$, the corresponding nonreversible chain reaches a stationary state in $O(n)$ steps, compared to $O(n^2)$ steps for a similar reversible chain.

The Reduced Spectral Radius

The rate of convergence is determined by

$$\tilde{\rho}(A(x)) = \max\{|z| \;:\; \det(A(x) - zI) = 0,\; z \in \mathbf{C}, z \neq 1\}.$$

It is easy to prove that this is minimized over $x \in [0,1]$ by

$$x_{\mathrm{opt}} = \frac{\sin(\pi/n)}{1 + \sin(\pi/n)} > \frac{1}{n}.$$

The Reduced Spectral Radius

The rate of convergence is determined by

$$\tilde{\rho}(A(x)) = \max\{|z| \;:\; \det(A(x) - zI) = 0,\; z \in \mathbf{C}, z \neq 1\}.$$

It is easy to prove that this is minimized over $x \in [0,1]$ by

$$x_{\text{opt}} = \frac{\sin(\pi/n)}{1 + \sin(\pi/n)} > \frac{1}{n}.$$

For $x < x_{\text{opt}}$, the active eigenvalues all occur in conjugate pairs.

The Reduced Spectral Radius

The rate of convergence is determined by

$$\tilde{\rho}(A(x)) = \max\{|z| \; : \; \det(A(x) - zI) = 0, \; z \in \mathbf{C}, z \neq 1\}.$$

It is easy to prove that this is minimized over $x \in [0, 1]$ by

$$x_{\text{opt}} = \frac{\sin(\pi/n)}{1 + \sin(\pi/n)} > \frac{1}{n}.$$

For $x < x_{\text{opt}}$, the active eigenvalues all occur in conjugate pairs.

For $x = x_{\text{opt}}$, one conjugate pair has coalesced to a double real nonderogatory eigenvalue (with a 2×2 Jordan block).

The Reduced Spectral Radius

The rate of convergence is determined by

$$\tilde{\rho}(A(x)) = \max\left\{|z| \ : \ \det(A(x) - zI) = 0, \ z \in \mathbf{C}, z \neq 1\right\}.$$

It is easy to prove that this is minimized over $x \in [0,1]$ by

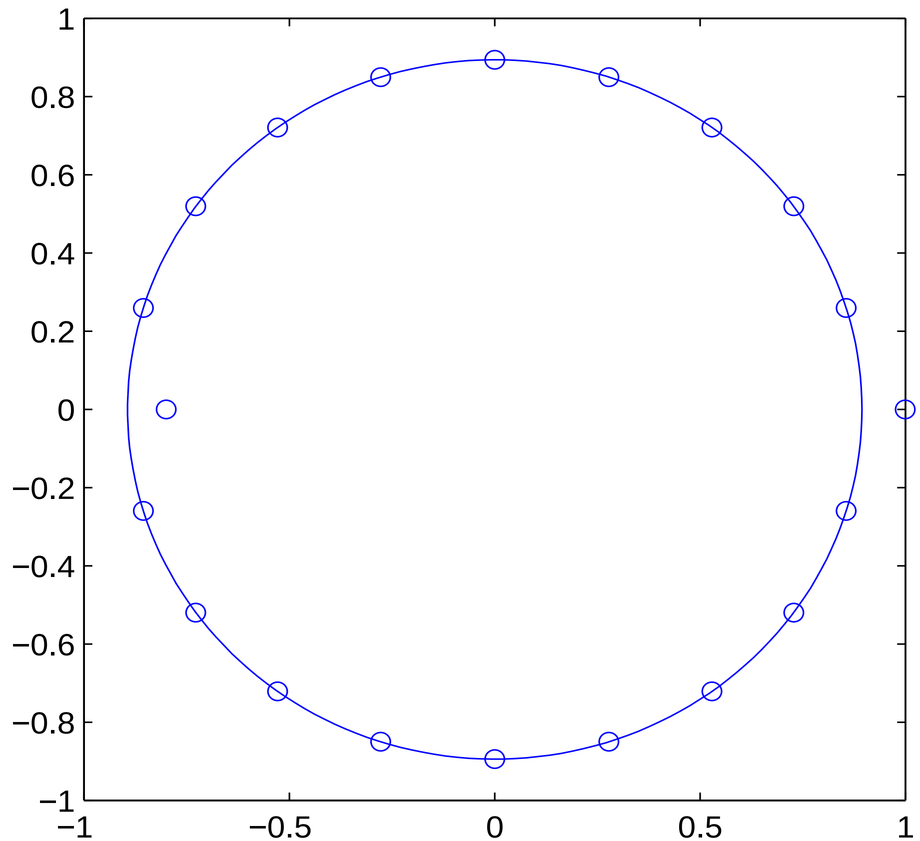
$$x_{\rm opt} = \frac{\sin(\pi/n)}{1+\sin(\pi/n)} > \frac{1}{n}.$$

For $x < x_{\rm opt}$, the active eigenvalues all occur in conjugate pairs.

For $x = x_{\rm opt}$, one conjugate pair has coalesced to a double real nonderogatory eigenvalue (with a 2×2 Jordan block).

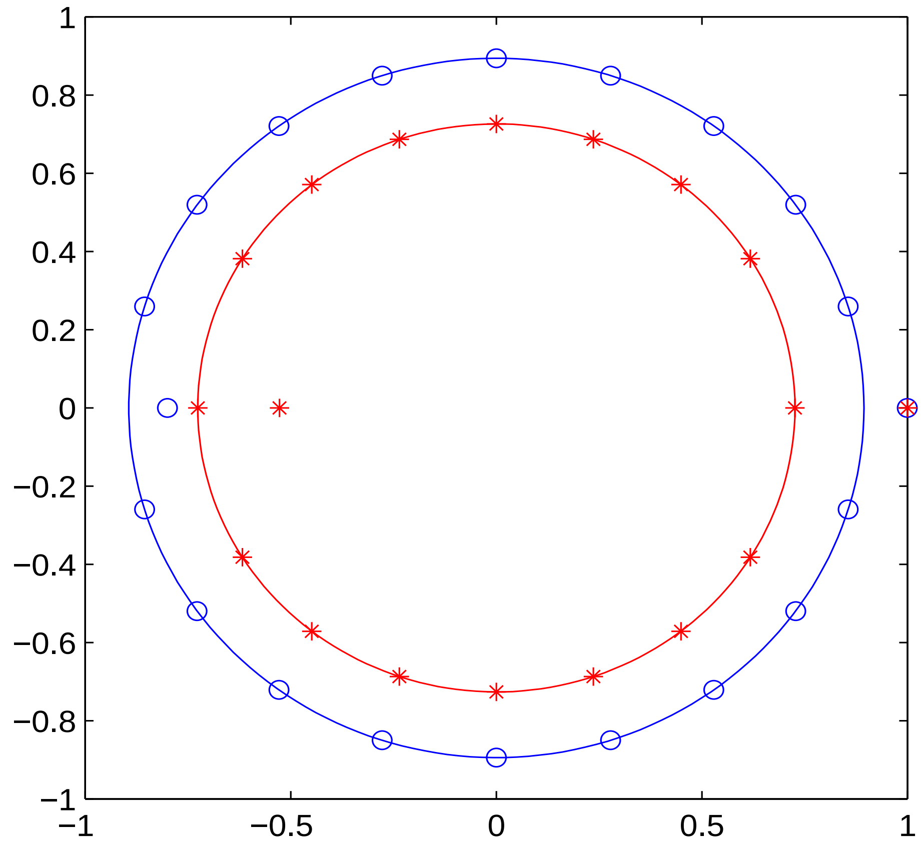
For $x > x_{\rm opt}$, this splits into two real eigenvalues, increasing $\tilde{\rho}$ by $O(|x - x_{\rm opt}|^{1/2})$.

Eigenvalues of the Transition Matrix, $n = 10$



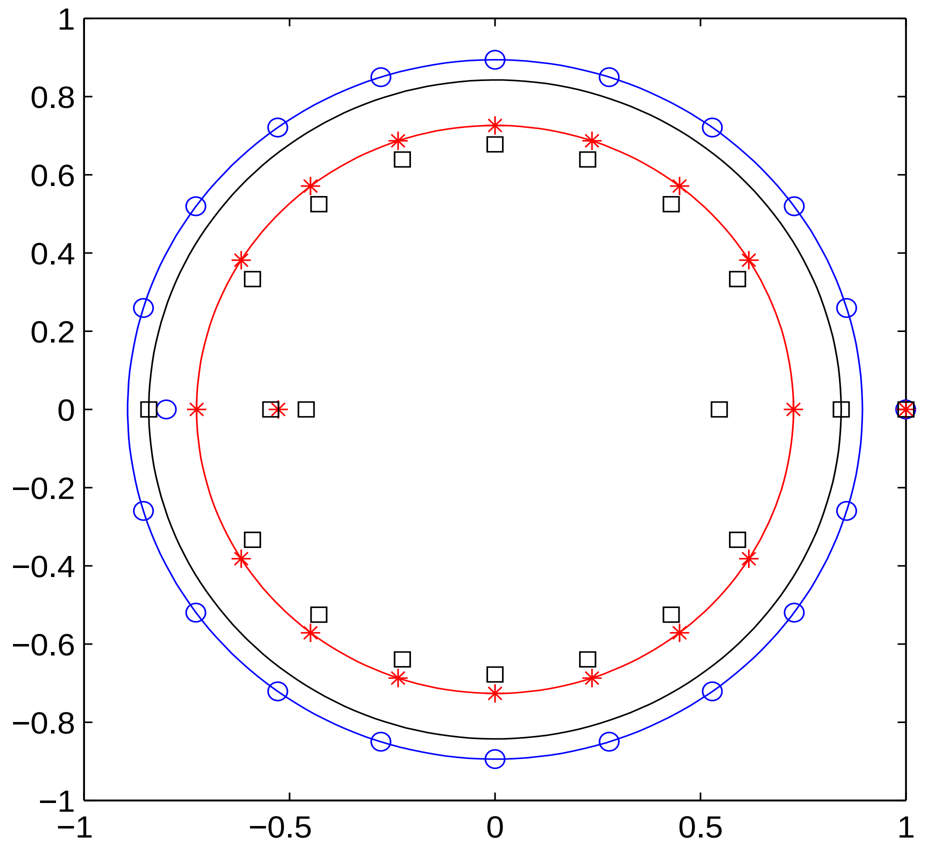
- Blue: eigenvalues when $x = 1/n$ (all complex)

Eigenvalues of the Transition Matrix, $n = 10$



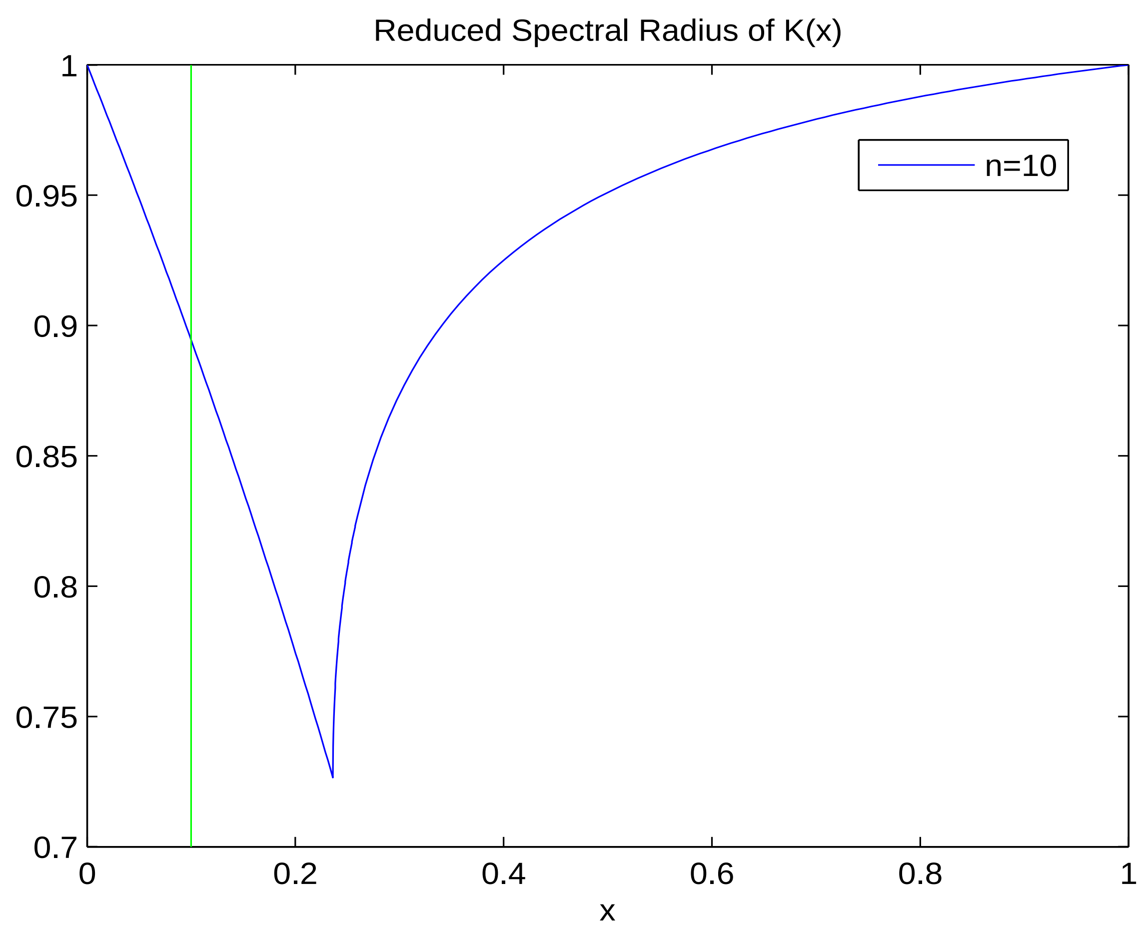
- Blue: eigenvalues when $x = 1/n$ (all complex)
- Red: eigenvalues when $x = x_{\mathrm{opt}}$ (one double real eigenvalue)

Eigenvalues of the Transition Matrix, $n = 10$

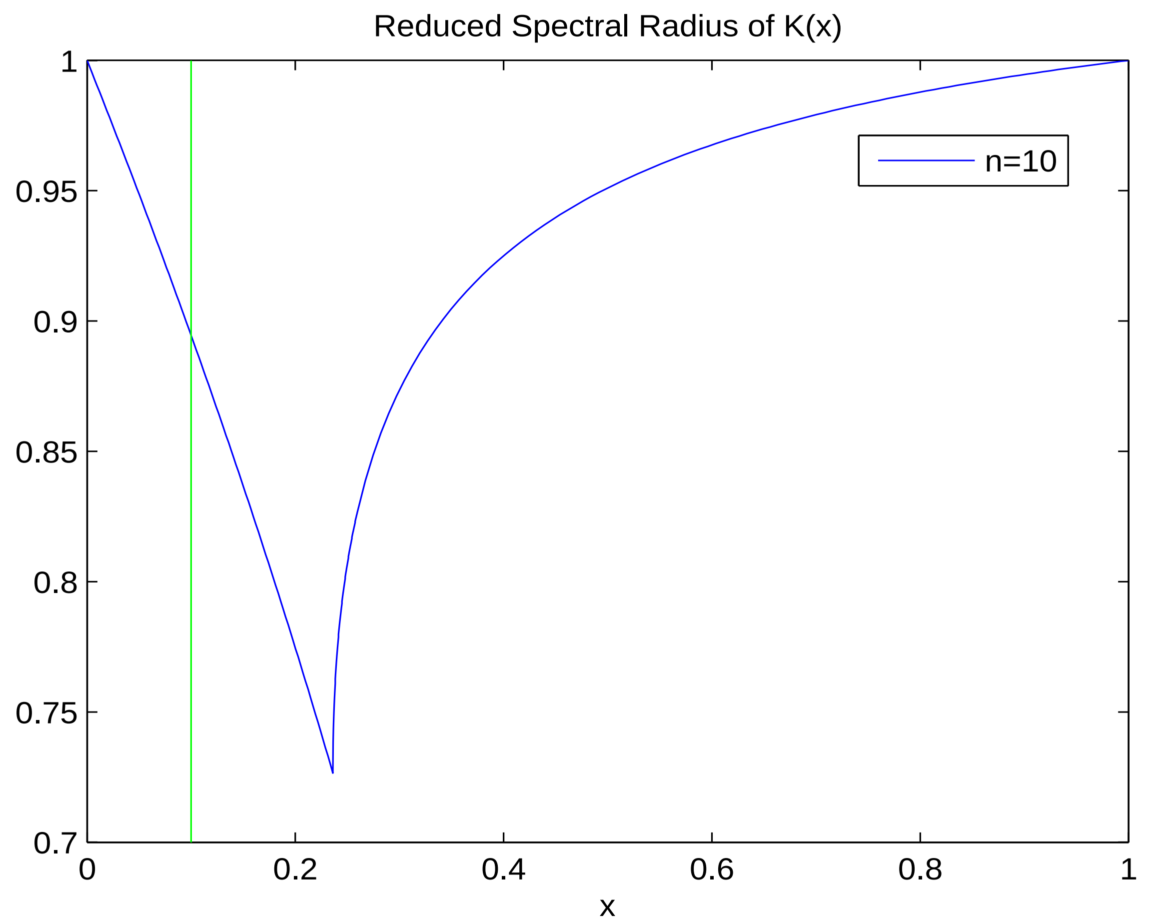


- Blue: eigenvalues when $x = 1/n$ (all complex)
- Red: eigenvalues when $x = x_{\text{opt}}$ (one double real eigenvalue)
- Black: eigenvalues when $x > x_{\text{opt}}$ ($\tilde{\rho}$ increases rapidly)

Reduced Spectral Radius as a Function of x

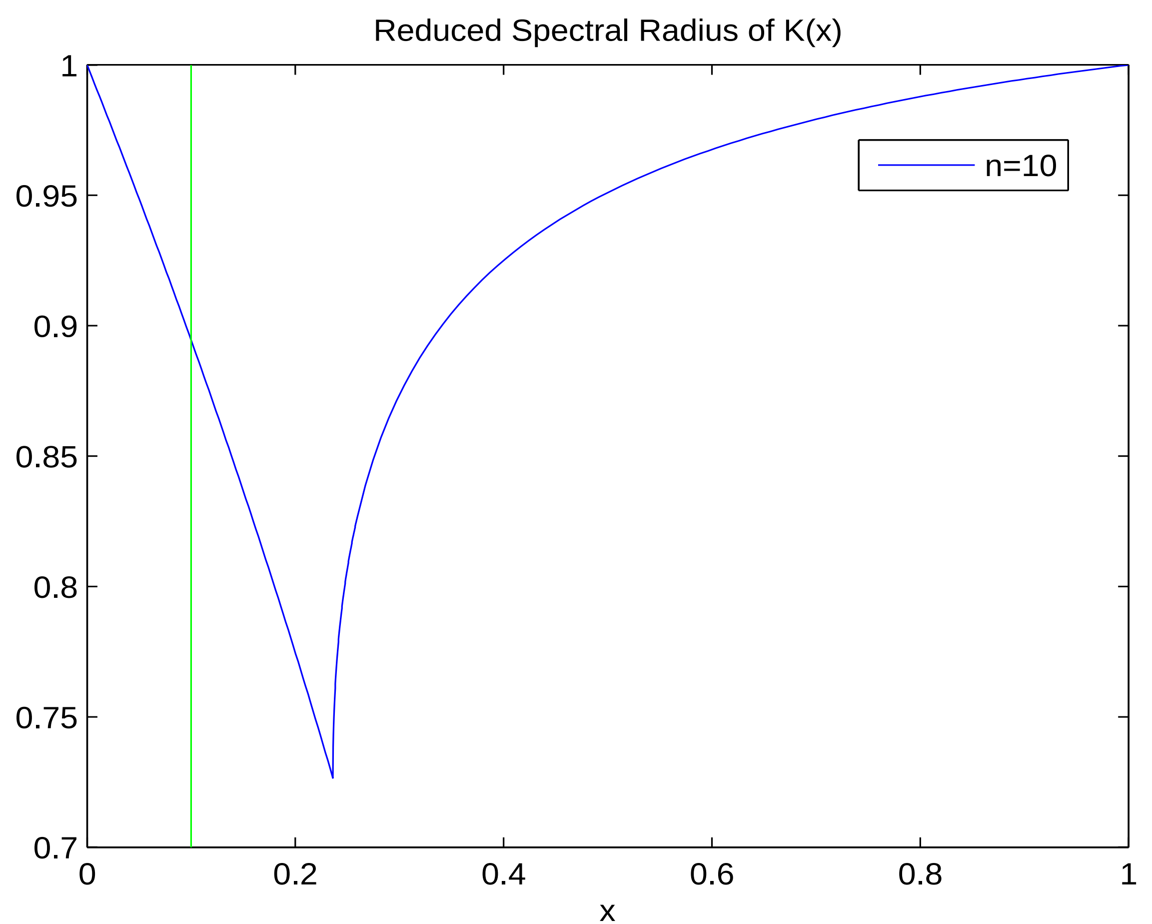


Reduced Spectral Radius as a Function of x



Note the big improvement changing x from $1/n$ to x_{opt}.

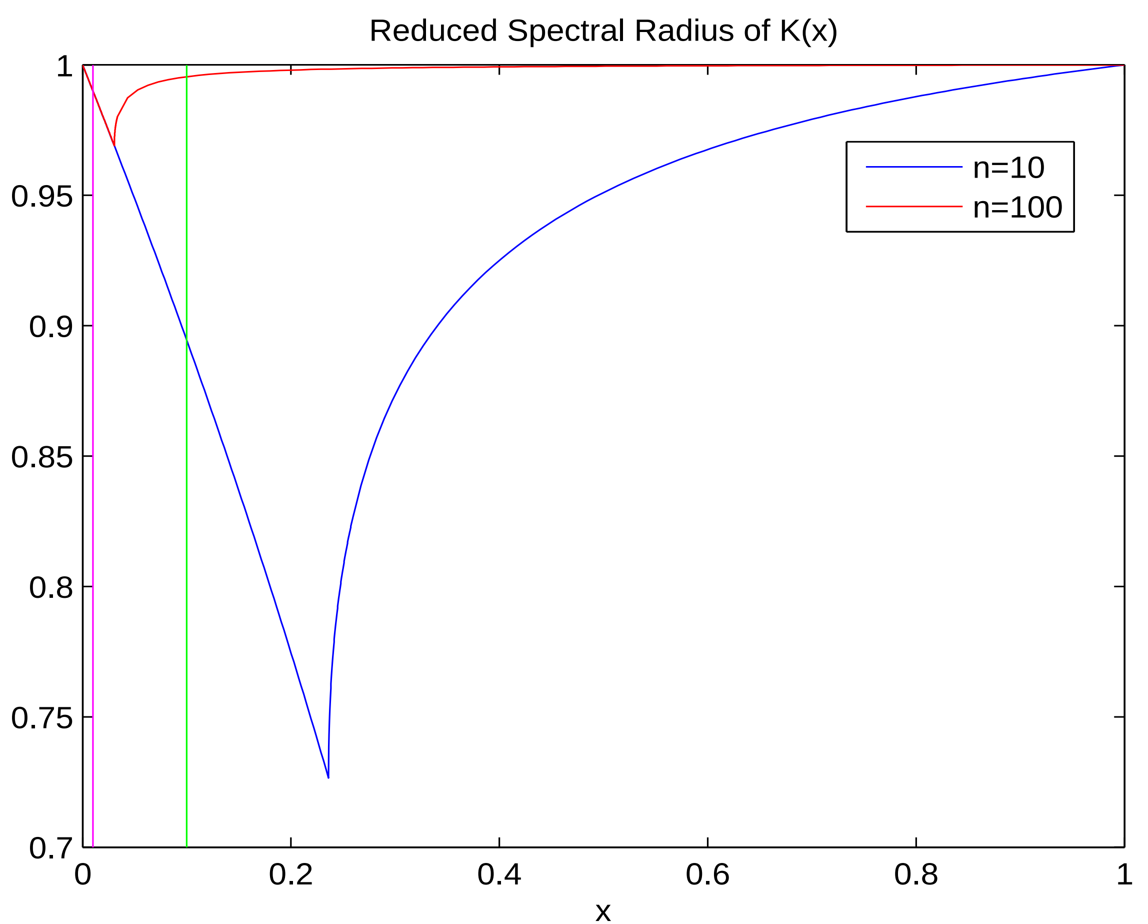
Reduced Spectral Radius as a Function of x



Note the big improvement changing x from $1/n$ to x_{opt}.

Much better to underestimate x_{opt} than overestimate. Similar plots apply to optimal damping for one-dimensional wave equation, optimal choice of parameter for SOR (successive over-relaxation), etc etc.

Reduced Spectral Radius as a Function of x



Note the big improvement changing x from $1/n$ to x_{opt}.

Much better to underestimate x_{opt} than overestimate. Similar plots apply to optimal damping for one-dimensional wave equation, optimal choice of parameter for SOR (successive over-relaxation), etc etc.

Convergence rate deteriorates as n increases.

Adding More Parameters

Not surprising that with one free parameter, we can only make one pair of eigenvalues coalesce.

Adding More Parameters

Not surprising that with one free parameter, we can only make one pair of eigenvalues coalesce.

Let's change $K(x)$ to have multiple parameters:

$$\begin{bmatrix}
0 & 1-x_1 & & & & & & & x_1 & 0 \\
 & & 1-x_2 & & & & & x_2 & & \\
 & & & \ddots & & & \iddots & & & \\
 & & & & 1-x_{n-1} & x_{n-1} & & & & \\
 & & & & x_{n-1} & 1-x_{n-1} & & & & \\
 & & & \iddots & & & \ddots & & & \\
 & & x_2 & & & & & 1-x_2 & & \\
0 & x_1 & & & & & & & 1-x_1 & 0 \\
x_n & & & & & & & & & 1-x_n \\
1-x_n & & & & & & & & & x_n
\end{bmatrix}.$$

Adding More Parameters

Not surprising that with one free parameter, we can only make one pair of eigenvalues coalesce.

Let's change $K(x)$ to have multiple parameters:

$$
\begin{bmatrix}
0 & 1-x_1 & & & & & & & x_1 & 0 \\
 & & 1-x_2 & & & & & x_2 & & \\
 & & & \ddots & & & \cdot^{\cdot^{\cdot}} & & & \\
 & & & & 1-x_{n-1} & x_{n-1} & & & & \\
 & & & & x_{n-1} & 1-x_{n-1} & & & & \\
 & & & \cdot^{\cdot^{\cdot}} & & & \ddots & & & \\
 & & x_2 & & & & & 1-x_2 & & \\
0 & x_1 & & & & & & & 1-x_1 & 0 \\
x_n & & & & & & & & & 1-x_n \\
1-x_n & & & & & & & & & x_n
\end{bmatrix}.
$$

Still doubly stochastic. Can we now reduce $\tilde{\rho}$ further?

Adding More Parameters

Not surprising that with one free parameter, we can only make one pair of eigenvalues coalesce.

Let's change $K(x)$ to have multiple parameters:

$$\begin{bmatrix}
0 & 1-x_1 & & & & & & & x_1 & 0 \\
 & & 1-x_2 & & & & & x_2 & & \\
 & & & \ddots & & & \iddots & & & \\
 & & & & 1-x_{n-1} & x_{n-1} & & & & \\
 & & & & x_{n-1} & 1-x_{n-1} & & & & \\
 & & & \iddots & & & \ddots & & & \\
 & & x_2 & & & & & 1-x_2 & & \\
0 & x_1 & & & & & & & 1-x_1 & 0 \\
x_n & & & & & & & & & 1-x_n \\
1-x_n & & & & & & & & & x_n
\end{bmatrix}.$$

Still doubly stochastic. Can we now reduce $\tilde{\rho}$ further?

No! It appears that $\mathbf{x}_{\text{opt}} = [x_{\text{opt}}, \ldots, x_{\text{opt}}]^T$ is locally optimal.

Checking Local Optimality

Numerically: by running an optimization method suitable for nonsmooth objectives at randomly generated points near $\mathbf{x}_{\mathrm{opt}}$. We used the "gradient sampling" method, and repeatedly obtained convergence to $\mathbf{x}_{\mathrm{opt}}$.

Checking Local Optimality

Numerically: by running an optimization method suitable for nonsmooth objectives at randomly generated points near $\mathbf{x}_{\text{opt}}$. We used the "gradient sampling" method, and repeatedly obtained convergence to $\mathbf{x}_{\text{opt}}$.

Theoretically: by subdifferential analysis. We found that

- $\mathbf{x}_{\text{opt}}$ satisfies a *necessary* condition for local optimality
- if we remove some redundancy by setting $x_j = x_{n-1-j}$ for $j = 1, 2, \ldots, \lfloor \frac{n-1}{2} \rfloor$ and $x_{n-1} = x_n$, we find $\mathbf{x}_{\text{opt}}$ satisfies a *sufficient* condition for local optimality.

Checking Local Optimality

Numerically: by running an optimization method suitable for nonsmooth objectives at randomly generated points near $\mathbf{x}_{\rm opt}$. We used the “gradient sampling” method, and repeatedly obtained convergence to $\mathbf{x}_{\rm opt}$.

Theoretically: by subdifferential analysis. We found that

- $\mathbf{x}_{\rm opt}$ satisfies a *necessary* condition for local optimality
- if we remove some redundancy by setting $x_j = x_{n-1-j}$ for $j = 1, 2, \ldots, \lfloor \frac{n-1}{2} \rfloor$ and $x_{n-1} = x_n$, we find $\mathbf{x}_{\rm opt}$ satisfies a *sufficient* condition for local optimality.

Because the analysis of this example is complicated, we explain the concepts on a much simpler example.

Subdifferential Analysis

Given a Euclidean space $\mathbf{E}$ and a continuous function $f : \mathbf{E} \to \mathbf{R}$, the *regular subdifferential* of f at a point x in $\mathbf{E}$, denoted $\hat{\partial} f(x)$, is

$$\{y \in \mathbf{E} : f(x+z) - f(x) \geq \langle y, z \rangle + o(\|z\|) \ \text{ for small } z \in \mathbf{E}\}.$$

Subdifferential Analysis

Given a Euclidean space $\mathbf{E}$ and a continuous function $f : \mathbf{E} \to \mathbf{R}$, the *regular subdifferential* of f at a point x in $\mathbf{E}$, denoted $\hat{\partial} f(x)$, is

$$\{y \in \mathbf{E} : f(x+z) - f(x) \geq \langle y, z \rangle + o(\|z\|) \ \text{ for small } z \in \mathbf{E}\}.$$

The *subdifferential* $\partial f(x)$ is

$$\left\{ y \in \mathbf{E} : \exists x_k, y_k \in \mathbf{E} \text{ with } x_k \to x, \quad y_k \to y, \quad y_k \in \hat{\partial} f(x_k) \right\}.$$

Subdifferential Analysis

Given a Euclidean space $\mathbf{E}$ and a continuous function $f : \mathbf{E} \to \mathbf{R}$, the *regular subdifferential* of f at a point x in $\mathbf{E}$, denoted $\hat{\partial} f(x)$, is

$$\{y \in \mathbf{E} : f(x+z) - f(x) \geq \langle y, z \rangle + o(\|z\|) \text{ for small } z \in \mathbf{E}\}.$$

The *subdifferential* $\partial f(x)$ is

$$\left\{y \in \mathbf{E} : \exists x_k, y_k \in \mathbf{E} \text{ with } x_k \to x, \quad y_k \to y, \quad y_k \in \hat{\partial} f(x_k)\right\}.$$

The *horizon subdifferential* $\partial^\infty f(x)$ is

$$\left\{y \in \mathbf{E} : \exists t_k \in \mathbf{R}, x_k, y_k \in \mathbf{E} \text{ with } t_k \downarrow 0, x_k \to x, t_k y_k \to y, y_k \in \hat{\partial} f(x_k)\right\}.$$

Subdifferential Analysis

Given a Euclidean space $\mathbf{E}$ and a continuous function $f : \mathbf{E} \to \mathbf{R}$, the *regular subdifferential* of f at a point x in $\mathbf{E}$, denoted $\hat{\partial} f(x)$, is

$$\{y \in \mathbf{E} : f(x+z) - f(x) \geq \langle y, z \rangle + o(\|z\|) \ \text{ for small } z \in \mathbf{E}\}.$$

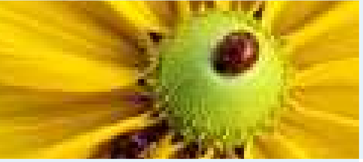
The *subdifferential* $\partial f(x)$ is

$$\left\{y \in \mathbf{E} : \exists x_k, y_k \in \mathbf{E} \text{ with } x_k \to x, \quad y_k \to y, \quad y_k \in \hat{\partial} f(x_k)\right\}.$$

The *horizon subdifferential* $\partial^\infty f(x)$ is

$$\left\{y \in \mathbf{E} : \exists t_k \in \mathbf{R}, x_k, y_k \in \mathbf{E} \text{ with } t_k \downarrow 0, x_k \to x, t_k y_k \to y, y_k \in \hat{\partial} f(x_k)\right\}.$$

Clearly, $\partial(x) \supseteq \hat{\partial} f(x)$ and $0 \in \partial^\infty f(x)$.



Subdifferential Analysis

Given a Euclidean space $\mathbf{E}$ and a continuous function $f : \mathbf{E} \to \mathbf{R}$, the *regular subdifferential* of f at a point x in $\mathbf{E}$, denoted $\hat{\partial} f(x)$, is

$$\{y \in \mathbf{E} : f(x+z) - f(x) \geq \langle y, z \rangle + o(\|z\|) \ \text{ for small } z \in \mathbf{E}\}.$$

The *subdifferential* $\partial f(x)$ is

$$\left\{y \in \mathbf{E} : \exists x_k, y_k \in \mathbf{E} \text{ with } x_k \to x, \quad y_k \to y, \quad y_k \in \hat{\partial} f(x_k)\right\}.$$

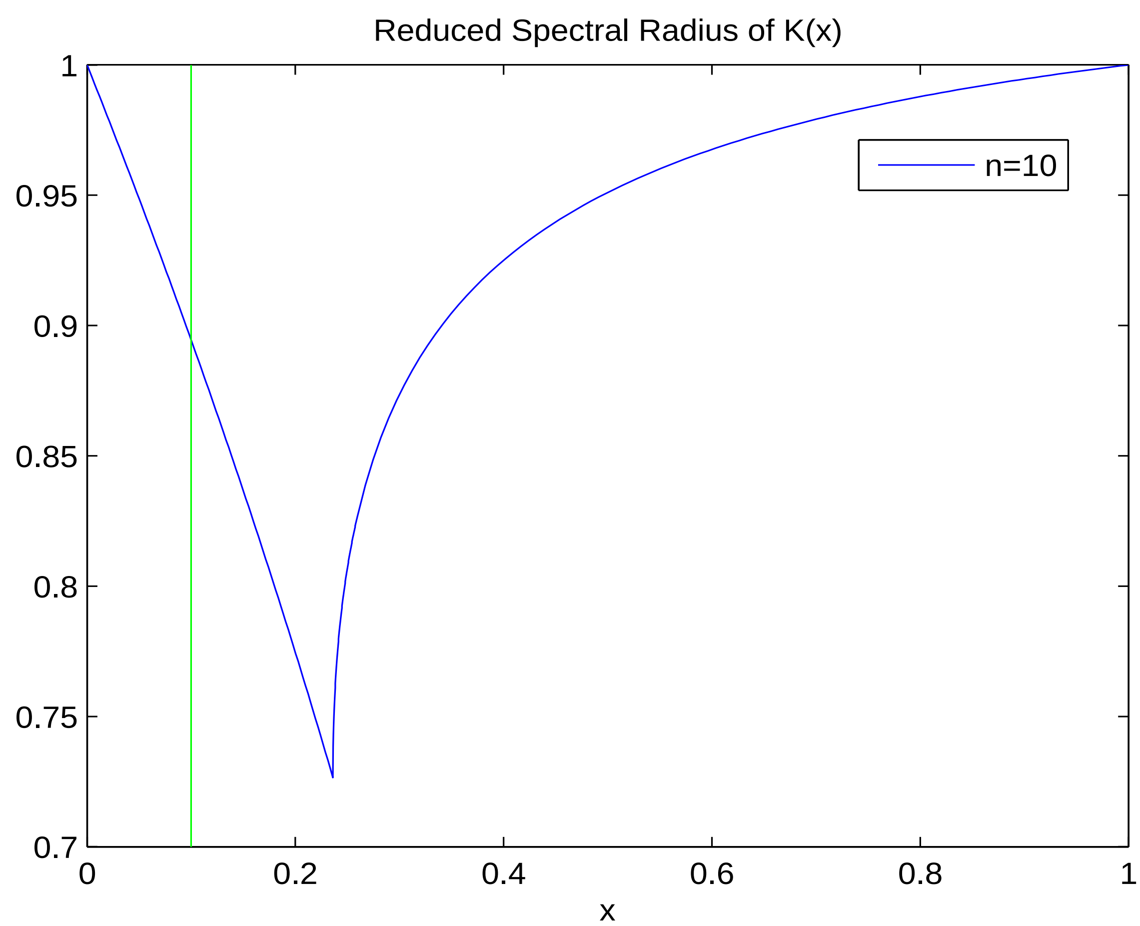
The *horizon subdifferential* $\partial^\infty f(x)$ is

$$\left\{y \in \mathbf{E} : \exists t_k \in \mathbf{R}, x_k, y_k \in \mathbf{E} \text{ with } t_k \downarrow 0, x_k \to x, t_k y_k \to y, y_k \in \hat{\partial} f(x_k)\right\}.$$

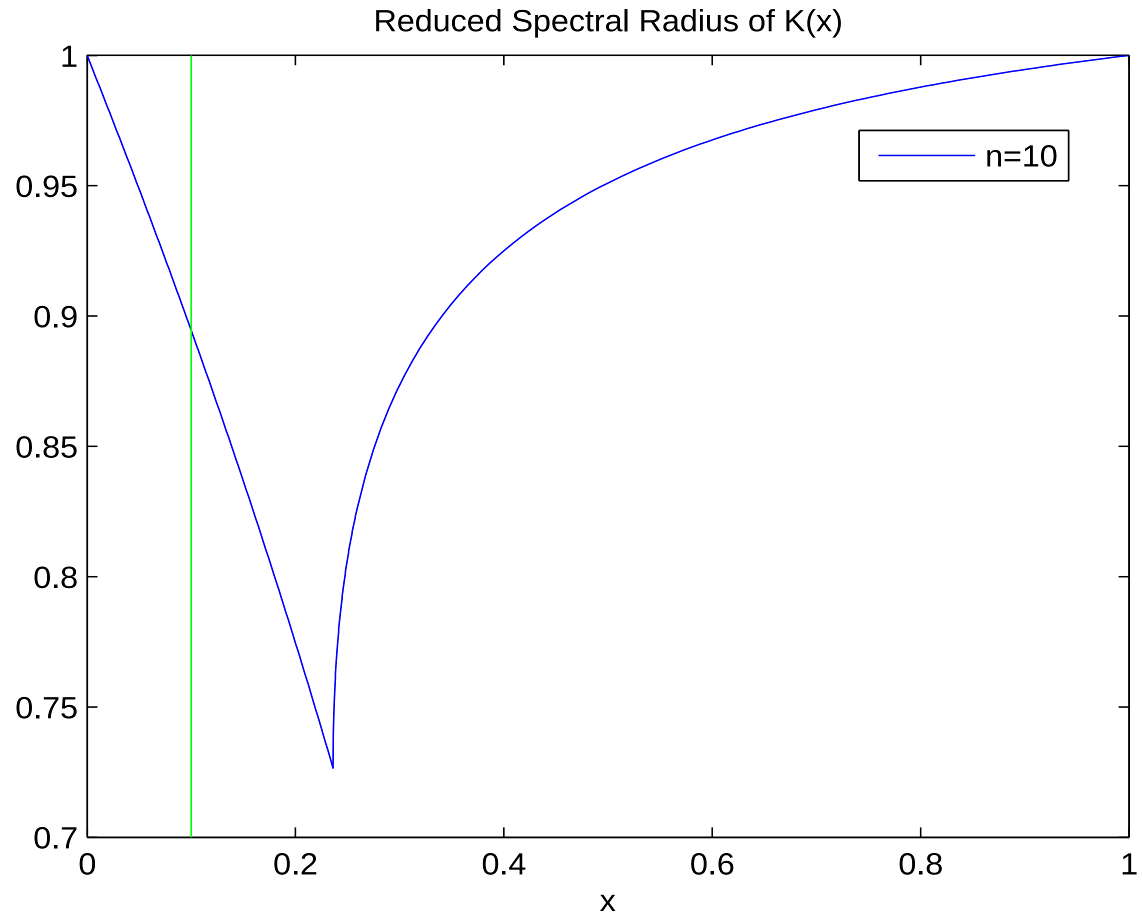
Clearly, $\partial(x) \supseteq \hat{\partial} f(x)$ and $0 \in \partial^\infty f(x)$.

We say f is *subdifferentially regular* at x if $\partial f(x) = \hat{\partial} f(x)$ and $\partial^\infty f(x)$ is the recession cone of $\hat{\partial} f(x)$, hence $\{0\}$ if $\hat{\partial} f(x)$ is compact.

One Variable Example



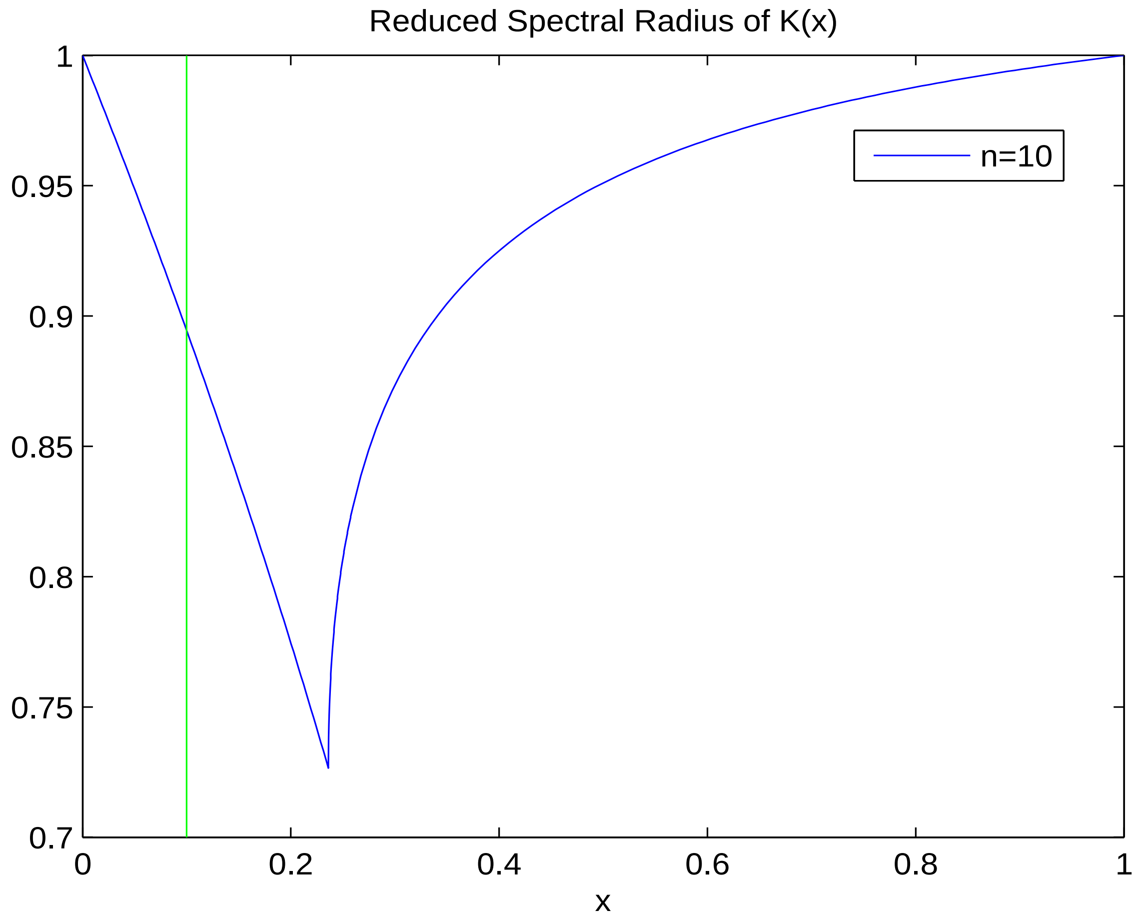
One Variable Example



By direct observation

$$\partial f(x_{\text{opt}}) = \hat{\partial} f(x_{\text{opt}}) = [-1, \infty)$$

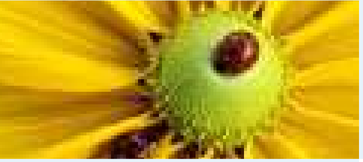
One Variable Example



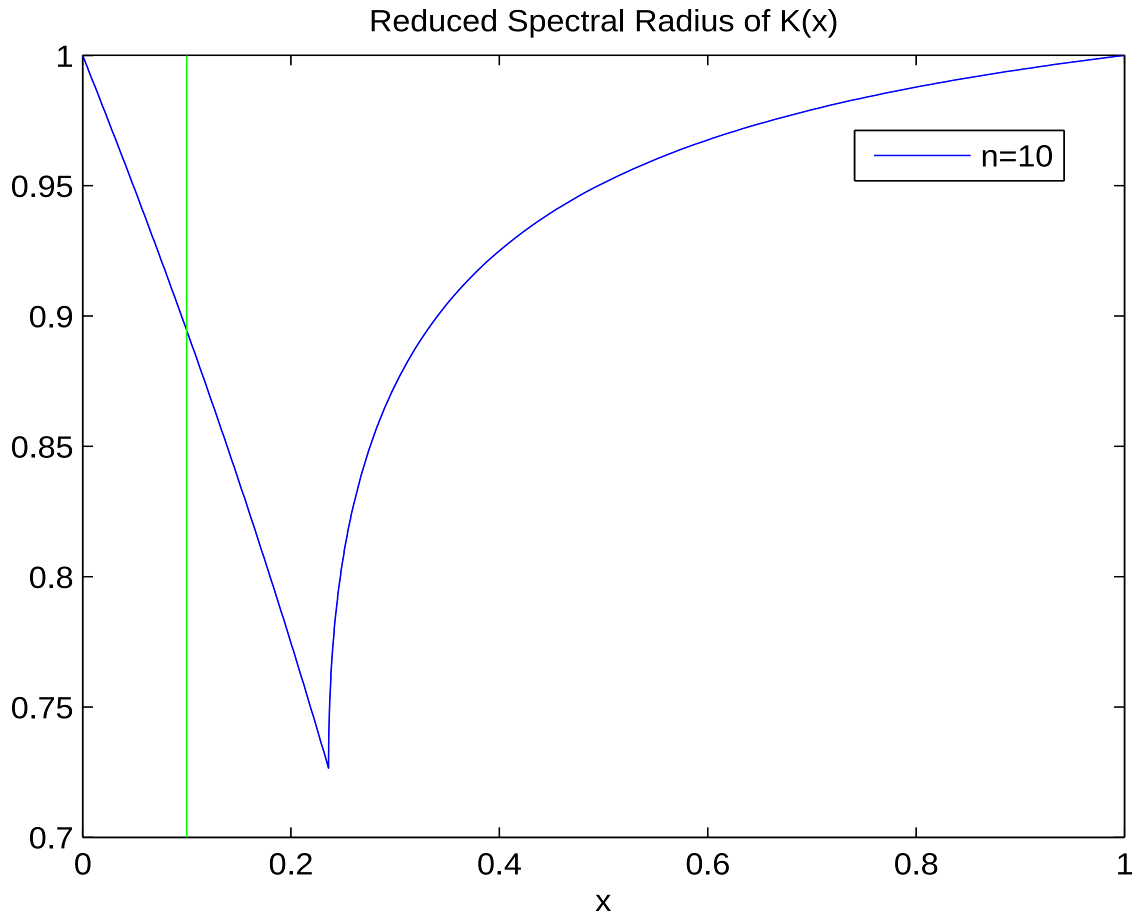
By direct observation

$$\partial f(x_{\rm opt}) = \hat{\partial} f(x_{\rm opt}) = [-1, \infty)$$

$$\partial^\infty f(x_{\rm opt}) = [0, \infty) = \text{ recession cone of } [-1, \infty)$$



One Variable Example



By direct observation

$$\partial f(x_{\text{opt}}) = \hat{\partial} f(x_{\text{opt}}) = [-1, \infty)$$

$$\partial^{\infty} f(x_{\text{opt}}) = [0, \infty) = \text{ recession cone of } [-1, \infty)$$

Hence, this function of one variable is subdifferentially regular.

The Subdifferential of the Spectral Abscissa

Theorem SSA. *The spectral abscissa $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ is subdifferentially regular at $X \in \mathbf{C}^{n\times n}$ if and only if all the active eigenvalues are nonderogatory. When this is the case, let $X = PJP^{-1}$, where $J = \mathrm{Diag}(J_1, \ldots, J_k, J_{k+1}, \ldots)$,*

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & \lambda_i \end{bmatrix} \quad (m_i \times m_i), \quad i = 1 \ldots, k$$

with Re $\lambda_i = \alpha(X)$, $i = 1, \ldots, k$ (the active eigenvalues, with multiplicity m_i).

The Subdifferential of the Spectral Abscissa

Theorem SSA. *The spectral abscissa $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ is subdifferentially regular at $X \in \mathbf{C}^{n\times n}$ if and only if all the active eigenvalues are nonderogatory. When this is the case, let $X = PJP^{-1}$, where $J = \mathrm{Diag}(J_1, \ldots, J_k, J_{k+1}, \ldots)$,*

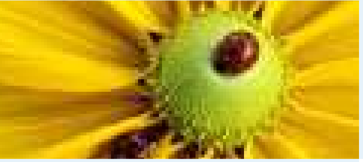
$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix} \quad (m_i \times m_i), \quad i = 1\ldots, k$$

with Re $\lambda_i = \alpha(X)$, $i = 1, \ldots, k$ (the active eigenvalues, with multiplicity m_i). Then:

$$\partial\alpha(X) = \Big\{P^{-*}WP^* : W = \mathrm{Diag}(W_1, \ldots, W_k, 0, \ldots, 0),$$

$$W_i = \begin{bmatrix} \theta_1^i & & & & \\ \theta_2^i & \ddots & & & \\ \vdots & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \theta_{m_i-1}^i & \cdots & \cdots & \theta_2^i & \theta_1^i \end{bmatrix},$$

$$\theta_1^i \in \mathbf{R}^+,\ m_1\theta_1^1 + \cdots m_k\theta_1^k = 1,\ \text{Re } \theta_2^i \geq 0\Big\}.$$



The Subdifferential of the Spectral Abscissa

Theorem SSA. *The spectral abscissa* $\alpha : \mathbf{C}^{n\times n} \to \mathbf{R}$ *is subdifferentially regular at* $X \in \mathbf{C}^{n\times n}$ *if and only if all the active eigenvalues are nonderogatory. When this is the case, let* $X = PJP^{-1}$, *where* $J = \mathrm{Diag}(J_1, \ldots, J_k, J_{k+1}, \ldots)$,

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & \lambda_i \end{bmatrix} \quad (m_i \times m_i), \quad i = 1\ldots, k$$

with Re $\lambda_i = \alpha(X)$, $i = 1, \ldots, k$ *(the active eigenvalues, with multiplicity* m_i*). Then:*

$$\partial\alpha(X) = \big\{ P^{-*} W P^{*} : W = \mathrm{Diag}(W_1, \ldots, W_k, 0, \ldots, 0),$$

$$W_i = \begin{bmatrix} \theta_1^i & & & & \\ \theta_2^i & \cdot & & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \\ \theta_{m_i-1}^i & \cdot & \cdot & \theta_2^i & \theta_1^i \end{bmatrix},$$

$$\theta_1^i \in \mathbf{R}^+,\ m_1\theta_1^1 + \cdots m_k\theta_1^k = 1,\ \text{Re } \theta_2^i \geq 0 \big\}.$$

$$\partial^\infty \alpha(X) :\ \text{same, except } \theta_1^i \equiv 0.$$

When All Active Eigenvalues are Simple

Then $X = PJP^{-1}$, where $J = \mathrm{Diag}(\lambda_1, \ldots, \lambda_k, J_{k+1}, \ldots)$ with Re $\lambda_i = \alpha(X), i = 1, \ldots, k$ (the active eigenvalues),

$$P = [v_1, \ldots, v_k, \ldots], \quad P^{-*} = [u_1, \ldots, u_k, \ldots]$$

with v_i and $u_i, i = 1, \ldots, k$, the corresponding right and left eigenvectors, normalized so $u_i^* v_i = 1$.

When All Active Eigenvalues are Simple

Then $X = PJP^{-1}$, where $J = \mathrm{Diag}(\lambda_1, \ldots, \lambda_k, J_{k+1}, \ldots)$ with Re $\lambda_i = \alpha(X), i = 1, \ldots, k$ (the active eigenvalues),

$$P = [v_1, \ldots, v_k, \ldots], \quad P^{-*} = [u_1, \ldots, u_k, \ldots]$$

with v_i and $u_i, i = 1, \ldots, k$, the corresponding right and left eigenvectors, normalized so $u_i^* v_i = 1$.
Then

$$\partial\alpha(X) = \Big\{P^{-*}WP^* : W = \mathrm{Diag}(\theta_1, \ldots, \theta_k, 0, \ldots, 0),$$

$$\theta_i \in \mathbf{R}^+, \ \ \theta_1 + \cdots + \theta_k = 1\Big\}$$

When All Active Eigenvalues are Simple

Then $X = PJP^{-1}$, where $J = \mathrm{Diag}(\lambda_1, \dots, \lambda_k, J_{k+1}, \dots)$ with Re $\lambda_i = \alpha(X), i = 1, \dots, k$ (the active eigenvalues),

$$P = [v_1, \dots, v_k, \dots], \quad P^{-*} = [u_1, \dots, u_k, \dots]$$

with v_i and $u_i, i = 1, \dots, k$, the corresponding right and left eigenvectors, normalized so $u_i^* v_i = 1$.
Then

$$\partial\alpha(X) = \{P^{-*}WP^* : W = \mathrm{Diag}(\theta_1, \dots, \theta_k, 0, \dots, 0),$$

$$\theta_i \in \mathbf{R}^+, \quad \theta_1 + \cdots + \theta_k = 1\}$$

$$= \left\{ \sum_{i=1}^{k} \theta_i u_i v_i^* : \theta_i \in \mathbf{R}^+, \quad \sum_{i=1}^{k} \theta_i = 1 \right\}.$$

When All Active Eigenvalues are Simple

Then $X = PJP^{-1}$, where $J = \mathrm{Diag}(\lambda_1, \ldots, \lambda_k, J_{k+1}, \ldots)$ with $\mathrm{Re}\,\lambda_i = \alpha(X), i = 1, \ldots, k$ (the active eigenvalues),

$$P = [v_1, \ldots, v_k, \ldots], \quad P^{-*} = [u_1, \ldots, u_k, \ldots]$$

with v_i and $u_i, i = 1, \ldots, k$, the corresponding right and left eigenvectors, normalized so $u_i^* v_i = 1$.
Then

$$\partial\alpha(X) = \{P^{-*}WP^* : W = \mathrm{Diag}(\theta_1, \ldots, \theta_k, 0, \ldots, 0),$$

$$\theta_i \in \mathbf{R}^+, \quad \theta_1 + \cdots + \theta_k = 1\}$$

$$= \left\{ \sum_{i=1}^k \theta_i u_i v_i^* : \theta_i \in \mathbf{R}^+, \quad \sum_{i=1}^k \theta_i = 1 \right\}.$$

If $k = 1$, we have

$$\partial\alpha(X) = \{u_1 v_1^*\} = \{\nabla\alpha(X)\}.$$

Optimality Conditions for a Matrix Family

Let $A : \mathbf{C}^m \to \mathbf{C}^{n\times n}$ be a smooth map satisfying the regularity condition

$$\mathcal{N}\big((DA(\bar{x}))^*\big) \cap \partial^\infty \alpha\big(A(\bar{x})\big) = 0,$$

for $\bar{x} \in \mathbf{C}^m$.

Optimality Conditions for a Matrix Family

Let $A : \mathbf{C}^m \to \mathbf{C}^{n\times n}$ be a smooth map satisfying the regularity condition

$$\mathcal{N}\big((DA(\bar{x}))^*\big) \cap \partial^\infty \alpha\big(A(\bar{x})\big) = 0,$$

for $\bar{x} \in \mathbf{C}^m$.

By a standard chain rule of nonsmooth analysis, a necessary condition for $\bar{x}$ to minimize $\alpha(A(x))$ is, if all active eigenvalues are nonderogatory

$$0 \in \partial(\alpha \circ A)(\bar{x}) = (DA)^* \partial\alpha(A(\bar{x})),$$

Optimality Conditions for a Matrix Family

Let $A : \mathbf{C}^m \to \mathbf{C}^{n\times n}$ be a smooth map satisfying the regularity condition

$$\mathcal{N}\big((DA(\bar{x}))^*\big) \cap \partial^\infty \alpha\big(A(\bar{x})\big) = 0,$$

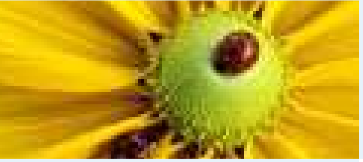
for $\bar{x} \in \mathbf{C}^m$.

By a standard chain rule of nonsmooth analysis, a necessary condition for $\bar{x}$ to minimize $\alpha(A(x))$ is, if all active eigenvalues are nonderogatory

$$0 \in \partial(\alpha \circ A)(\bar{x}) = (DA)^* \partial\alpha(A(\bar{x})),$$

and a sufficient condition is

$$0 \in \text{int } \partial(\alpha \circ A)(\bar{x}) = \text{int } (DA)^* \partial\alpha(A(\bar{x})).$$



Optimality Conditions for a Matrix Family

Let $A : \mathbf{C}^m \to \mathbf{C}^{n\times n}$ be a smooth map satisfying the regularity condition

$$\mathcal{N}\big((DA(\bar{x}))^*\big) \cap \partial^\infty \alpha\big(A(\bar{x})\big) = 0,$$

for $\bar{x} \in \mathbf{C}^m$.

By a standard chain rule of nonsmooth analysis, a necessary condition for $\bar{x}$ to minimize $\alpha(A(x))$ is, if all active eigenvalues are nonderogatory

$$0 \in \partial(\alpha \circ A)(\bar{x}) = (DA)^* \partial\alpha(A(\bar{x})),$$

and a sufficient condition is

$$0 \in \operatorname{int}\, \partial(\alpha \circ A)(\bar{x}) = \operatorname{int}\, (DA)^* \partial\alpha(A(\bar{x})).$$

Furthermore, if the last condition holds, small perturbations to the data make only small perturbations in the minimizer.

A Simple Example

Define the affine map $A : \mathbf{C}^{n-1} \to \mathbf{C}^{n\times n}$ by

$$A(x) = \begin{pmatrix} -x_1 & 1 & 0 & \dots & 0 \\ x_1 & 0 & 1 & \dots & 0 \\ x_2 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ x_{n-1} & 0 & 0 & \dots & 0 \end{pmatrix}.$$

A Simple Example

Define the affine map $A : \mathbf{C}^{n-1} \to \mathbf{C}^{n\times n}$ by

$$A(x) = \begin{pmatrix} -x_1 & 1 & 0 & \dots & 0 \\ x_1 & 0 & 1 & \dots & 0 \\ x_2 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ x_{n-1} & 0 & 0 & \dots & 0 \end{pmatrix}.$$

The adjoint of the derivative is $(DA)^* : \mathbf{C}^{n\times n} \to \mathbf{C}^{n-1}$, given by

$$(y_{rs}) \in \mathbf{C}^{n\times n} \mapsto (y_{21} - y_{11}, y_{31}, y_{41}, \dots, y_{n1})^T.$$

A Simple Example

Define the affine map $A : \mathbf{C}^{n-1} \to \mathbf{C}^{n\times n}$ by

$$A(x) = \begin{pmatrix} -x_1 & 1 & 0 & \dots & 0 \\ x_1 & 0 & 1 & \dots & 0 \\ x_2 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ x_{n-1} & 0 & 0 & \dots & 0 \end{pmatrix}.$$

The adjoint of the derivative is $(DA)^* : \mathbf{C}^{n\times n} \to \mathbf{C}^{n-1}$, given by

$$(y_{rs}) \in \mathbf{C}^{n\times n} \mapsto (y_{21} - y_{11}, y_{31}, y_{41}, \dots, y_{n1})^T.$$

So, by Theorem SSA and the chain rule,

$$\hat{\partial}(\alpha \circ A)(0) = \partial(\alpha \circ A)(0) = (DA)^*\partial\alpha(A(0))$$

$$= \left\{(\theta_2 - 1/n, \theta_3, \theta_4, \dots, \theta_n)^T : \text{ Re } \theta_2 \geq 0\right\},$$

and since the interior of this set contains 0, we see $\bar{x} = 0$ is a sharp local minimizer of the spectral abscissa of $A(x)$.

Active Eigenvalue Structures at Optimizers

This example: one active eigenvalue, nonderogatory, with multiplicity n.

Active Eigenvalue Structures at Optimizers

This example: one active eigenvalue, nonderogatory, with multiplicity n.

Markov Chain example: $n - 2$ active eigenvalues, all nonderogatory, with multiplicities $2, 1, 1, \ldots, 1$: same kind of analysis applies.

Active Eigenvalue Structures at Optimizers

This example: one active eigenvalue, nonderogatory, with multiplicity n.

Markov Chain example: $n-2$ active eigenvalues, all nonderogatory, with multiplicities $2, 1, 1, \ldots, 1$: same kind of analysis applies.

An example from surface approximation: several fixed eigenvalues, want to reduce modulus of others: after much numerical computation, found that all can be reduced to zero

Active Eigenvalue Structures at Optimizers

This example: one active eigenvalue, nonderogatory, with multiplicity n.

Markov Chain example: $n-2$ active eigenvalues, all nonderogatory, with multiplicities $2, 1, 1, \ldots, 1$: same kind of analysis applies.

An example from surface approximation: several fixed eigenvalues, want to reduce modulus of others: after much numerical computation, found that all can be reduced to zero

- triangular mesh case: verified analytically, with the derogatory zero eigenvalue having Jordan block sizes 2, 1, 1, 1.

Active Eigenvalue Structures at Optimizers

This example: one active eigenvalue, nonderogatory, with multiplicity n.

Markov Chain example: $n-2$ active eigenvalues, all nonderogatory, with multiplicities $2, 1, 1, \ldots, 1$: same kind of analysis applies.

An example from surface approximation: several fixed eigenvalues, want to reduce modulus of others: after much numerical computation, found that all can be reduced to zero

- triangular mesh case: verified analytically, with the derogatory zero eigenvalue having Jordan block sizes 2, 1, 1, 1.
- quadrilateral mesh case: numerically reduced moduli of eigenvalues to about 10^{-4} and determined that the derogatory zero eigenvalue has Jordan block sizes 5, 3, 2, 2.

Active Eigenvalue Structures at Optimizers

This example: one active eigenvalue, nonderogatory, with multiplicity n.

Markov Chain example: $n-2$ active eigenvalues, all nonderogatory, with multiplicities $2, 1, 1, \ldots, 1$: same kind of analysis applies.

An example from surface approximation: several fixed eigenvalues, want to reduce modulus of others: after much numerical computation, found that all can be reduced to zero

- triangular mesh case: verified analytically, with the derogatory zero eigenvalue having Jordan block sizes 2, 1, 1, 1.
- quadrilateral mesh case: numerically reduced moduli of eigenvalues to about 10^{-4} and determined that the derogatory zero eigenvalue has Jordan block sizes 5, 3, 2, 2.

Any structure is possible, but expect nonderogatory eigenvalues generically.

Steepest descent (gradient) method with line search: fails.

Steepest descent (gradient) method with line search: fails.

Gradient sampling with line search: nice convergence theory but computationally intensive.

Numerical Optimization Methods

Steepest descent (gradient) method with line search: fails.

Gradient sampling with line search: nice convergence theory but computationally intensive.

BFGS quasi-Newton method with line search: very effective in practice.

Numerical Optimization Methods

Steepest descent (gradient) method with line search: fails.

Gradient sampling with line search: nice convergence theory but computationally intensive.

BFGS quasi-Newton method with line search: very effective in practice.

In both cases, make no attempt to *predict active eigenvalues and their multiplicities*: simply use gradients which exist a.e.:

$$\nabla(\alpha \circ A)(x) = \big(DA(x)\big)^* \nabla\alpha\big(A(x)\big) = \big(DA(x)\big)^* v_1 u_1^*$$

so

$$\big(\nabla(\alpha \circ A)(x)\big)_k = \langle \frac{\partial A}{\partial x_k}(x), v_1 u_1^* \rangle = u_1^* \frac{\partial A}{\partial x_k}(x) v_1$$

where u_1 and v_1 are left and right eigenvectors for the dominant eigenvalue λ_1.

References for Part II

Optimizing Matrix Stability
J.V. Burke, A.S. Lewis and M.L. Overton
Proceedings of the American Mathematical Society (2001)

References for Part II

Optimizing Matrix Stability
J.V. Burke, A.S. Lewis and M.L. Overton
Proceedings of the American Mathematical Society (2001)

Variational Analysis of Non-Lipschitz Spectral Functions
J.V. Burke and M.L. Overton
Math. Programming (2001)

References for Part II

Optimizing Matrix Stability
J.V. Burke, A.S. Lewis and M.L. Overton
Proceedings of the American Mathematical Society (2001)

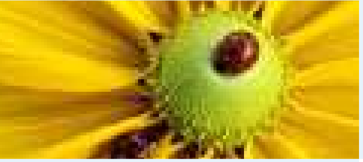
Variational Analysis of Non-Lipschitz Spectral Functions
J.V. Burke and M.L. Overton
Math. Programming (2001)

Variational Analysis
R.T. Rockafellar and R.J.B. Wets
Springer (1998)

The subdifferential we use is due to Mordukhovich. It coincides with the Clarke generalized gradient in the subdifferentially regular case.

References for Part II, Continued

A Robust Gradient Sampling Method for Nonsmooth, Nonconvex Optimization
J.V. Burke, A.S. Lewis and M.L. Overton
SIAM J. Optimization (2006)



References for Part II, Continued

A Robust Gradient Sampling Method for Nonsmooth, Nonconvex Optimization
J.V. Burke, A.S. Lewis and M.L. Overton
SIAM J. Optimization (2006)

Nonsmooth Optimization via Quasi-Newton Methods
A.S. Lewis and M.L. Overton
In revision for *Math. Programming*.

References for Part II, Continued

A Robust Gradient Sampling Method for Nonsmooth, Nonconvex Optimization
J.V. Burke, A.S. Lewis and M.L. Overton
SIAM J. Optimization (2006)

Nonsmooth Optimization via Quasi-Newton Methods
A.S. Lewis and M.L. Overton
In revision for *Math. Programming*.

Optimizing the Asymptotic Convergence Rate of the Diaconis-Holmes-Neal Sampler
K. K. Gade and M.L. Overton
Advances in Applied Mathematics (2007)

References for Part II, Continued

A Robust Gradient Sampling Method for Nonsmooth, Nonconvex Optimization
J.V. Burke, A.S. Lewis and M.L. Overton
SIAM J. Optimization (2006)

Nonsmooth Optimization via Quasi-Newton Methods
A.S. Lewis and M.L. Overton
In revision for *Math. Programming*.

Optimizing the Asymptotic Convergence Rate of the Diaconis-Holmes-Neal Sampler
K. K. Gade and M.L. Overton
Advances in Applied Mathematics (2007)

Eigenvalue Optimization in C^2 Subdivision and Boundary Subdivision
Sara Grundel, Ph.D. thesis, NYU, 2011.

Part III
Optimization of Pseudospectra
with
M. Gürbüzbalaban (NYU)
A.S. Lewis (Cornell)

Pseudospectra

The area swept out in the complex plane by the eigenvalues under perturbation.

$$\Lambda_\epsilon(A) = \{z \in \mathbf{C} \ : \ \det(A+E-zI) = 0 \text{ for some } E \text{ with } \|E\| \le \epsilon\}$$

Pseudospectra

The area swept out in the complex plane by the eigenvalues under perturbation.

$$\Lambda_\epsilon(A) = \{z \in \mathbf{C} \;:\; \det(A+E-zI) = 0 \text{ for some } E \text{ with } \|E\| \le \epsilon\}$$

Motivation: a more robust measure of system behaviour than eigenvalues.

Pseudospectra

The area swept out in the complex plane by the eigenvalues under perturbation.

$$\Lambda_\epsilon(A) = \{z \in \mathbf{C} \ : \ \det(A+E-zI) = 0 \text{ for some } E \text{ with } \|E\| \leq \epsilon\}$$

Motivation: a more robust measure of system behaviour than eigenvalues.

For the spectral and Frobenius norms

$$\Lambda_\epsilon(A) = \{z \in \mathbf{C} \ : \ \sigma_n(A - zI) \leq \epsilon\}$$

where σ_n denotes smallest singular value. (Proof: SVD)

Pseudospectra

The area swept out in the complex plane by the eigenvalues under perturbation.

$$\Lambda_\epsilon(A) = \{z \in \mathbf{C} \ : \ \det(A+E-zI) = 0 \text{ for some } E \text{ with } \|E\| \leq \epsilon\}$$

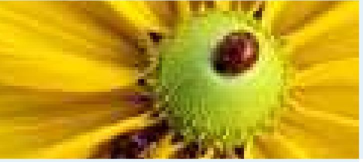
Motivation: a more robust measure of system behaviour than eigenvalues.

For the spectral and Frobenius norms

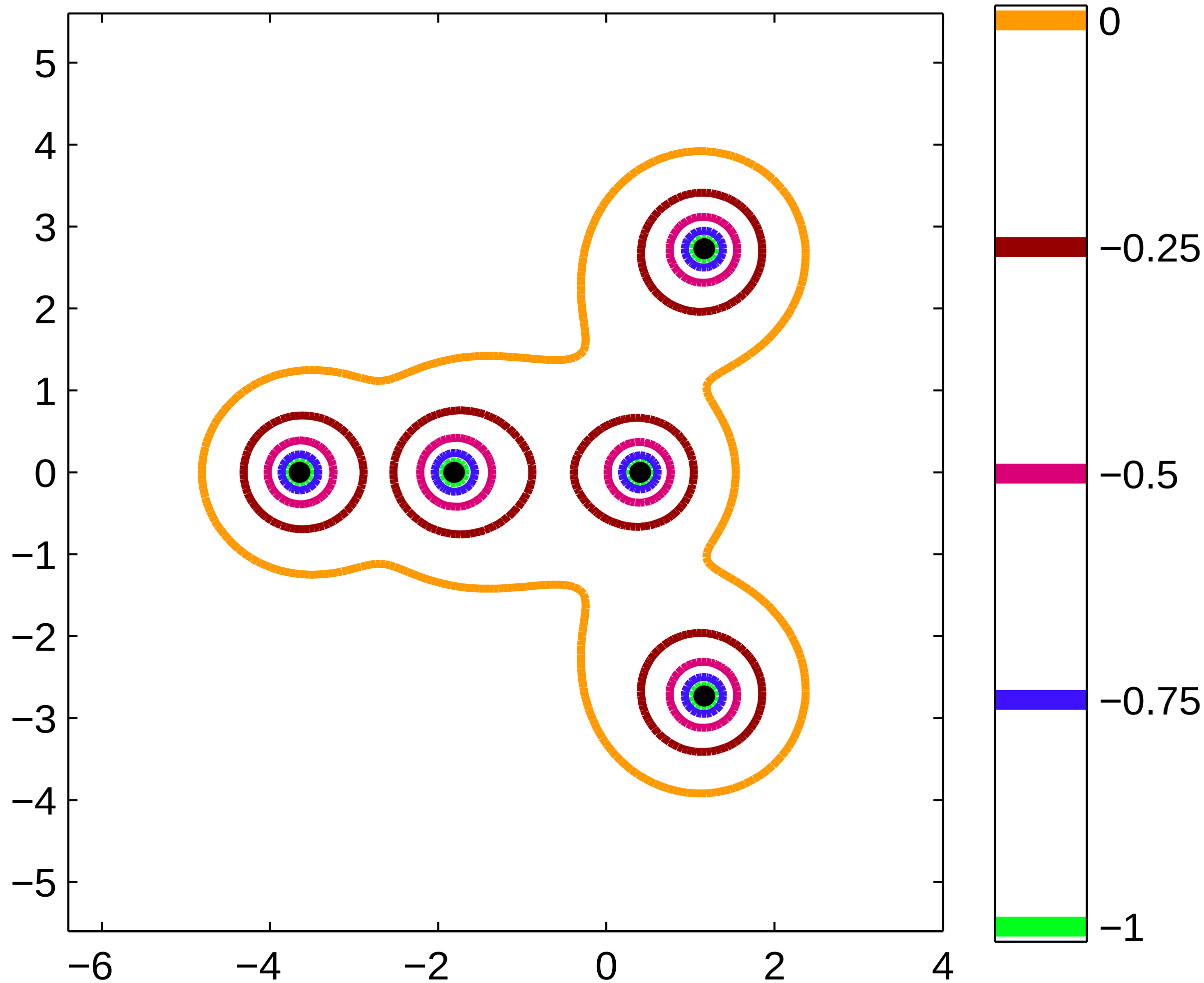
$$\Lambda_\epsilon(A) = \{z \in \mathbf{C} \ : \ \sigma_n(A - zI) \leq \epsilon\}$$

where σ_n denotes smallest singular value. (Proof: SVD)

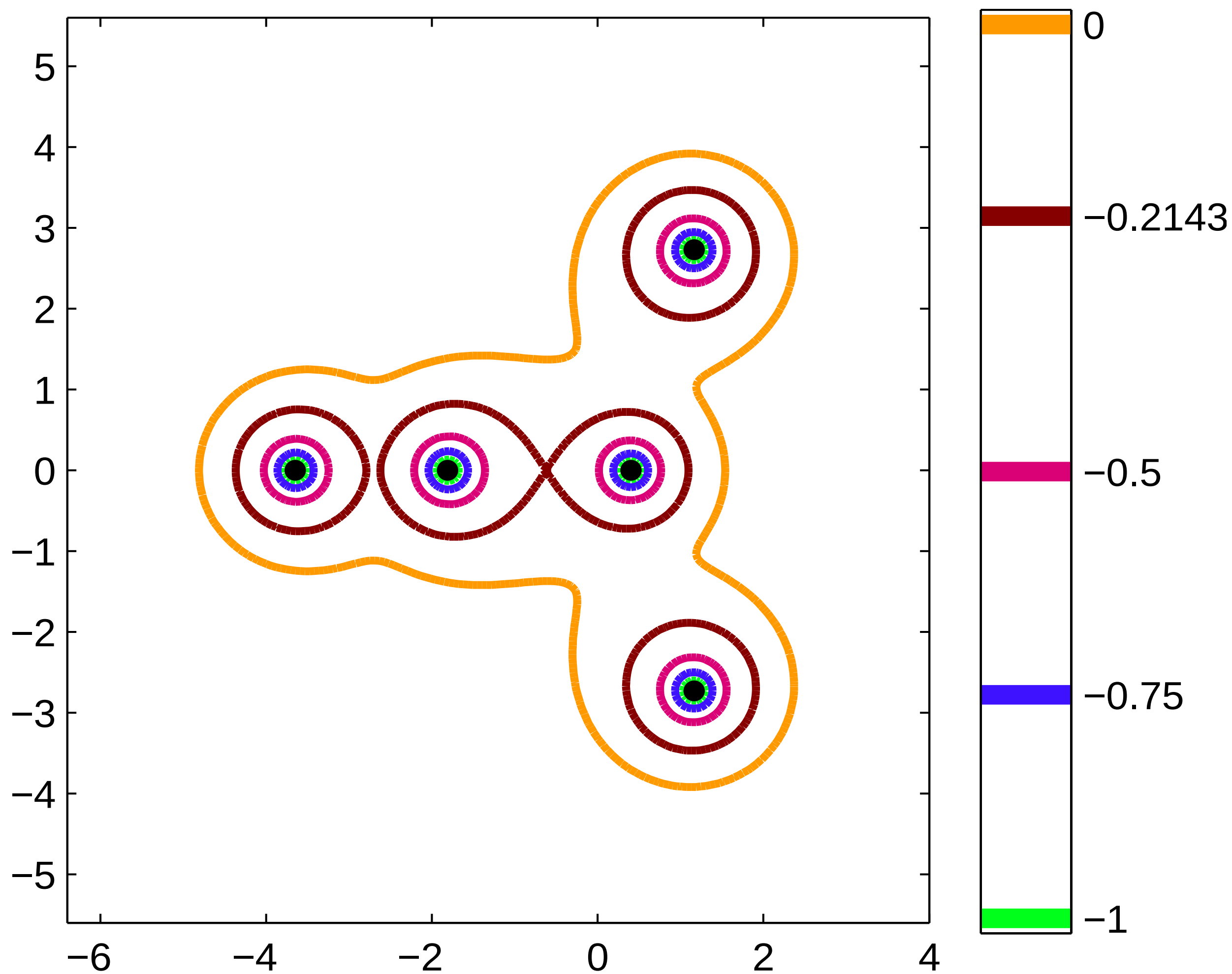
Let $f(x, y) = \sigma_n(A - (x + iy)I)$. Then pseudospectra are lower level sets of f.



A Randomly Generated Example, $n = 5$

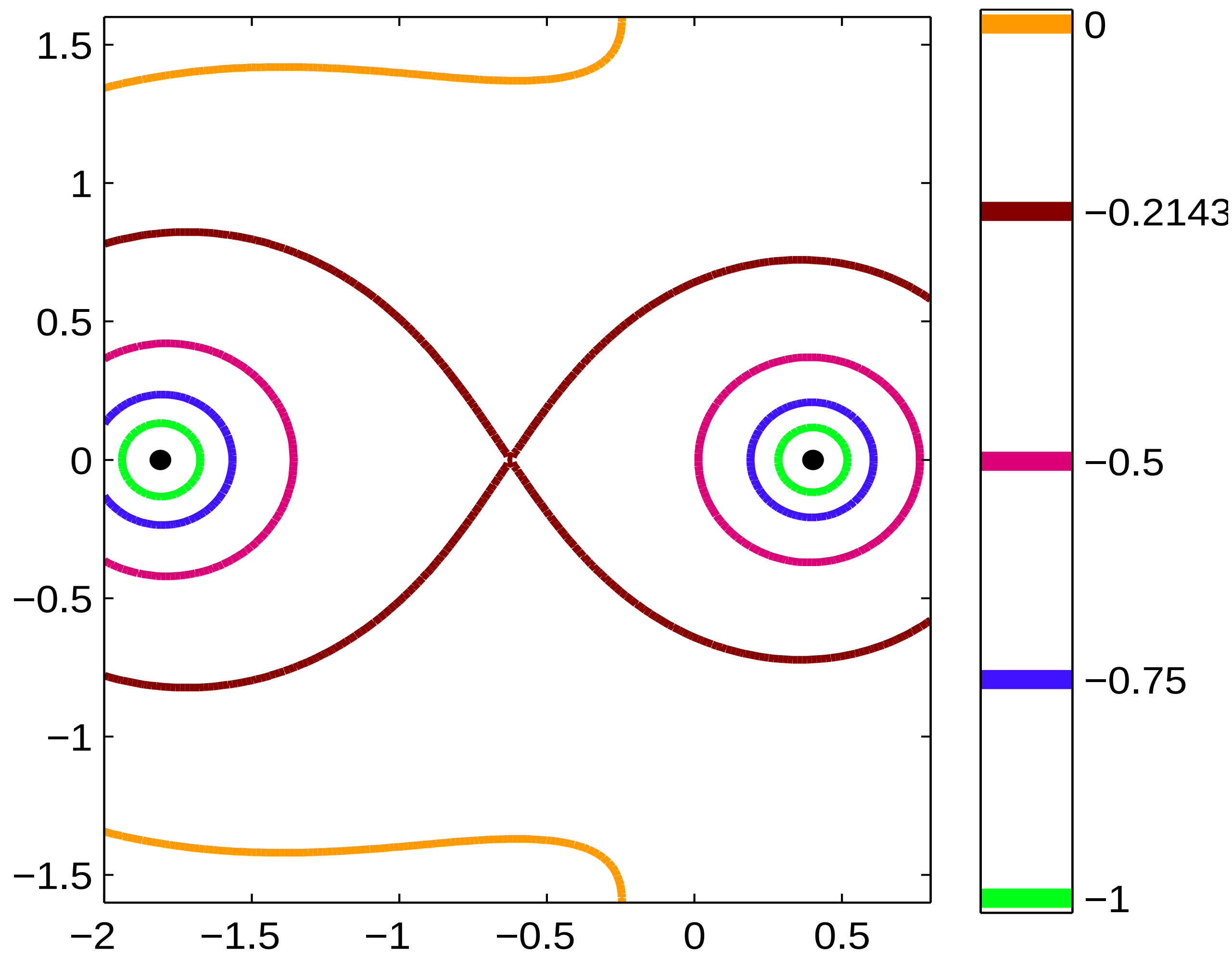


The black dots are the eigenvalues and the colored curves are the boundaries of the ϵ-pseudospectra. The scale on the right is log 10.



The black dots are the eigenvalues and the colored curves are the boundaries of the ϵ-pseudospectra. The scale on the right is log 10.

Same, Close-up



The Pseudospectrum is not Lipschitz

The set-valued map $\Lambda_\epsilon(A)$ is not Lipschitz continuous w.r.t. either A or ϵ near critical values of ϵ.

The Pseudospectrum is not Lipschitz

The set-valued map $\Lambda_\epsilon(A)$ is not Lipschitz continuous w.r.t. either A or ϵ near critical values of ϵ.

What about pseudospectral max functions?

Pseudospectral abscissa:

$$\alpha_\epsilon(A) = \max\{\text{Re } z \ : \ z \in \Lambda_\epsilon(A)\}$$

The Pseudospectrum is not Lipschitz

The set-valued map $\Lambda_\epsilon(A)$ is not Lipschitz continuous w.r.t. either A or ϵ near critical values of ϵ.

What about pseudospectral max functions?

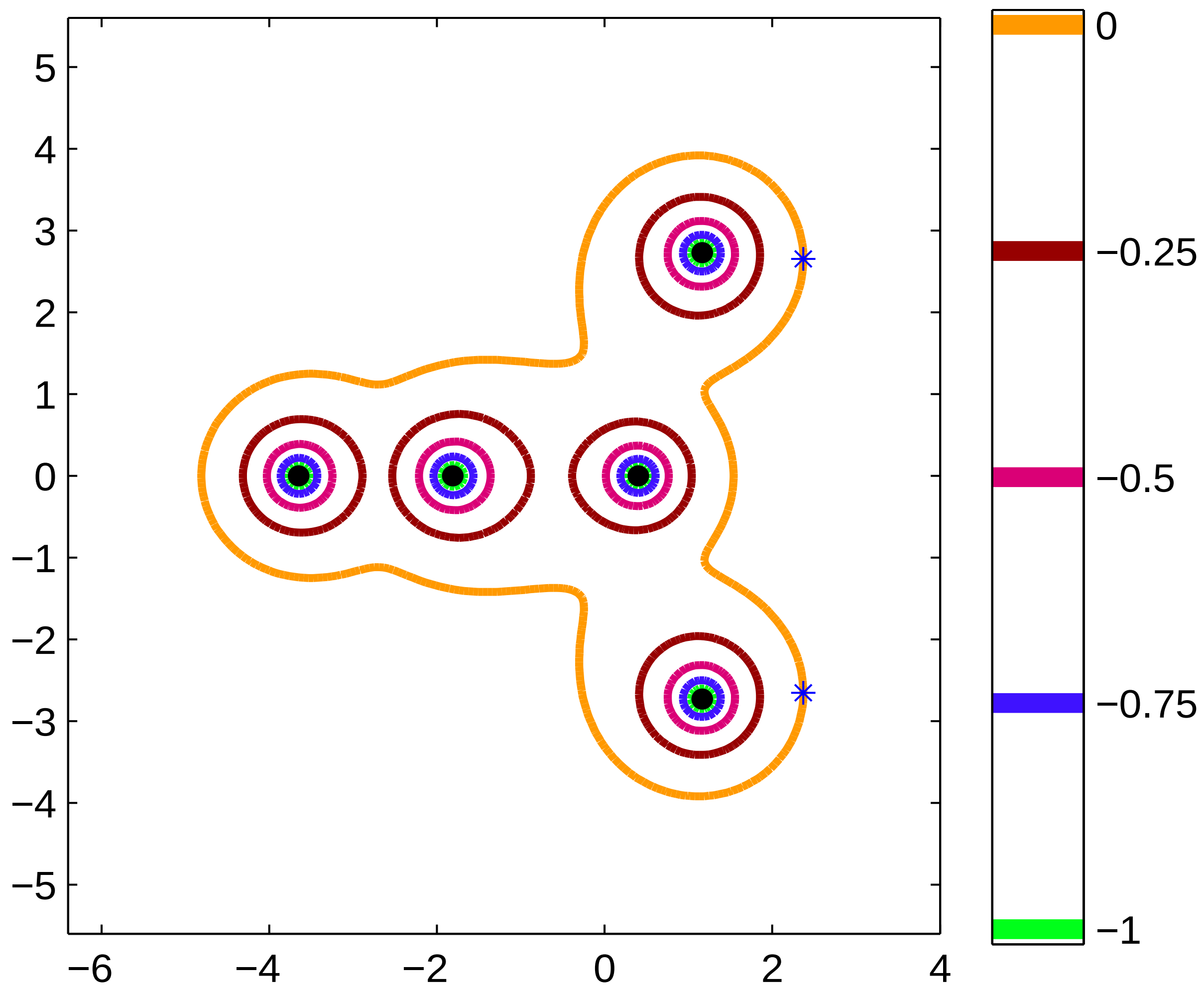
Pseudospectral abscissa:

$$\alpha_\epsilon(A) = \max\{\text{Re } z \ : \ z \in \Lambda_\epsilon(A)\}$$

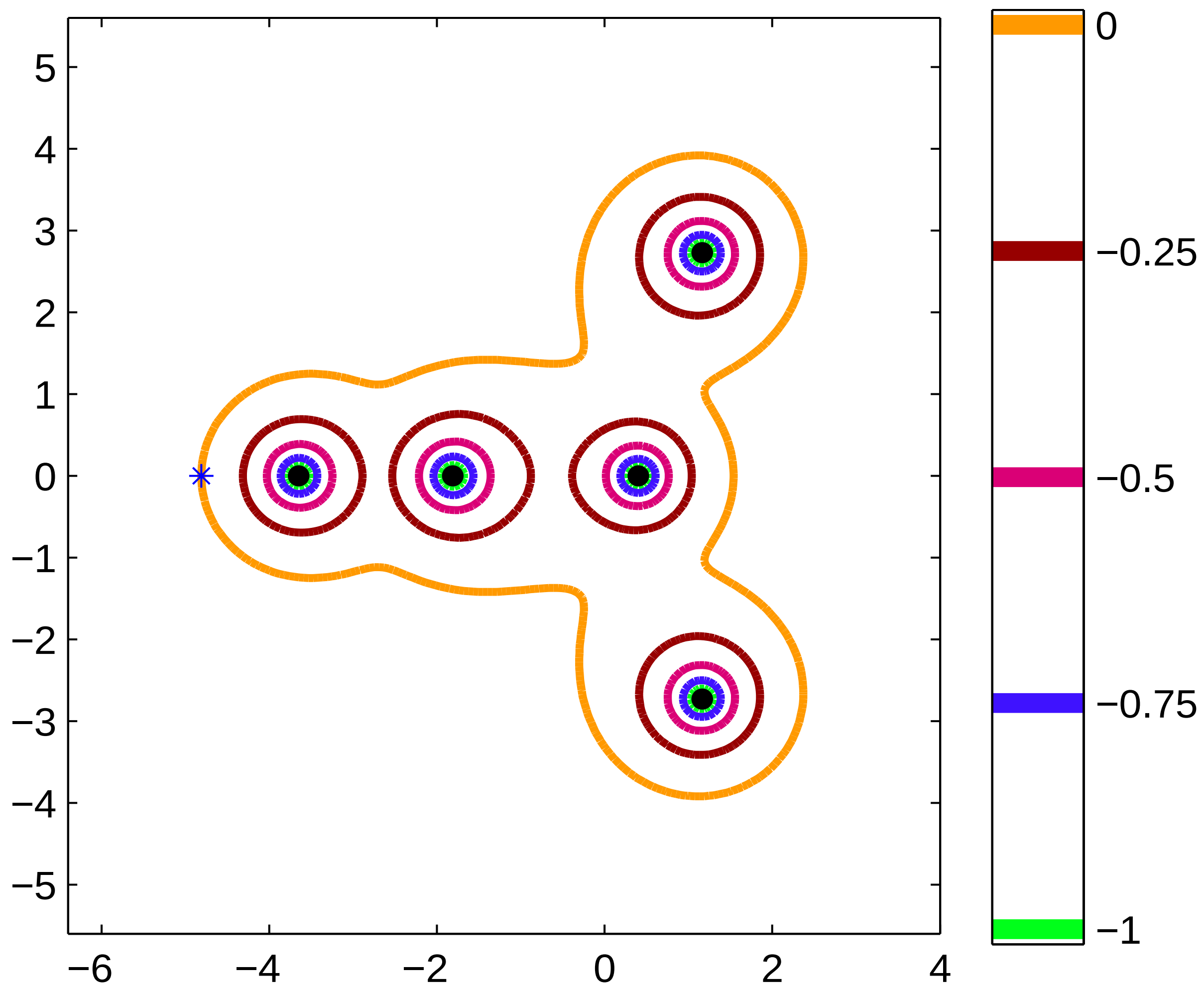
Pseudospectral radius:

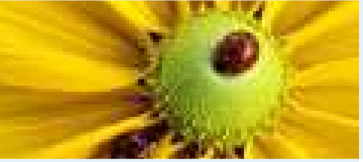
$$\alpha_\epsilon(A) = \max\{|z| \ : \ z \in \Lambda_\epsilon(A)\}$$

Same Example: Pseudospectral Abscissa, $\epsilon = 10^0$



Same Example: Pseudospectral Radius, $\epsilon = 10^0$





Pseudospectral Max Functions are Locally Lipschitz

Theorem PMFLL. *The pseudospectral abscissa and radius functions α_ϵ and ρ_ϵ are locally Lipschitz at all matrices A and at all $\epsilon > 0$.*

Pseudospectral Max Functions are Locally Lipschitz

Theorem PMFLL. *The pseudospectral abscissa and radius functions α_ϵ and ρ_ϵ are locally Lipschitz at all matrices A and at all $\epsilon > 0$.*

Proof: surprisingly complex, as one must allow for the possibility that the pseudospectral max function is attained at a point $z \in \Lambda_\epsilon(A)$ where $\sigma_{n-1}(A - zI) = \sigma_n(A - zI)$.

Pseudospectral Max Functions are Locally Lipschitz

Theorem PMFLL. *The pseudospectral abscissa and radius functions α_ϵ and ρ_ϵ are locally Lipschitz at all matrices A and at all $\epsilon > 0$.*

Proof: surprisingly complex, as one must allow for the possibility that the pseudospectral max function is attained at a point $z \in \Lambda_\epsilon(A)$ where $\sigma_{n-1}(A - zI) = \sigma_n(A - zI)$.

Computing α_ϵ and ρ_ϵ: requires solving nonconvex global optimization problem in two real variables. Specialized algorithms are based on computing eigenvalues of Hamiltonian matrices and checking whether any are imaginary.

Pseudospectral Max Functions are Locally Lipschitz

Theorem PMFLL. *The pseudospectral abscissa and radius functions α_ϵ and ρ_ϵ are locally Lipschitz at all matrices A and at all $\epsilon > 0$.*

Proof: surprisingly complex, as one must allow for the possibility that the pseudospectral max function is attained at a point $z \in \Lambda_\epsilon(A)$ where $\sigma_{n-1}(A - zI) = \sigma_n(A - zI)$.

Computing α_ϵ and ρ_ϵ: requires solving nonconvex global optimization problem in two real variables. Specialized algorithms are based on computing eigenvalues of Hamiltonian matrices and checking whether any are imaginary.

Well known related quantity: the Distance to Instability. For a matrix whose eigenvalues are in the left-half plane: $d_{\mathrm{ci}}(A)$ is the largest ϵ such that $\Lambda_\epsilon(A)$ is in the left-half plane.

Pseudospectral Max Functions are Locally Lipschitz

Theorem PMFLL. *The pseudospectral abscissa and radius functions α_ϵ and ρ_ϵ are locally Lipschitz at all matrices A and at all $\epsilon > 0$.*

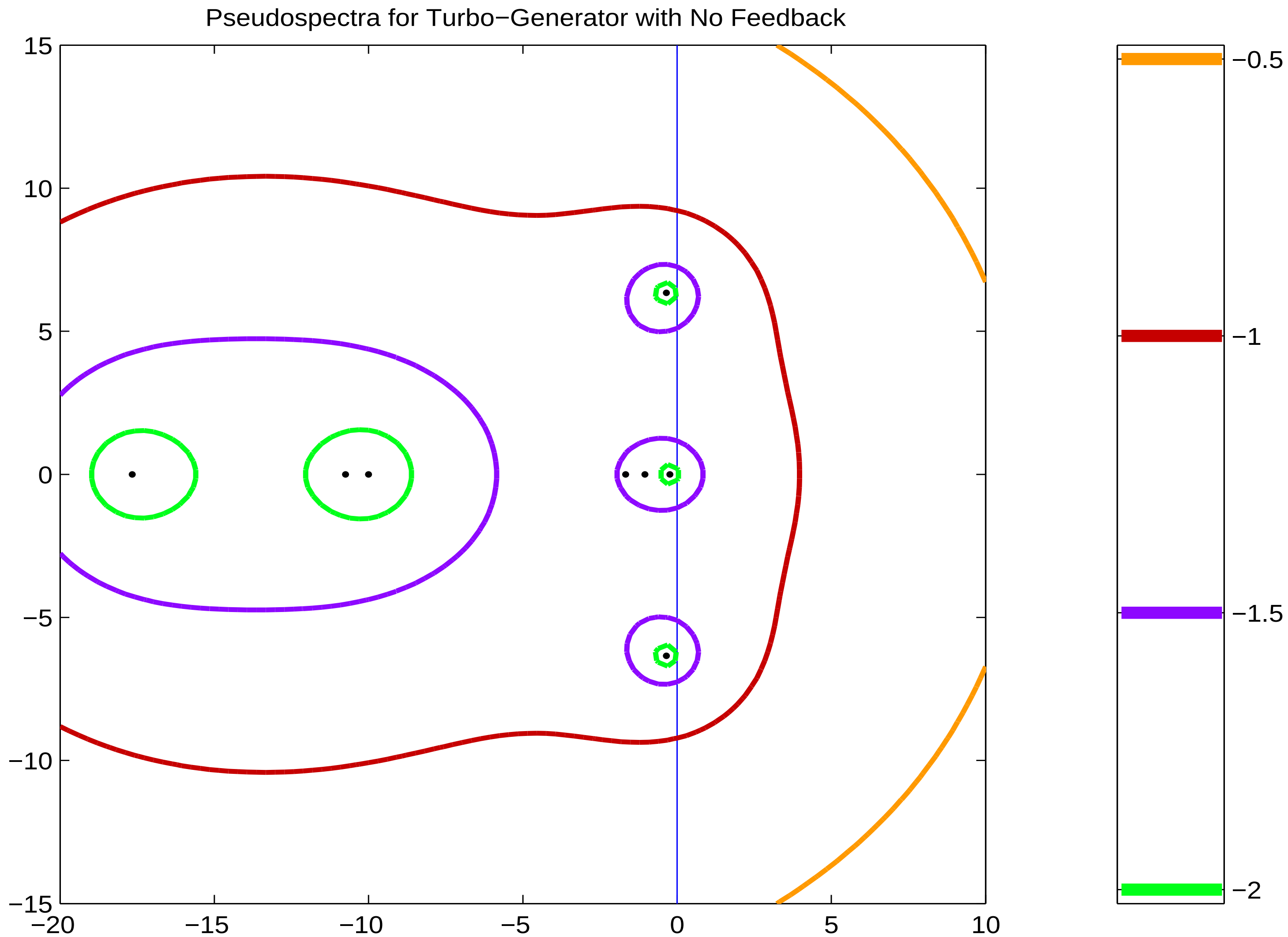
Proof: surprisingly complex, as one must allow for the possibility that the pseudospectral max function is attained at a point $z \in \Lambda_\epsilon(A)$ where $\sigma_{n-1}(A - zI) = \sigma_n(A - zI)$.

Computing α_ϵ and ρ_ϵ: requires solving nonconvex global optimization problem in two real variables. Specialized algorithms are based on computing eigenvalues of Hamiltonian matrices and checking whether any are imaginary.

Well known related quantity: the Distance to Instability. For a matrix whose eigenvalues are in the left-half plane: $d_{\mathrm{ci}}(A)$ is the largest ϵ such that $\Lambda_\epsilon(A)$ is in the left-half plane.

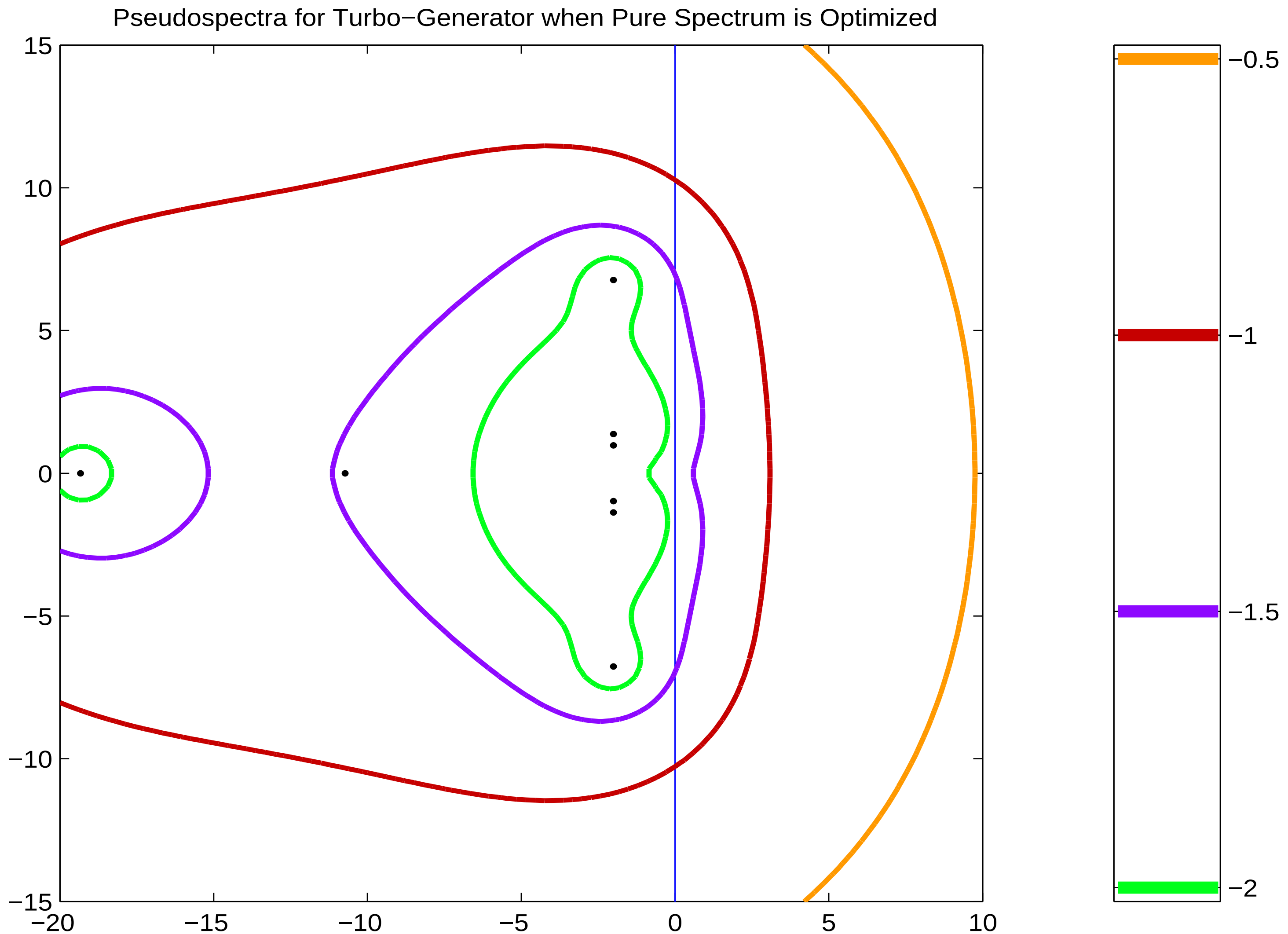
Numerical optimization using gradients: now left and right singular vectors appear instead of left and right eigenvectors.

A Turbo Generator Control Problem



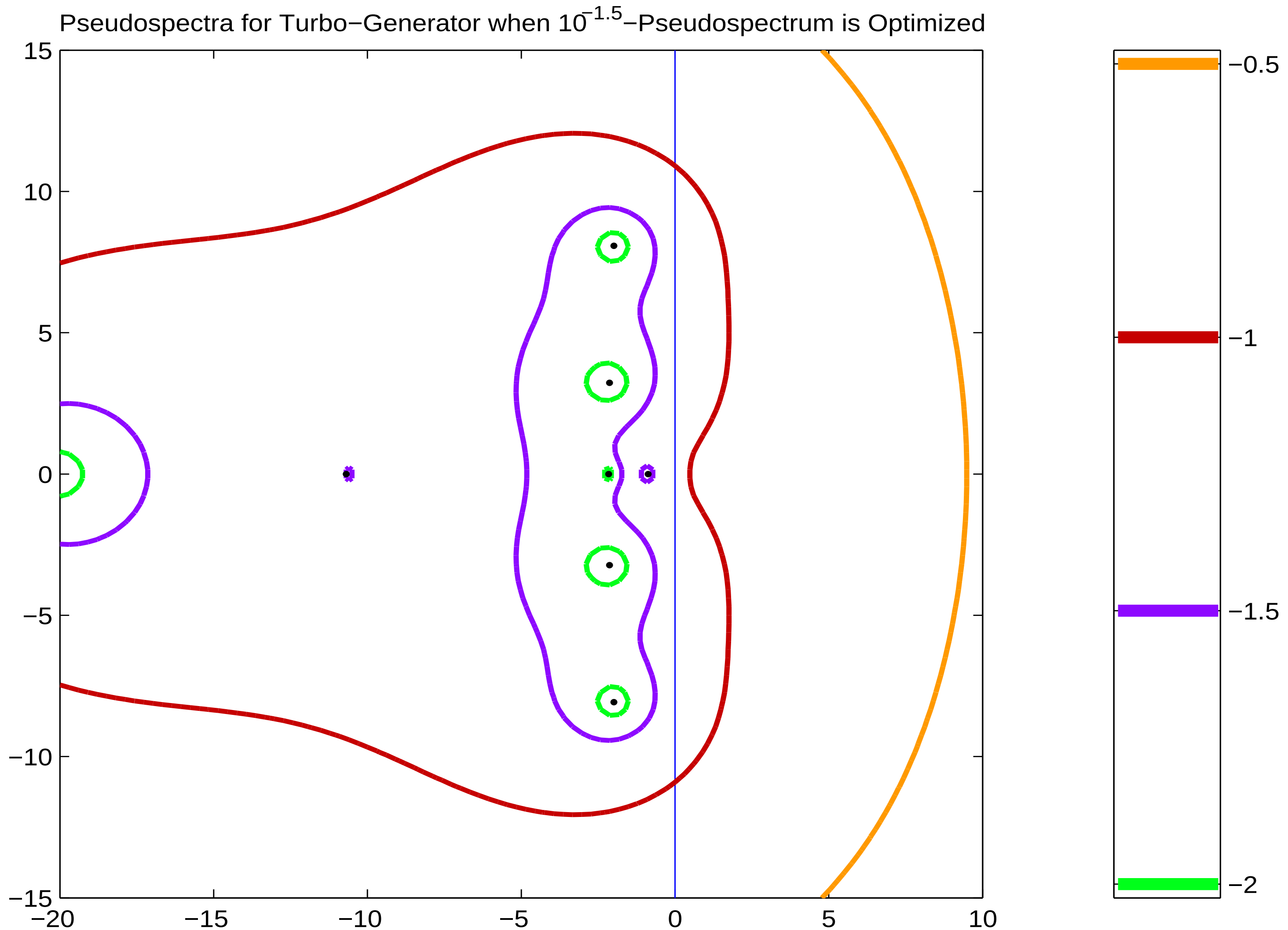
Pseudospectra for open-loop turbo generator plant with no feedback.

Turbo Generator with Optimized Eigenvalues



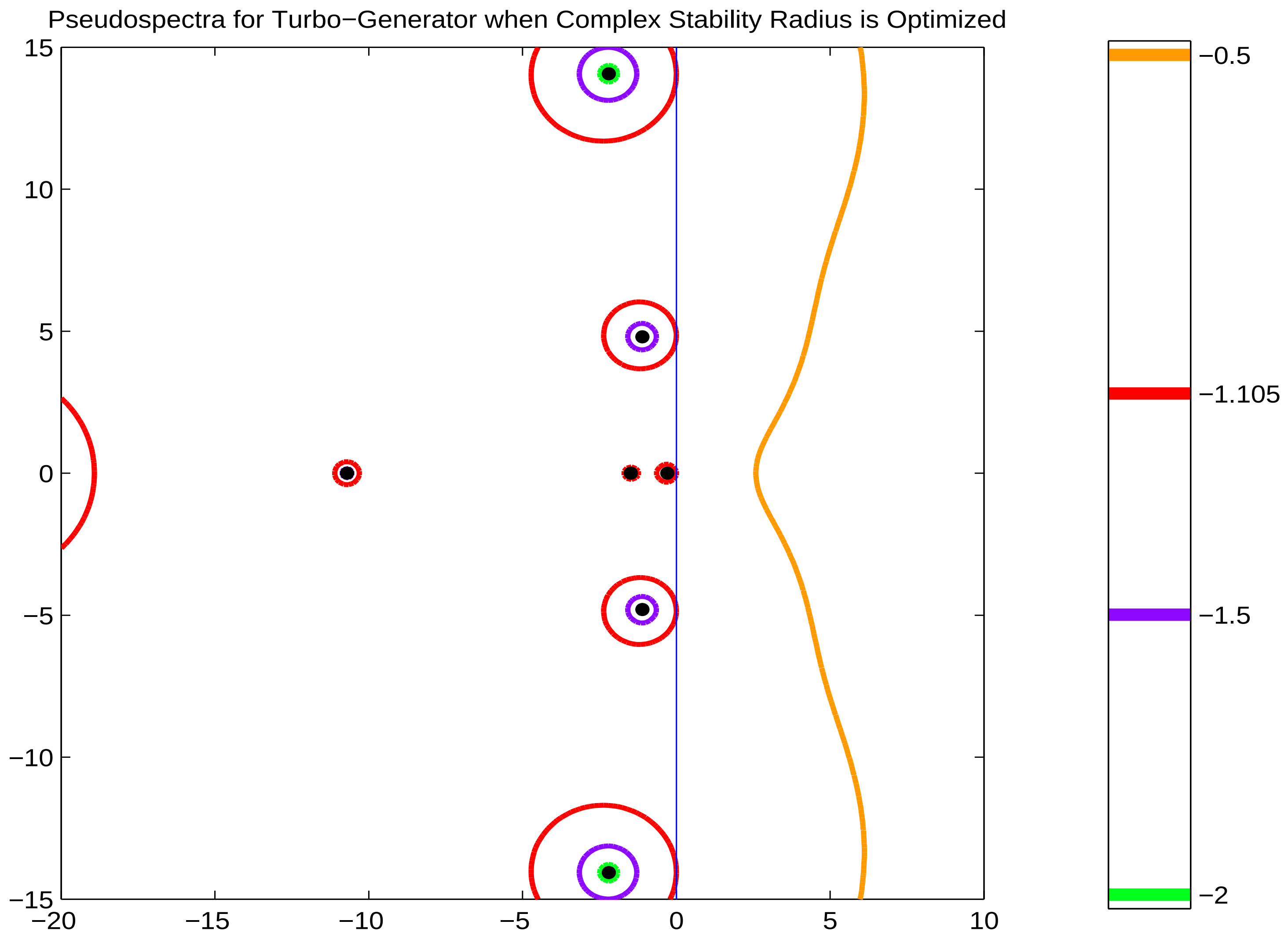
Pseudospectra for turbo generator plant with feedback computed by minimizing the spectral abscissa

Turbo Generator with Optimized ϵ-Pseudospectrum



Pseudospectra for turbo generator plant with feedback computed by minimizing α_ϵ with $\epsilon = 10^{-1.5}$

Turbo Generator with Optimized Dist. to Instability



Pseudospectra for turbo generator plant with feedback computed by minimizing $d_{\mathrm{ci}}(A)$

References for Part III

Spectra and Pseudospectra
L. N. Trefethen and M. Embree
Princeton University Press (2005).
Origins: Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, etc.

References for Part III

Spectra and Pseudospectra
L. N. Trefethen and M. Embree
Princeton University Press (2005).

Origins: Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, etc.

Pseudospectral plots: EigTool (T. Wright and L.N. Trefethen, 2004).

References for Part III

Spectra and Pseudospectra
L. N. Trefethen and M. Embree
Princeton University Press (2005).
Origins: Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, etc.

Pseudospectral plots: EigTool (T. Wright and L.N. Trefethen, 2004).

Some Regularity Results for the Pseudospectral Abscissa and Pseudospectral Radius of a Matrix
M. Gürbüzbalaban and M.L. Overton
Submitted to SIAM J. Optimization.

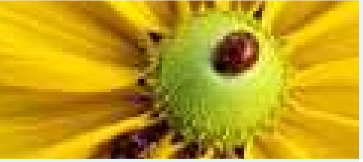
References for Part III

Spectra and Pseudospectra
L. N. Trefethen and M. Embree
Princeton University Press (2005).
Origins: Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, etc.

Pseudospectral plots: EigTool (T. Wright and L.N. Trefethen, 2004).

Some Regularity Results for the Pseudospectral Abscissa and Pseudospectral Radius of a Matrix
M. Gürbüzbalaban and M.L. Overton
Submitted to SIAM J. Optimization.

Our paper builds on and proves a conjecture in:
Variational Analysis of Pseudospectra
A.S. Lewis and C.-H. J. Pang
SIAM J. Optimization (2008).



References for Part III

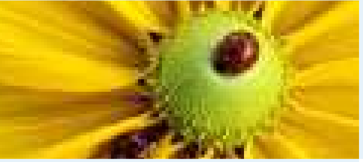
Spectra and Pseudospectra
L. N. Trefethen and M. Embree
Princeton University Press (2005).
Origins: Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, etc.

Pseudospectral plots: EigTool (T. Wright and L.N. Trefethen, 2004).

Some Regularity Results for the Pseudospectral Abscissa and Pseudospectral Radius of a Matrix
M. Gürbüzbalaban and M.L. Overton
Submitted to SIAM J. Optimization.

Our paper builds on and proves a conjecture in:
Variational Analysis of Pseudospectra
A.S. Lewis and C.-H. J. Pang
SIAM J. Optimization (2008).

We also need a result in:
Characterization and Construction of the Nearest Defective Matrix via Coalescence of Pseudospectral Components
R. Alam, S. Bora, R. Byers and M.L. Overton
Linear Algebra and its Applications (2011).



Thanks a lot for your attention!