Optimization of Polynomial Roots, Eigenvalues and Pseudospectra

Michael L. Overton

Courant Institute of Mathematical Sciences
NYU

Banff Stability Workshop
Nov 5, 2012

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:

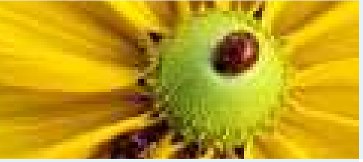
$$\rho(p) = \max\{|z| \ : \ p(z) = 0, \ z \in \mathbf{C}\}.$$

The Root Radius and the Root Abscissa

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

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

We say p is Schur stable if $\rho(p) < 1$.



The Root Radius and the Root Abscissa

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

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

We say p is Schur stable if $\rho(p) < 1$.

Let α denote the *root abscissa*:

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

The Root Radius and the Root Abscissa

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

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

We say p is Schur stable if $\rho(p) < 1$.

Let α denote the *root abscissa*:

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

We say p is Hurwitz stable if $\alpha(p) < 0$.

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

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.

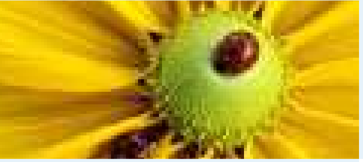
But if an affine family of monic polynomials of degree n has $n-1$ free parameters, this question can be answered efficiently.

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

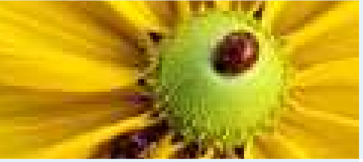
$$\rho^* := \inf_{p \in P} \rho(p)$$

has a globally optimal solution of the form

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

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

Proof: uses implicit function theorem.



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: for each $k = 0, \ldots, n$, substitute coefficients of $(z - \gamma)^{n-k} (z + \gamma)^k$ into the constraint to give a polynomial with n roots that are candidates for γ. Choose smallest such $|\gamma|$.

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

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

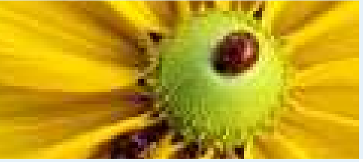
has an optimal solution of the form

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

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: A complicated inductive argument.



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*

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

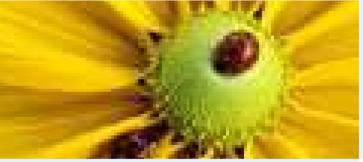
Optimizing the Root Abscissa: Real Case

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

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



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

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

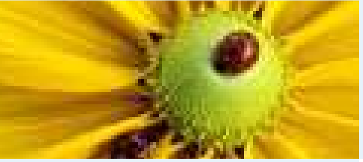
where $h^{(i)}$ is the i-th derivative of h.

Root Abscissa, Real Case, Continued

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 $\gamma = \alpha^*$.



Root Abscissa, Real Case, Continued

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 $\gamma = \alpha^*$.

When the optimal abscissa is not attained, for all $\epsilon > 0$ can find an approximately optimal polynomial

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

with $0 < \ell \leq n$ and $M_\epsilon \to -\infty$ as $\epsilon \to 0$.

Root Abscissa, Real Case, Continued

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

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

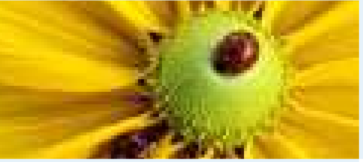
with $\gamma = \alpha^*$.

When the optimal abscissa is not attained, for all $\epsilon > 0$ can find an approximately optimal polynomial

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

with $0 < \ell \le n$ and $M_\epsilon \to -\infty$ as $\epsilon \to 0$.

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



Root Abscissa, Real Case, Continued

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 $\gamma = \alpha^*$.

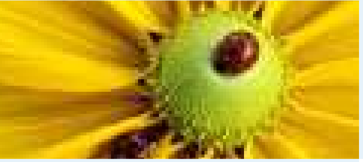
When the optimal abscissa is not attained, for all $\epsilon > 0$ can find an approximately optimal polynomial

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

with $0 < \ell \leq n$ and $M_\epsilon \to -\infty$ as $\epsilon \to 0$.

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

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



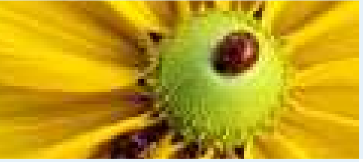
Optimizing the Abscissa: Real vs. Complex Case

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

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

Example: Static Output Feedback Stabilization

Given the dynamical system with input and output:

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

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

Example: Static Output Feedback Stabilization

Given the dynamical system with input and output:

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

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

SOF: find a controller $K \in \mathbf{R}^{\ell \times m}$ 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.

Example: Static Output Feedback Stabilization

Given the dynamical system with input and output:

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

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

SOF: find a controller $K \in \mathbf{R}^{\ell\times m}$ 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.

Example: Static Output Feedback Stabilization

Given the dynamical system with input and output:

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

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

SOF: find a controller $K \in \mathbf{R}^{\ell\times m}$ 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.

Example: Frequency Domain Stabilization

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

Example: 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.

Example: 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

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

Example: 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.

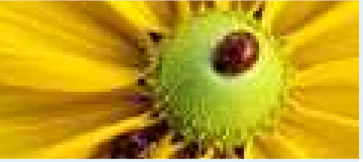
Then, maximizing the closed-loop asymptotic decay rate is equivalent to solving the optimization problem

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

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: see Part III.

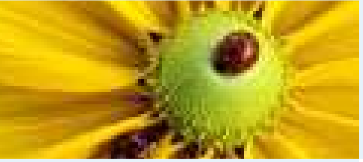
Caveats

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: see Part III.

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: see Part III.

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.

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: see Part III.

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.

AFFPOLYMIN: publicly available MATLAB code implementing the algorithms implicit in Theorems RRR, RRC, RAR, RAC.

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, to appear in 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, to appear in IEEE Trans. Auto. Control.

Based in part on a remarkable 1979 Ph.D. thesis by Raymond Chen, University of Florida.

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, to appear in IEEE Trans. Auto. Control.

Based in part on a remarkable 1979 Ph.D. thesis by Raymond Chen, University of Florida.

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

References for Part I

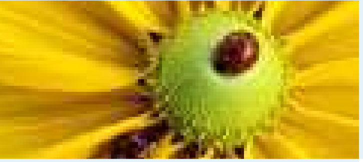
Explicit Solutions for Root Optimization of a Polynomial Family with One Affine Constraint
V.D. Blondel, M. Gürbüzbalaban, A. Megretski, M.L. Overton, to appear in IEEE Trans. Auto. Control.

Based in part on a remarkable 1979 Ph.D. thesis by Raymond Chen, University of Florida.

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/

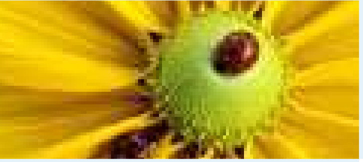


Part II
Optimization of Eigenvalues
with
J.V. Burke (Wash.)
K.K. Gade (NYU)
S. Grundel (NYU)
A.S. Lewis (Cornell)

The Spectral Radius and the Spectral Abscissa

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

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



The Spectral Radius and the Spectral Abscissa

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

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

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

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

The Spectral Radius and the Spectral Abscissa

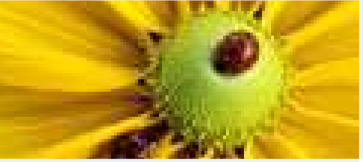
Now let $\rho : \mathbf{C}^{n\times n} \to \mathbf{R}$ denote *spectral radius*:

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

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

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

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*:

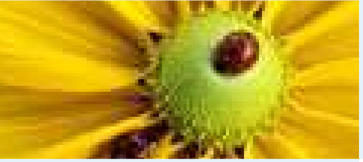
$$\rho(A) = \max \{ |z| \ : \ \det(A - zI) = 0, \ z \in \mathbf{C} \} .$$

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

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

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

$\rho(A) < 1$ is the stability condition for the discrete-time dynamical system $\xi_{k+1} = A\xi_k$.



The Spectral Radius and the Spectral Abscissa

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

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

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

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

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

$\rho(A) < 1$ is the stability condition for the discrete-time dynamical system $\xi_{k+1} = A\xi_k$.

$\alpha(A) < 0$ is the stability condition for the continuous-time dynamical system $\dot{\xi} = A\xi$.

The Spectral Radius and the Spectral Abscissa

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

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

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

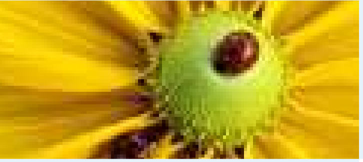
$$\alpha(A) = \max\{\mathrm{Re}(z) \ : \ \det(A - zI) = 0, \ z \in \mathbf{C}\}.$$

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

$\rho(A) < 1$ is the stability condition for the discrete-time dynamical system $\xi_{k+1} = A\xi_k$.

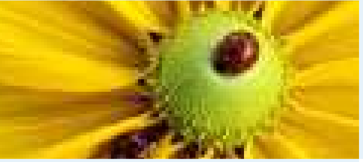
$\alpha(A) < 0$ is the stability condition for the continuous-time dynamical system $\dot{\xi} = A\xi$.

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



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.



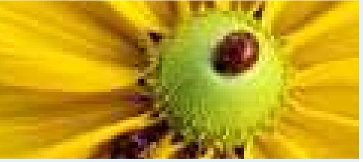
No Extension of Part I

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

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

This matrix depends affinely on a single parameter x, 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

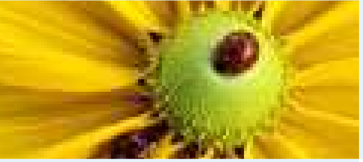
The spectral radius and abscissa are the radius and abscissa of the characteristic polynomial of a matrix, but the results of Part I do not extend to the more general case of an affine family of $n \times n$ matrices depending on $n - 1$ parameters.

For example, consider the matrix family

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

This matrix depends affinely on a single parameter x, 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 $A(x)$ is attained by $x = 0$, 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

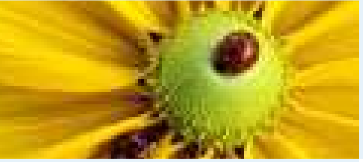
$$\begin{bmatrix} 0 & 1-x & & & & & & & x & 0 \\ & & 1-x & & & & & x & & \\ & & & \ddots & & & \cdot^{\cdot^{\cdot}} & & & \\ & & & & 1-x & x & & & & \\ & & & & x & 1-x & & & & \\ & & & \cdot^{\cdot^{\cdot}} & & & \ddots & & & \\ & & x & & & & & 1-x & & \\ & x & & & & & & & 1-x & \\ x & & & & & & & &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

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

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

The Reduced Spectral Radius

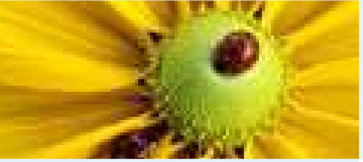
The rate of convergence is determined by

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

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 (the ones with largest modulus excluding 1) 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 (the ones with largest modulus excluding 1) all occur in conjugate pairs.

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

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

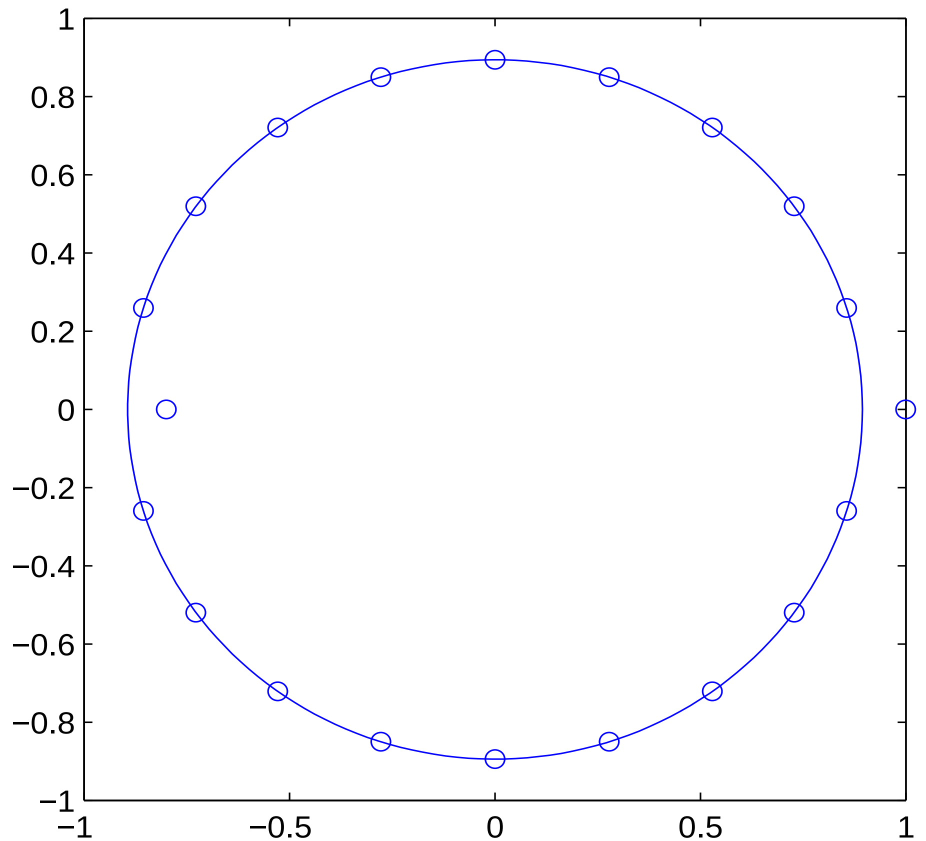
$$x_{\text{opt}} = \frac{\sin(\pi/n)}{1 + \sin(\pi/n)} > \frac{1}{n}.$$

For $x < x_{\text{opt}}$, the active eigenvalues (the ones with largest modulus excluding 1) all occur in conjugate pairs.

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

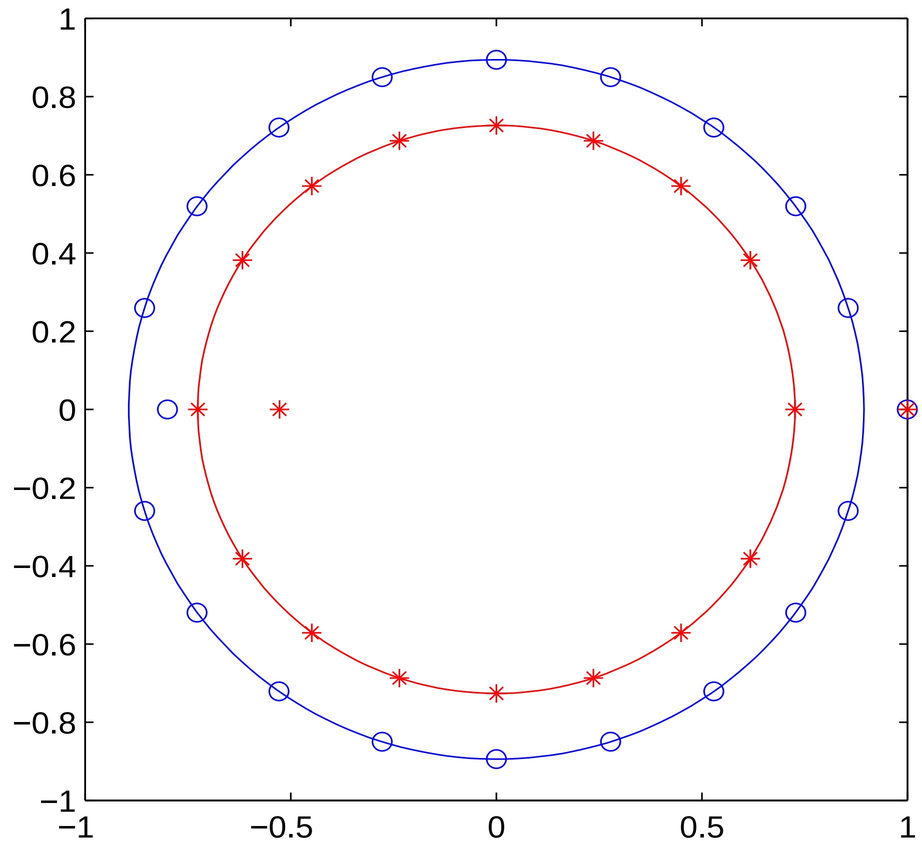
For $x > x_{\text{opt}}$, this splits into two real eigenvalues, increasing $\tilde{\rho}$ by $O(|x - x_{\text{opt}}|^{1/2})$.

Eigenvalues of the Transition Matrix, $n = 10$

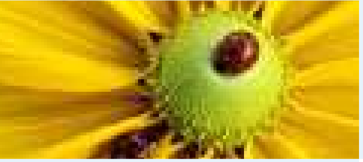


- Blue: eigenvalues when $x = 1/n$ (all complex)

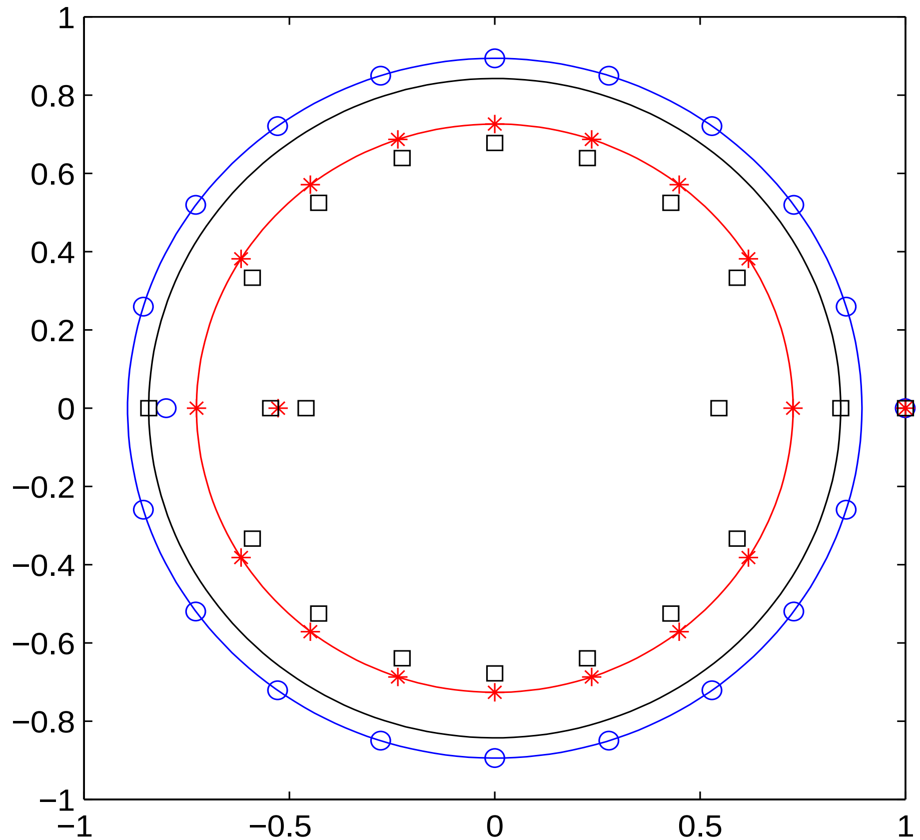
Eigenvalues of the Transition Matrix, $n = 10$



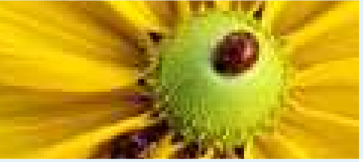
- Blue: eigenvalues when $x = 1/n$ (all complex)
- Red: eigenvalues when $x = x_{\text{opt}}$ (one double real eigenvalue)



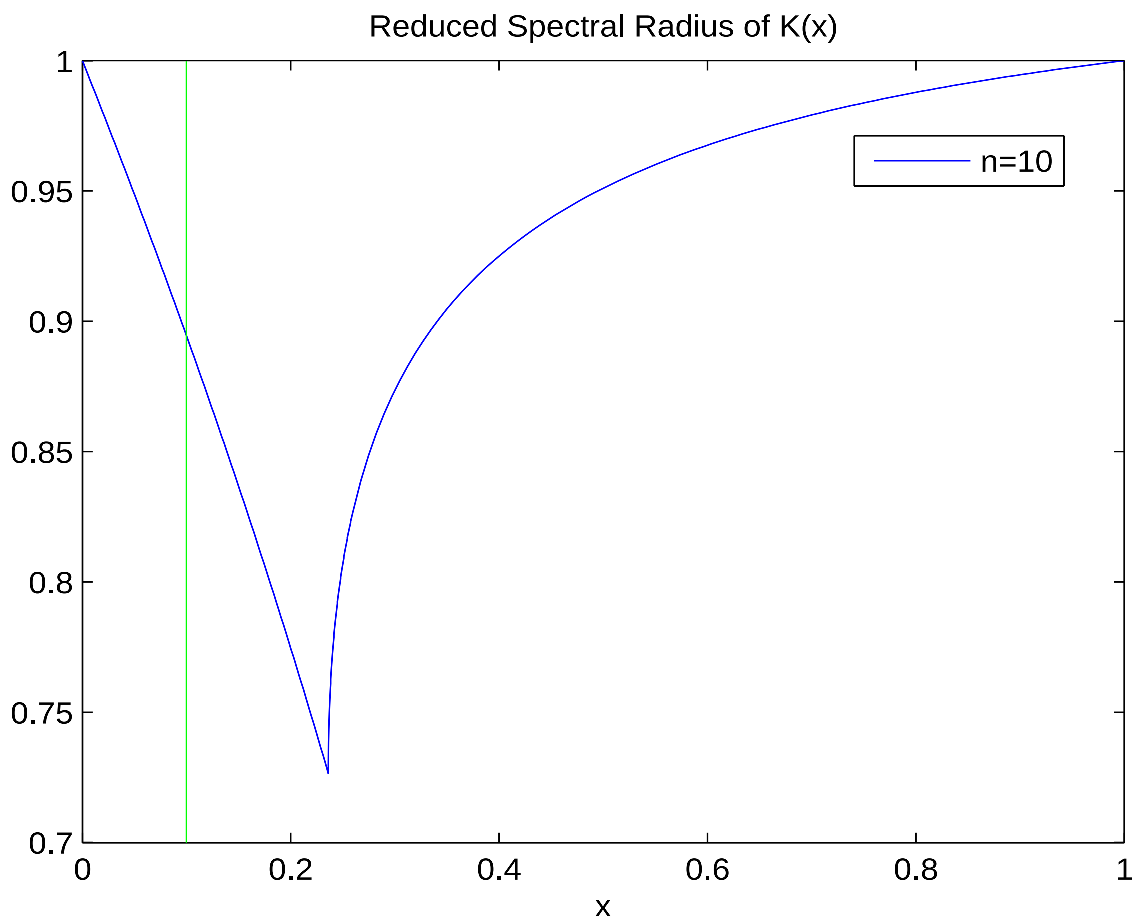
Eigenvalues of the Transition Matrix, $n = 10$



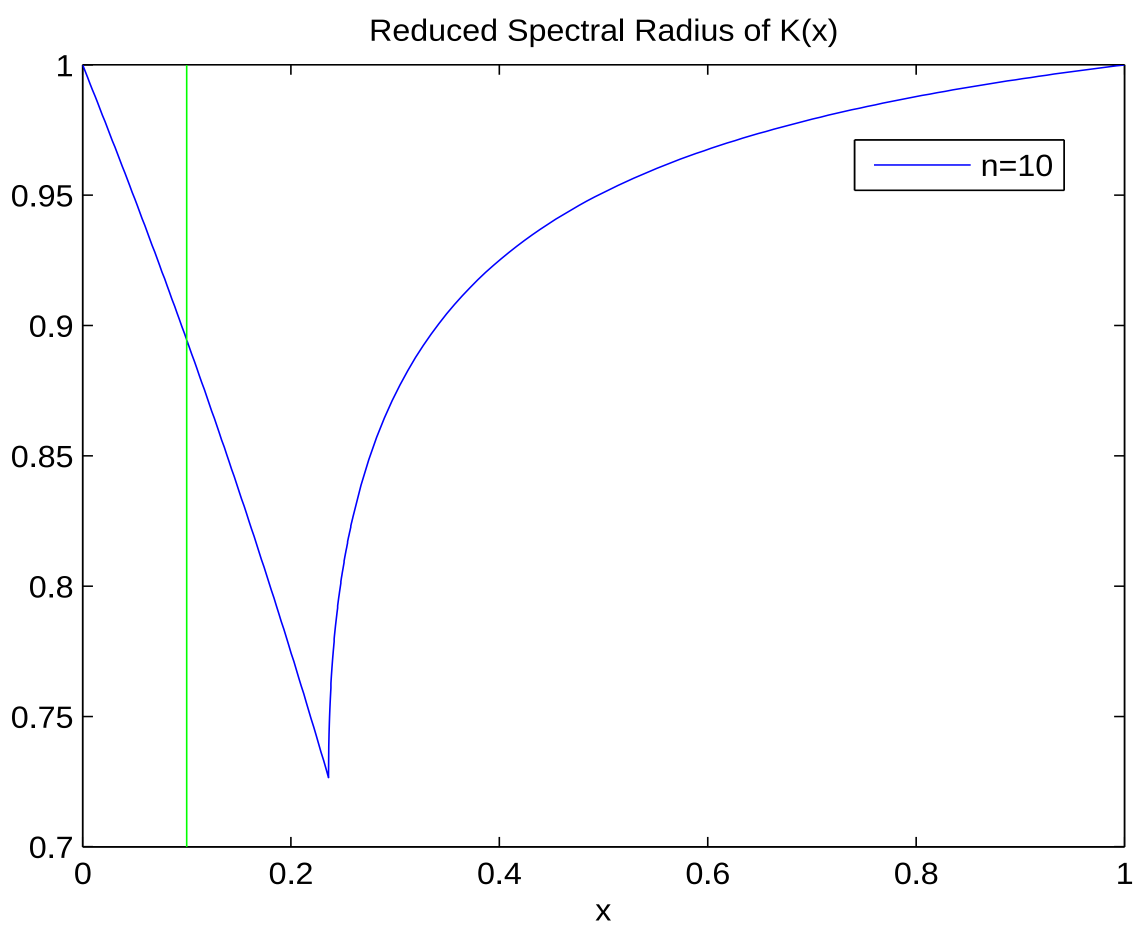
- Blue: eigenvalues when $x = 1/n$ (all complex)
- Red: eigenvalues when $x = x_{\text{opt}}$ (one double real eigenvalue)
- Black: eigenvalues when $x > x_{\text{opt}}$ ($\tilde{\rho}$ increases rapidly)



Reduced Spectral Radius as a Function of x

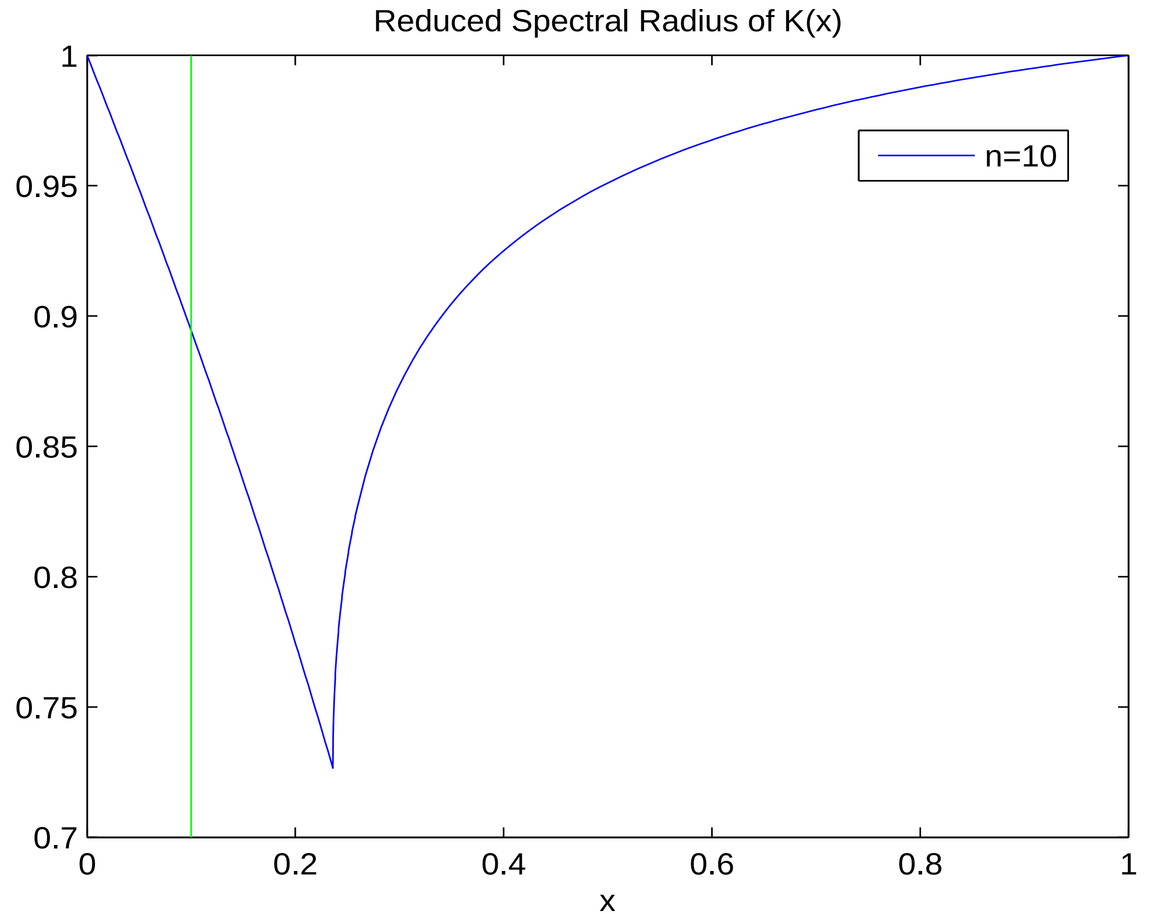


Reduced Spectral Radius as a Function of x



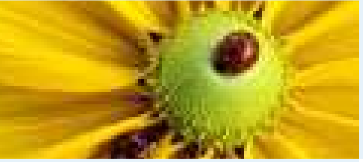
Note the big improvement changing x from $1/n$ to x_{opt}.

Reduced Spectral Radius as a Function of x

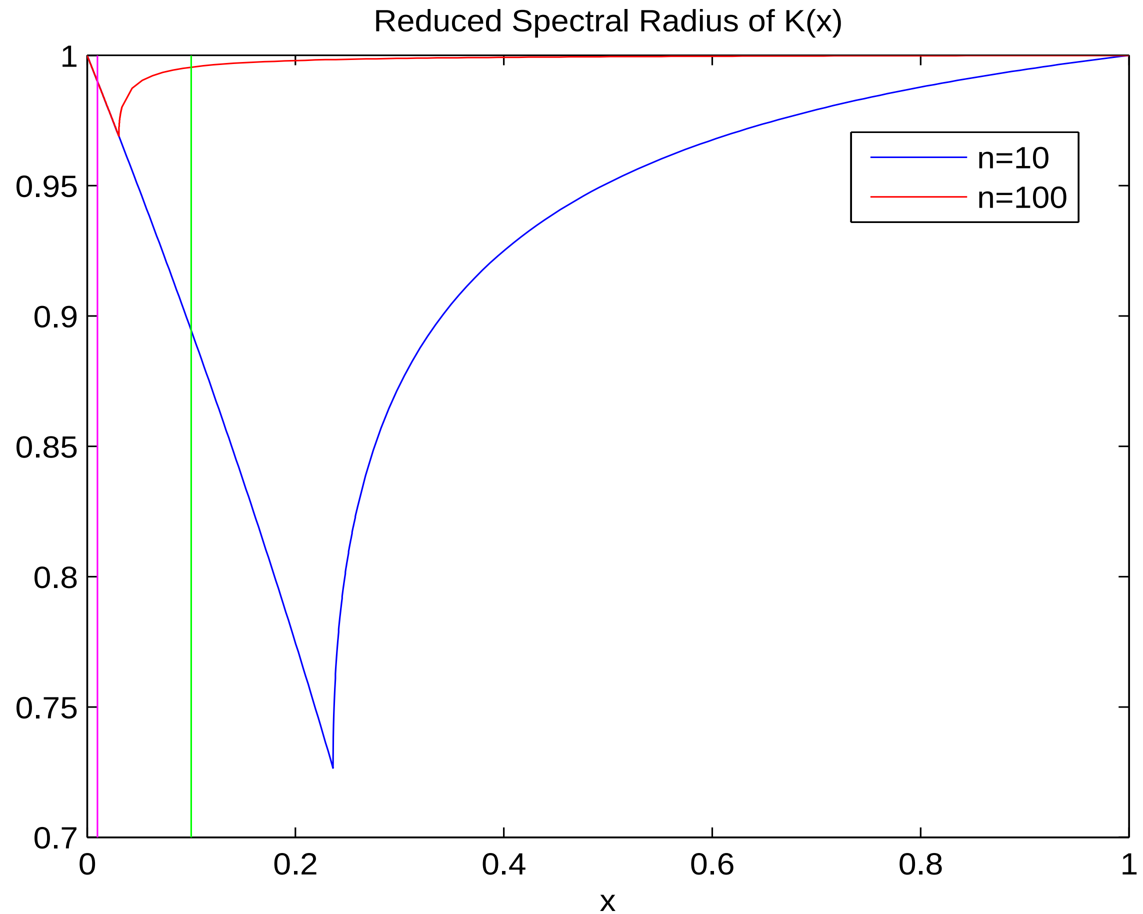


Note the big improvement changing x from $1/n$ to x_{opt}.

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 $x_{\rm opt}$.

Much better to underestimate $x_{\rm opt}$ 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 $A(x)$ to have multiple parameters:

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

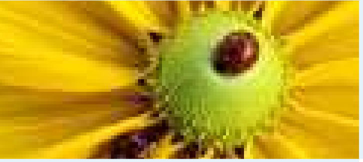
Adding More Parameters

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

Let's change $A(x)$ 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?



Adding More Parameters

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

Let's change $A(x)$ 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 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 $\mathbf{x}_{\mathrm{opt}}$. We repeatedly obtained convergence to $\mathbf{x}_{\mathrm{opt}}$.

Theoretically: by variational analysis. We found that

- $\mathbf{x}_{\mathrm{opt}}$ satisfies a *necessary* 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 repeatedly obtained convergence to $\mathbf{x}_{\rm opt}$.

Theoretically: by variational 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.

Checking Local Optimality

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

Theoretically: by variational 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.

Special analysis needed because optimization objective is not Lipschitz at a matrix with an active multiple eigenvalue.

Checking Local Optimality

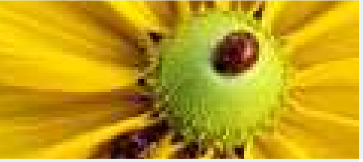
Numerically: by running an optimization method suitable for nonsmooth objectives at randomly generated points near $\mathbf{x}_{\rm opt}$. We repeatedly obtained convergence to $\mathbf{x}_{\rm opt}$.

Theoretically: by variational 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.

Special analysis needed because optimization objective is not Lipschitz at a matrix with an active multiple eigenvalue.

Essential to the analysis: each active eigenvalue corresponds to a single Jordan block, in this case with sizes 2, 1, . . . , 1.



Checking Local Optimality

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

Theoretically: by variational analysis. We found that

- $\mathbf{x}_{\mathrm{opt}}$ 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 $\mathbf{x}_{\mathrm{opt}}$ satisfies a *sufficient* condition for local optimality.

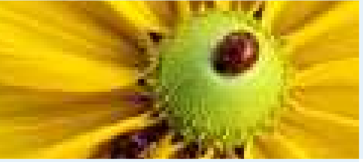
Special analysis needed because optimization objective is not Lipschitz at a matrix with an active multiple eigenvalue.

Essential to the analysis: each active eigenvalue corresponds to a single Jordan block, in this case with sizes 2, 1, . . . , 1.

Too complicated to explain in talk, but see references for more information.

Surface Approximation By Subdivision

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



Surface Approximation By Subdivision

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

- triangular mesh case: optimal multiple zero eigenvalue verified analytically, with multiple Jordan blocks of order 2, 1, 1, 1.

Surface Approximation By Subdivision

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

- triangular mesh case: optimal multiple zero eigenvalue verified analytically, with multiple Jordan blocks of order 2, 1, 1, 1.
- quadrilateral mesh case: numerically reduced moduli of eigenvalues to about 10^{-4} and estimated that the apparently optimal multiple zero eigenvalue has multiple Jordan blocks of order 5, 3, 2, 2.

Surface Approximation By Subdivision

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

- triangular mesh case: optimal multiple zero eigenvalue verified analytically, with multiple Jordan blocks of order 2, 1, 1, 1.
- quadrilateral mesh case: numerically reduced moduli of eigenvalues to about 10^{-4} and estimated that the apparently optimal multiple zero eigenvalue has multiple Jordan blocks of order 5, 3, 2, 2.

In both cases, the active eigenvalue zero has not only algebraic multiplicity > 1 but also geometric multiplicity > 1. The latter results from special structure and will not occur generically.

Ordinary gradient method with line search: fails, typically converges to some arbitrary point where f is not differentiable.

Numerical Optimization of Nonsmooth, Nonconvex f

Ordinary gradient method with line search: fails, typically converges to some arbitrary point where f is not differentiable.

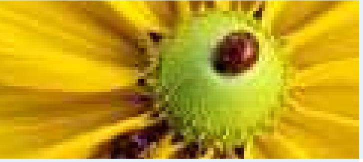
Gradient sampling with line search: for locally Lipschitz f can prove convergence to nonsmooth stationary points of f (typically local minimizers where f is not differentiable).

Numerical Optimization of Nonsmooth, Nonconvex f

Ordinary gradient method with line search: fails, typically converges to some arbitrary point where f is not differentiable.

Gradient sampling with line search: for locally Lipschitz f can prove convergence to nonsmooth stationary points of f (typically local minimizers where f is not differentiable).

BFGS quasi-Newton method with line search: empirically same property with much less computation.



Numerical Optimization of Nonsmooth, Nonconvex f

Ordinary gradient method with line search: fails, typically converges to some arbitrary point where f is not differentiable.

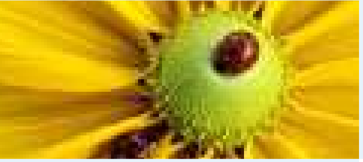
Gradient sampling with line search: for locally Lipschitz f can prove convergence to nonsmooth stationary points of f (typically local minimizers where f is not differentiable).

BFGS quasi-Newton method with line search: empirically same property with much less computation.

When using these methods to minimize the nonsmooth, nonconvex, non-Lipschitz functions $\rho(A(x))$ or $\alpha(A(x))$, *make no attempt to predict active eigenvalues or estimate their multiplicities*; just use gradients which exist at almost every x

$$\frac{\partial}{\partial x_k}\alpha\left(A(x)\right) = \left\langle \frac{\partial A}{\partial x_k}(x), \frac{1}{v^* u} v u^* \right\rangle = \mathrm{Re}\frac{u^* \frac{\partial A}{\partial x_k}(x) v}{u^* v}$$

where v and u are *right and left eigenvectors for the rightmost eigenvalue* λ.



Numerical Optimization of Nonsmooth, Nonconvex f

Ordinary gradient method with line search: fails, typically converges to some arbitrary point where f is not differentiable.

Gradient sampling with line search: for locally Lipschitz f can prove convergence to nonsmooth stationary points of f (typically local minimizers where f is not differentiable).

BFGS quasi-Newton method with line search: empirically same property with much less computation.

When using these methods to minimize the nonsmooth, nonconvex, non-Lipschitz functions $\rho(A(x))$ or $\alpha(A(x))$, *make no attempt to predict active eigenvalues or estimate their multiplicities*; just use gradients which exist at almost every x

$$\frac{\partial}{\partial x_k}\alpha\left(A(x)\right) = \left\langle \frac{\partial A}{\partial x_k}(x), \frac{1}{v^* u} v u^* \right\rangle = \mathrm{Re}\frac{u^* \frac{\partial A}{\partial x_k}(x) v}{u^* v}$$

where v and u are *right and left eigenvectors for the rightmost eigenvalue* λ.

HANSO (Hybrid Algorithm for Nonsmooth Optimization): publicly available MATLAB software.

References for Part II

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

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

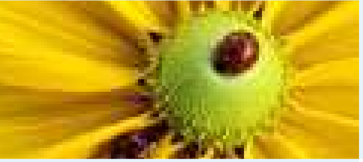
References for Part II

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

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

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.



References for Part II

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

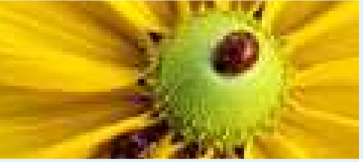
Variational Analysis of Non-Lipschitz Spectral Functions J.V. Burke and M.L. Overton, Math. Programming (2001)

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.

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, to appear in *Math. Programming*.

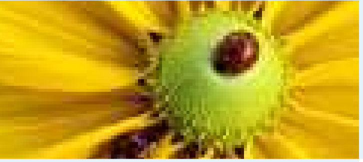


Part III
Optimization of Pseudospectra
with
J.V. Burke (U. Wash.)
N.Guglielmi (L'Aquila)
M. Gürbüzbalaban (NYU)
A.S. Lewis (Cornell)

Pseudospectra

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

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



Pseudospectra

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

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

A more robust measure of system behaviour than eigenvalues.

Pseudospectra

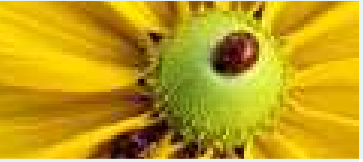
The area swept out in the complex plane by the eigenvalues under perturbation.

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

A more robust measure of system behaviour than eigenvalues.

For $\|\cdot\| = \|\cdot\|_2$,

$$\sigma_\epsilon(A) \quad = \quad \left\{z \in \mathbf{C} \;:\; \left\|(A - zI)^{-1}\right\| \geq \epsilon^{-1}\right\}$$



Pseudospectra

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

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

A more robust measure of system behaviour than eigenvalues.

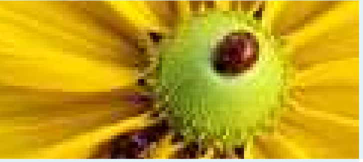
For $\|\cdot\| = \|\cdot\|_2$,

$$\begin{aligned} \sigma_\epsilon(A) &= \left\{z \in \mathbf{C} \ : \ \left\|(A - zI)^{-1}\right\| \geq \epsilon^{-1}\right\} \\ &= \left\{z \in \mathbf{C} \ : \ s_n(A - zI) \leq \epsilon\right\} \end{aligned}$$

where s_n denotes smallest singular value:

$$A - zI = U \operatorname{diag}(s) V^*$$

with $U^*U = V^*V = I$.



Pseudospectra

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

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

A more robust measure of system behaviour than eigenvalues.

For $\|\cdot\| = \|\cdot\|_2$,

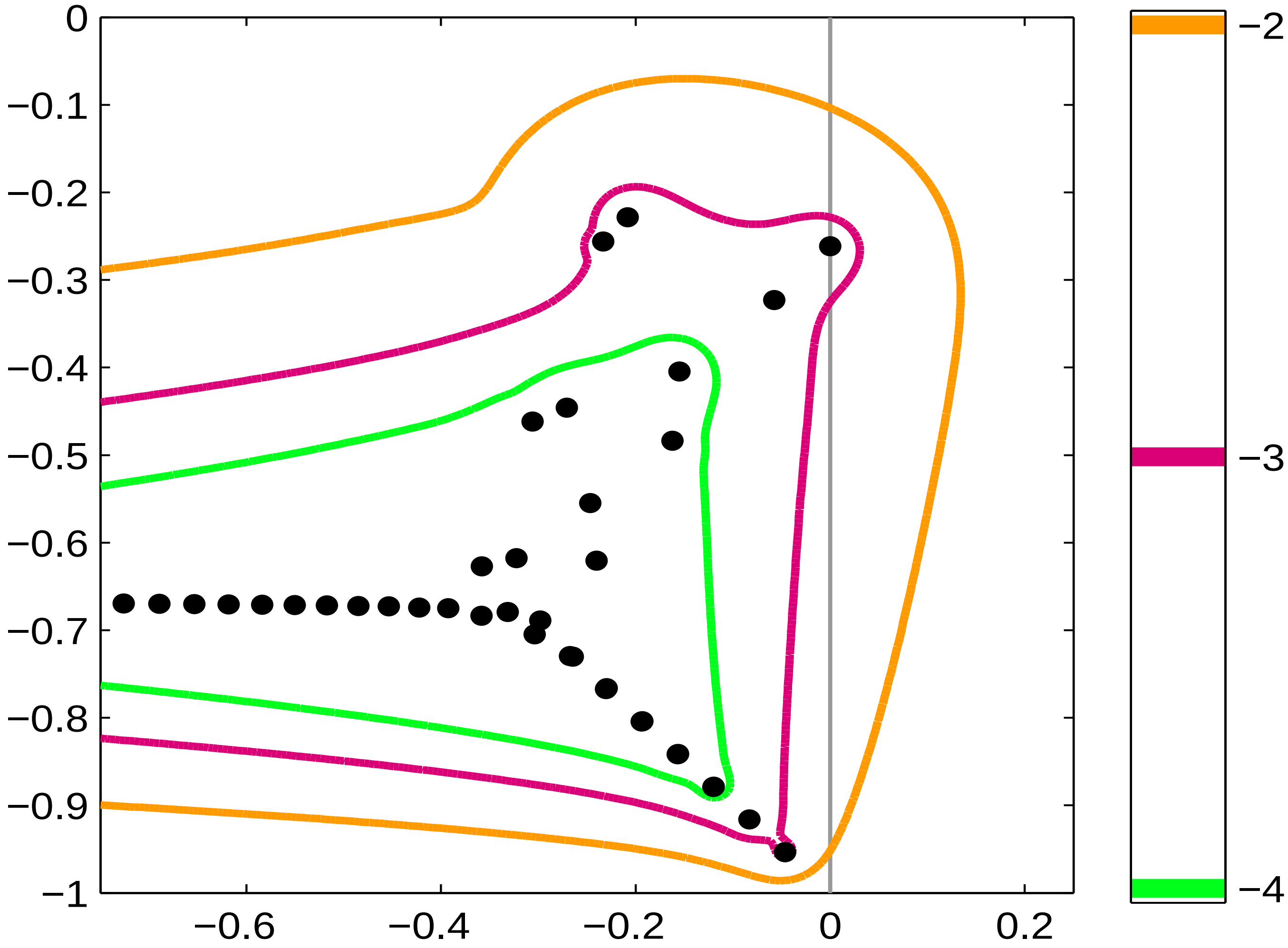
$$\begin{aligned}\sigma_\epsilon(A) &= \left\{z \in \mathbf{C} \ : \ \left\|(A - zI)^{-1}\right\| \ge \epsilon^{-1}\right\} \\ &= \left\{z \in \mathbf{C} \ : \ s_n(A - zI) \le \epsilon\right\}\end{aligned}$$

where s_n denotes smallest singular value:

$$A - zI = U \mathrm{diag}(s) V^*$$

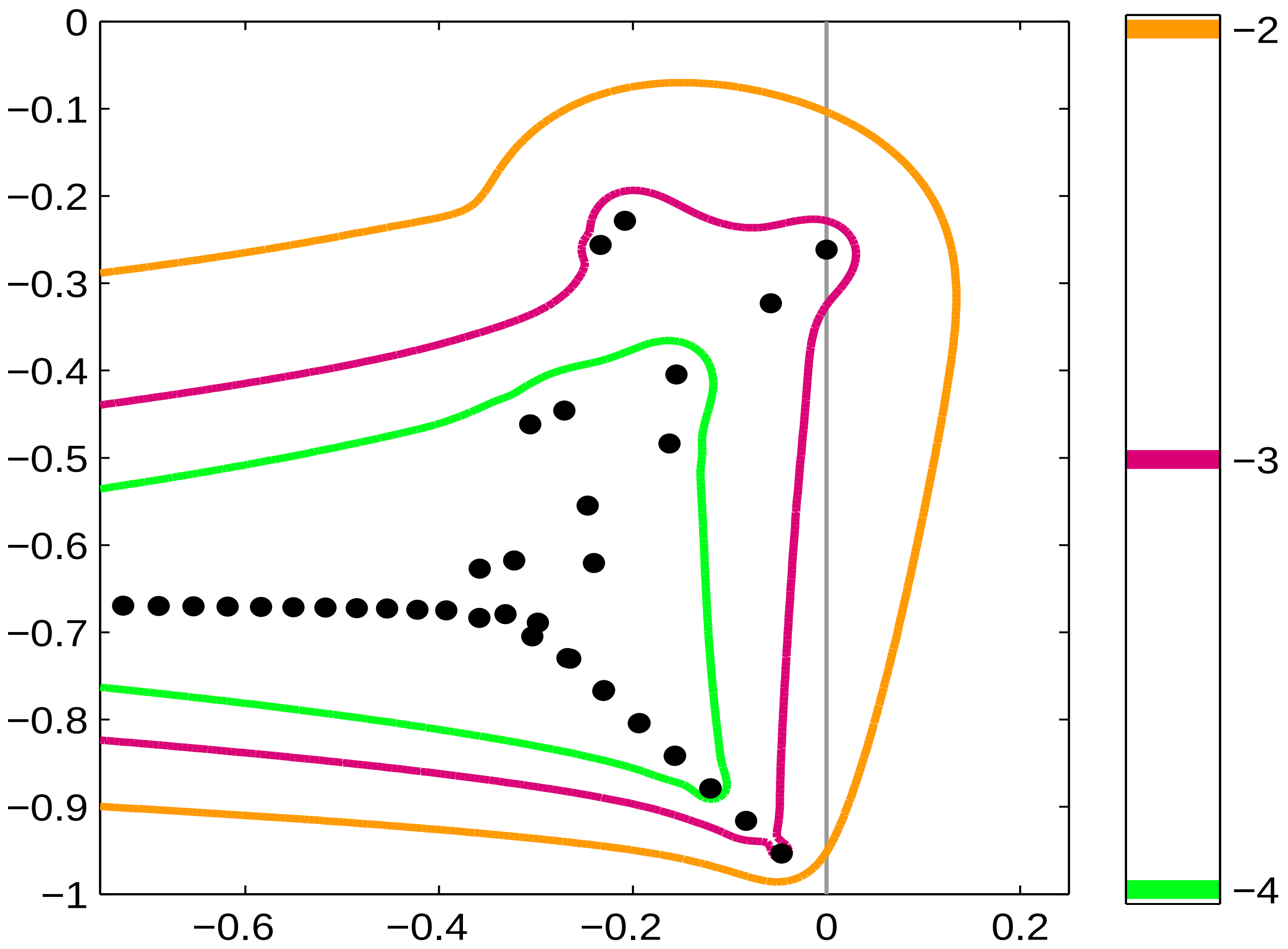
with $U^*U = V^*V = I$.

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



0
−0.1
−0.2
−0.3
−0.4
−0.5
−0.6
−0.7
−0.8
−0.9
−1
−0.6
−0.4
−0.2
0
0.2
−2
−3
−4

Orr-Sommerfeld Matrix ($n = 99$, $\epsilon = 10^{-4}, 10^{-3}, 10^{-2}$)



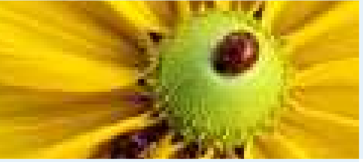
Black dots are eigenvalues and colored curves are pseudospectral boundaries. Note the pseudospectra are not convex.

Constructing E given $z \in \partial\sigma_\epsilon(A)$

Let

$$A - zI = U\mathrm{diag}(s)V^* = \sum_{j=1}^{n} s_j u_j v_j^*, \quad s_n = \epsilon$$

with $U^*U = V^*V = I$.



Constructing E given $z \in \partial\sigma_\epsilon(A)$

Let

$$A - zI = U\mathrm{diag}(s)V^* = \sum_{j=1}^{n} s_j u_j v_j^*, \quad s_n = \epsilon$$

with $U^*U = V^*V = I$.

Then if we set $u = u_n$, $v = v_n$, $E = -\epsilon u v^*$ we have

$$\det(A - zI + E) = 0$$

so z is an eigenvalue of $A + E$.

Constructing E given $z \in \partial\sigma_\epsilon(A)$

Let

$$A - zI = U\mathrm{diag}(s)V^* = \sum_{j=1}^{n} s_j u_j v_j^*, \quad s_n = \epsilon$$

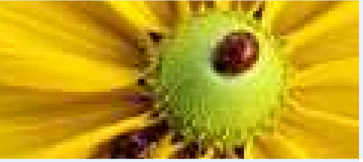
with $U^*U = V^*V = I$.

Then if we set $u = u_n$, $v = v_n$, $E = -\epsilon u v^*$ we have

$$\det(A - zI + E) = 0$$

so z is an eigenvalue of $A + E$.

Key point: can choose E to have rank one.



Constructing E given $z \in \partial\sigma_\epsilon(A)$

Let

$$A - zI = U\mathrm{diag}(s)V^* = \sum_{j=1}^{n} s_j u_j v_j^*, \quad s_n = \epsilon$$

with $U^*U = V^*V = I$.

Then if we set $u = u_n$, $v = v_n$, $E = -\epsilon u v^*$ we have

$$\det(A - zI + E) = 0$$

so z is an eigenvalue of $A + E$.

Key point: can choose E to have rank one. Furthermore

$$(A - zI)v = \epsilon u, \quad u^*(A - zI) = \epsilon v^*$$

Constructing E given $z \in \partial\sigma_\epsilon(A)$

Let

$$A - zI = U\mathrm{diag}(s)V^* = \sum_{j=1}^{n} s_j u_j v_j^*, \quad s_n = \epsilon$$

with $U^*U = V^*V = I$.

Then if we set $u = u_n$, $v = v_n$, $E = -\epsilon u v^*$ we have

$$\det(A - zI + E) = 0$$

so z is an eigenvalue of $A + E$.

Key point: can choose E to have rank one. Furthermore

$$(A - zI)v = \epsilon u, \quad u^*(A - zI) = \epsilon v^*$$

so

$$(A - zI + E)v = 0, \quad u^*(A - zI + E) = 0.$$

Thus the right and left singular vectors of $A - zI$ for the singular value ϵ are also right and left eigenvectors of $A + E$ for the eigenvalue z.

Pseudospectral Radius and Abscissa

Pseudospectral radius: modulus of outermost point in $\sigma_\epsilon(A)$

$$\rho_\epsilon(A) = \max\{|z| \ : \ z \in \sigma_\epsilon(A)\}$$

Pseudospectral Radius and Abscissa

Pseudospectral radius: modulus of outermost point in $\sigma_\epsilon(A)$

$$\rho_\epsilon(A) = \max\{|z| \ : \ z \in \sigma_\epsilon(A)\}$$

Pseudospectral abscissa: real part of rightmost point in $\sigma_\epsilon(A)$

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

Pseudospectral Radius and Abscissa

Pseudospectral radius: modulus of outermost point in $\sigma_\epsilon(A)$

$$\rho_\epsilon(A) = \max\{|z| \;:\; z \in \sigma_\epsilon(A)\}$$

Pseudospectral abscissa: real part of rightmost point in $\sigma_\epsilon(A)$

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

Computing these quantities: nontrivial because $\sigma_\epsilon(A)$ is not convex.

Pseudospectral Radius and Abscissa

Pseudospectral radius: modulus of outermost point in $\sigma_\epsilon(A)$

$$\rho_\epsilon(A) = \max\{|z| \ : \ z \in \sigma_\epsilon(A)\}$$

Pseudospectral abscissa: real part of rightmost point in $\sigma_\epsilon(A)$

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

Computing these quantities: nontrivial because $\sigma_\epsilon(A)$ is not convex.

Criss-cross algorithm for computing the pseudospectral abscissa $\alpha_\epsilon(A)$: based on repeatedly computing eigenvalues of $2n \times 2n$ Hamiltonian matrices and checking whether any are imaginary, and computing SVDs for each imaginary eigenvalue.

Pseudospectral Radius and Abscissa

Pseudospectral radius: modulus of outermost point in $\sigma_\epsilon(A)$

$$\rho_\epsilon(A) = \max\{|z| \ : \ z \in \sigma_\epsilon(A)\}$$

Pseudospectral abscissa: real part of rightmost point in $\sigma_\epsilon(A)$

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

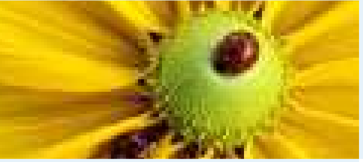
Computing these quantities: nontrivial because $\sigma_\epsilon(A)$ is not convex.

Criss-cross algorithm for computing the pseudospectral abscissa $\alpha_\epsilon(A)$: based on repeatedly computing eigenvalues of $2n \times 2n$ Hamiltonian matrices and checking whether any are imaginary, and computing SVDs for each imaginary eigenvalue.

Too expensive if n large.

Approximating the Pseudospectral Abscissa if n is Big

We want a rightmost point z of $\sigma_\epsilon(A)$, so $s_n(A - zI) = \epsilon$. Let v and u be corresponding right and left singular vectors. We know that z is an eigenvalue of $B = A - \epsilon uv^*$ with right and left eigenvectors v and u.



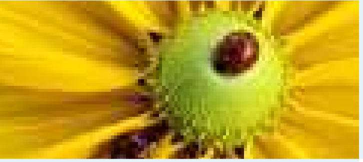
Approximating the Pseudospectral Abscissa if n is Big

We want a rightmost point z of $\sigma_\epsilon(A)$, so $s_n(A - zI) = \epsilon$. Let v and u be corresponding right and left singular vectors. We know that z is an eigenvalue of $B = A - \epsilon uv^*$ with right and left eigenvectors v and u.

Let us generate a sequence

$$B^{(k)} = A - \epsilon u^{(k)} \left(v^{(k)}\right)^*$$

with $\left\|u^{(k)}\right\| = \left\|v^{(k)}\right\| = 1$. We want $u^{(k)} \to u$, $v^{(k)} \to v$.



Approximating the Pseudospectral Abscissa if n is Big

We want a rightmost point z of $\sigma_\epsilon(A)$, so $s_n(A - zI) = \epsilon$. Let v and u be corresponding right and left singular vectors. We know that z is an eigenvalue of $B = A - \epsilon uv^*$ with right and left eigenvectors v and u.

Let us generate a sequence

$$B^{(k)} = A - \epsilon u^{(k)} \left(v^{(k)}\right)^*$$

with $\left\|u^{(k)}\right\| = \left\|v^{(k)}\right\| = 1$. We want $u^{(k)} \to u$, $v^{(k)} \to v$.

No Hamiltonian eigenvalue decompositions or SVDs allowed. The only matrix operations are the computation of eigenvalues with largest real part and their corresponding right and left eigenvectors, which can be done efficiently using the implicitly restarted Arnoldi method (ARPACK).

RP-Compatible Right and Left Eigenvectors

A pair of right and left eigenvectors p and q for a simple eigenvalue λ is called *RP-compatible* if $\|p\| = \|q\| = 1$ and p^*q is real and positive, and therefore in the interval $(0, 1]$.

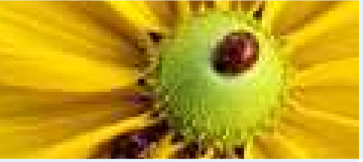
RP-Compatible Right and Left Eigenvectors

A pair of right and left eigenvectors p and q for a simple eigenvalue λ is called *RP-compatible* if $\|p\| = \|q\| = 1$ and p^*q is real and positive, and therefore in the interval $(0, 1]$.

This defines right and left eigenvectors uniquely up to $p \leftarrow e^{i\theta}p$, $q \leftarrow e^{i\theta}q$.

New Algorithm to Approximate $\alpha_\epsilon(A)$

1. Let $z^{(0)}$ be a rightmost eigenvalue of A, with RP-compatible right and left eigenvectors $v^{(0)}$ and $u^{(0)}$. Set $B^{(0)} = A - \epsilon u^{(0)} \left(v^{(0)}\right)^*$.



New Algorithm to Approximate $\alpha_\epsilon(A)$

1. Let $z^{(0)}$ be a rightmost eigenvalue of A, with RP-compatible right and left eigenvectors $v^{(0)}$ and $u^{(0)}$. Set $B^{(0)} = A - \epsilon u^{(0)} \left(v^{(0)}\right)^*$.

2. For $k = 1, 2, \ldots$: let $z^{(k)}$ be a rightmost eigenvalue of $B^{(k)}$ with RP-compatible right and left eigenvectors $v^{(k)}$ and $u^{(k)}$. Set $B^{(k+1)} = A - \epsilon u^{(k)} \left(v^{(k)}\right)^*$.

New Algorithm to Approximate $\alpha_\epsilon(A)$

1. Let $z^{(0)}$ be a rightmost eigenvalue of A, with RP-compatible right and left eigenvectors $v^{(0)}$ and $u^{(0)}$. Set $B^{(0)} = A - \epsilon u^{(0)} \left(v^{(0)}\right)^*$.

2. For $k = 1, 2, \ldots$: let $z^{(k)}$ be a rightmost eigenvalue of $B^{(k)}$ with RP-compatible right and left eigenvectors $v^{(k)}$ and $u^{(k)}$. Set $B^{(k+1)} = A - \epsilon u^{(k)} \left(v^{(k)}\right)^*$.

Clearly, Re $z^{(k)} \leq \alpha_\epsilon(A)$ for all k.

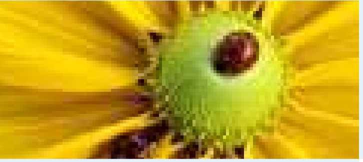
New Algorithm to Approximate $\alpha_\epsilon(A)$

1. Let $z^{(0)}$ be a rightmost eigenvalue of A, with RP-compatible right and left eigenvectors $v^{(0)}$ and $u^{(0)}$. Set $B^{(0)} = A - \epsilon u^{(0)} \left(v^{(0)}\right)^*$.

2. For $k = 1, 2, \ldots$: let $z^{(k)}$ be a rightmost eigenvalue of $B^{(k)}$ with RP-compatible right and left eigenvectors $v^{(k)}$ and $u^{(k)}$. Set $B^{(k+1)} = A - \epsilon u^{(k)} \left(v^{(k)}\right)^*$.

Clearly, Re $z^{(k)} \leq \alpha_\epsilon(A)$ for all k.

Almost always: $z^{(k)} \to z$, a locally rightmost point of $\sigma_\epsilon(A)$, and $v^{(k)}$ and $u^{(k)}$ converge to right and left singular vectors v and u corresponding to smallest singular value of $A - zI$.



New Algorithm to Approximate $\alpha_\epsilon(A)$

1. Let $z^{(0)}$ be a rightmost eigenvalue of A, with RP-compatible right and left eigenvectors $v^{(0)}$ and $u^{(0)}$. Set $B^{(0)} = A - \epsilon u^{(0)} \left(v^{(0)}\right)^*$.

2. For $k = 1, 2, \ldots$: let $z^{(k)}$ be a rightmost eigenvalue of $B^{(k)}$ with RP-compatible right and left eigenvectors $v^{(k)}$ and $u^{(k)}$. Set $B^{(k+1)} = A - \epsilon u^{(k)} \left(v^{(k)}\right)^*$.

Clearly, Re $z^{(k)} \leq \alpha_\epsilon(A)$ for all k.

Almost always: $z^{(k)} \to z$, a locally rightmost point of $\sigma_\epsilon(A)$, and $v^{(k)}$ and $u^{(k)}$ converge to right and left singular vectors v and u corresponding to smallest singular value of $A - zI$.

Often, but not always, z is a globally rightmost point so Re $z = \alpha_\epsilon(A)$.

New Algorithm to Approximate $\alpha_\epsilon(A)$

1. Let $z^{(0)}$ be a rightmost eigenvalue of A, with RP-compatible right and left eigenvectors $v^{(0)}$ and $u^{(0)}$. Set $B^{(0)} = A - \epsilon u^{(0)} \left(v^{(0)}\right)^*$.

2. For $k = 1, 2, \ldots$: let $z^{(k)}$ be a rightmost eigenvalue of $B^{(k)}$ with RP-compatible right and left eigenvectors $v^{(k)}$ and $u^{(k)}$. Set $B^{(k+1)} = A - \epsilon u^{(k)} \left(v^{(k)}\right)^*$.

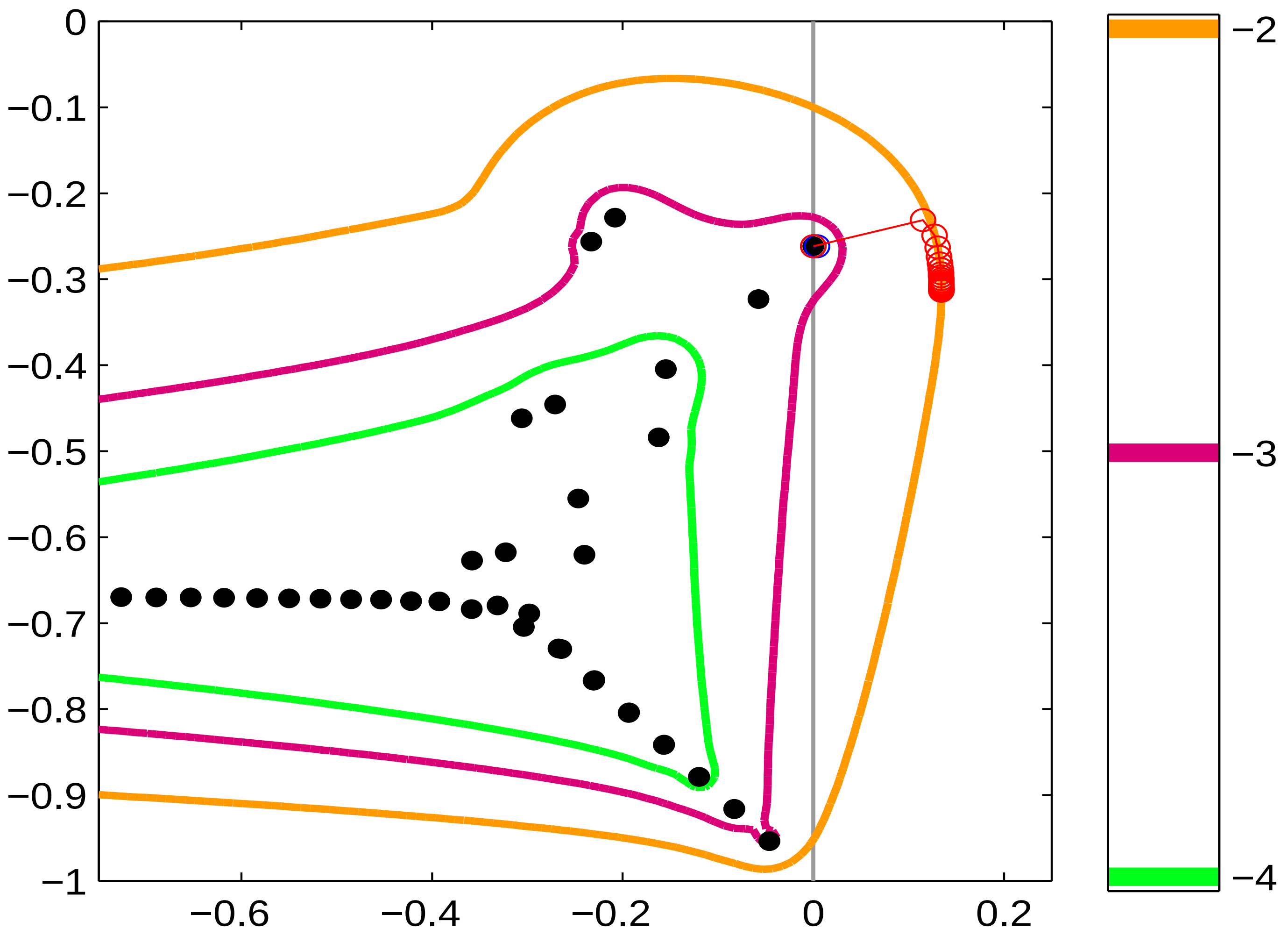
Clearly, Re $z^{(k)} \le \alpha_\epsilon(A)$ for all k.

Almost always: $z^{(k)} \to z$, a locally rightmost point of $\sigma_\epsilon(A)$, and $v^{(k)}$ and $u^{(k)}$ converge to right and left singular vectors v and u corresponding to smallest singular value of $A - zI$.

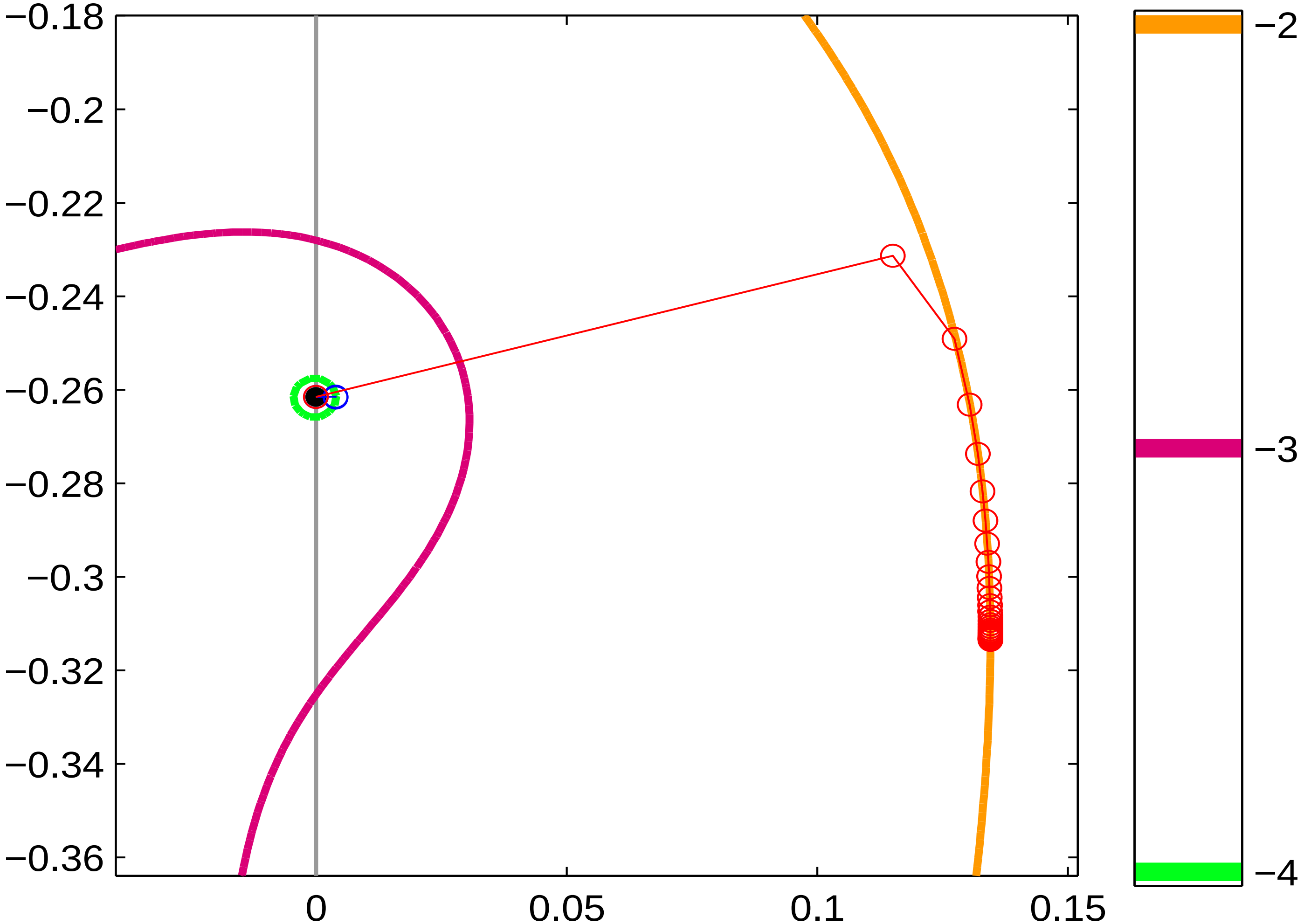
Often, but not always, z is a globally rightmost point so Re $z = \alpha_\epsilon(A)$.

We have theorems characterizing fixed points of the algorithm and proving local convergence at a geometric rate for ϵ small.

Orr-Sommerfeld Matrix ($n = 99$, $\epsilon = 10^{-4}, 10^{-2}$)



Orr-Sommerfeld Matrix ($n = 99$, $\epsilon = 10^{-4}, 10^{-2}$)



Minimizing $\alpha_\epsilon(A(x))$ over Parametrized Matrix $A(x)$

For given x in parameter space $\mathbf{R}^p$, compute $\alpha_\epsilon(A(x))$ by criss-cross algorithm if n small and otherwise by the new algorithm.

Minimizing $\alpha_\epsilon(A(x))$ over Parametrized Matrix $A(x)$

For given x in parameter space $\mathbf{R}^p$, compute $\alpha_\epsilon(A(x))$ by criss-cross algorithm if n small and otherwise by the new algorithm.

Like α, α_ϵ is nonsmooth and nonconvex, but unlike α, it is locally Lipschitz for $\epsilon > 0$ (although σ_ϵ is not).

Minimizing $\alpha_\epsilon(A(x))$ over Parametrized Matrix $A(x)$

For given x in parameter space $\mathbf{R}^p$, compute $\alpha_\epsilon(A(x))$ by criss-cross algorithm if n small and otherwise by the new algorithm.

Like α, α_ϵ is nonsmooth and nonconvex, but unlike α, it is locally Lipschitz for $\epsilon > 0$ (although σ_ϵ is not).

Derivatives:

$$\frac{\partial}{\partial x_k}\alpha_\epsilon(A(x)) = \left\langle \frac{\partial A}{\partial x_k}(x), \frac{1}{v^*u}vu^* \right\rangle = \mathrm{Re}\frac{u^*\frac{\partial A}{\partial x_k}(x)v}{u^*v}$$

where v and u are right and left singular vectors for the singular value ϵ of $A - zI$ with z the rightmost point of $\sigma_\epsilon(A)$, equivalently RP-compatible right and left eigenvectors for the eigenvalue z of $A - \epsilon uv^*$.

Minimizing $\alpha_\epsilon(A(x))$ over Parametrized Matrix $A(x)$

For given x in parameter space $\mathbf{R}^p$, compute $\alpha_\epsilon(A(x))$ by criss-cross algorithm if n small and otherwise by the new algorithm.

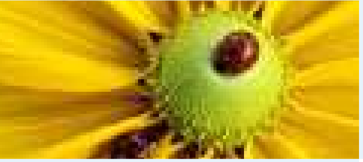
Like α, α_ϵ is nonsmooth and nonconvex, but unlike α, it is locally Lipschitz for $\epsilon > 0$ (although σ_ϵ is not).

Derivatives:

$$\frac{\partial}{\partial x_k}\alpha_\epsilon(A(x)) = \left\langle \frac{\partial A}{\partial x_k}(x), \frac{1}{v^* u} v u^* \right\rangle = \mathrm{Re}\,\frac{u^* \frac{\partial A}{\partial x_k}(x) v}{u^* v}$$

where v and u are right and left singular vectors for the singular value ϵ of $A - zI$ with z the rightmost point of $\sigma_\epsilon(A)$, equivalently RP-compatible right and left eigenvectors for the eigenvalue z of $A - \epsilon u v^*$.

As earlier, use Gradient Sampling or BFGS.



Minimizing $\alpha_\epsilon(A(x))$ over Parametrized Matrix $A(x)$

For given x in parameter space $\mathbf{R}^p$, compute $\alpha_\epsilon(A(x))$ by criss-cross algorithm if n small and otherwise by the new algorithm.

Like α, α_ϵ is nonsmooth and nonconvex, but unlike α, it is locally Lipschitz for $\epsilon > 0$ (although σ_ϵ is not).

Derivatives:

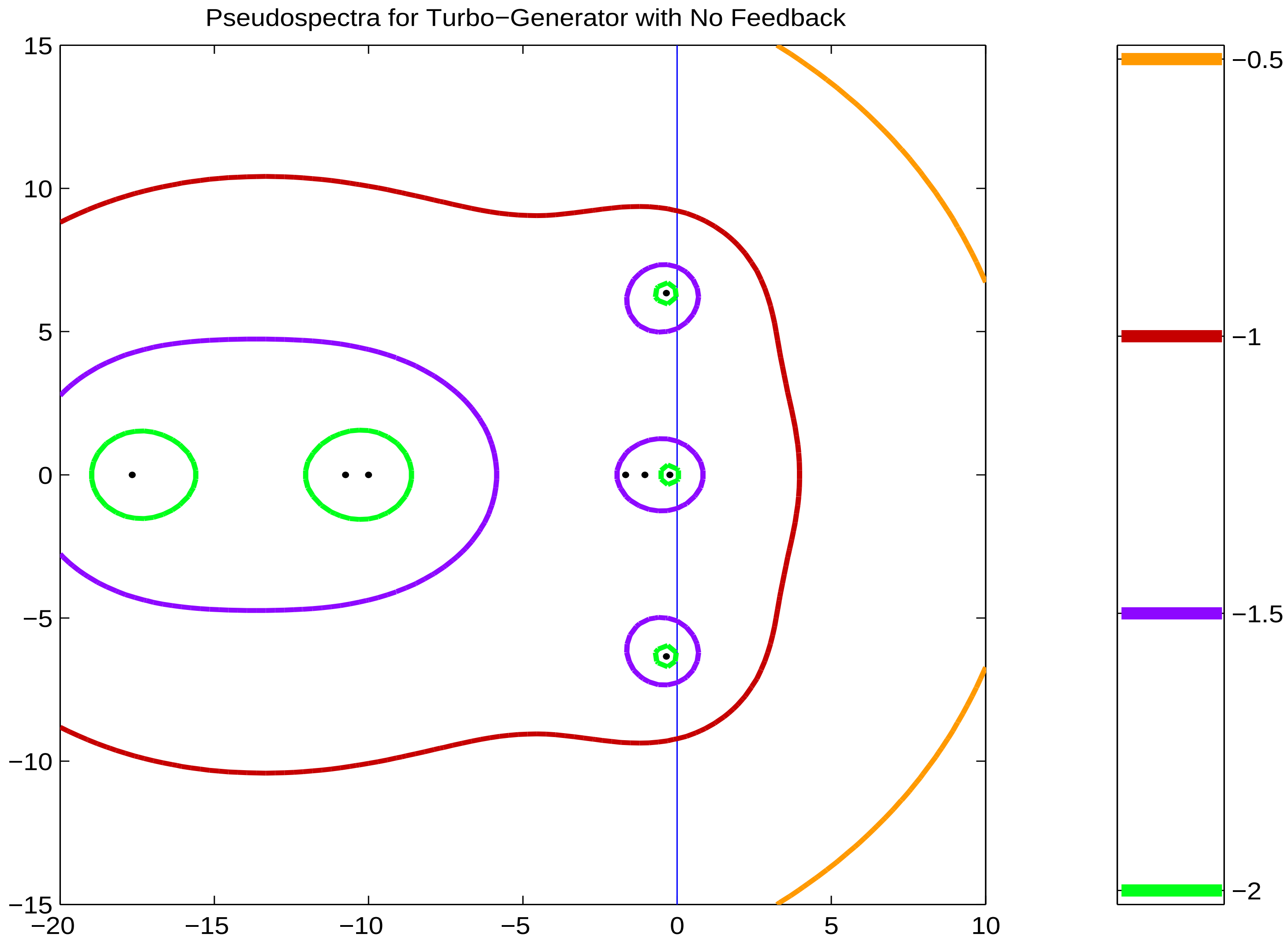
$$\frac{\partial}{\partial x_k}\alpha_\epsilon(A(x)) = \left\langle \frac{\partial A}{\partial x_k}(x), \frac{1}{v^* u} v u^* \right\rangle = \mathrm{Re}\frac{u^* \frac{\partial A}{\partial x_k}(x) v}{u^* v}$$

where v and u are right and left singular vectors for the singular value ϵ of $A - zI$ with z the rightmost point of $\sigma_\epsilon(A)$, equivalently RP-compatible right and left eigenvectors for the eigenvalue z of $A - \epsilon u v^*$.

As earlier, use Gradient Sampling or BFGS.

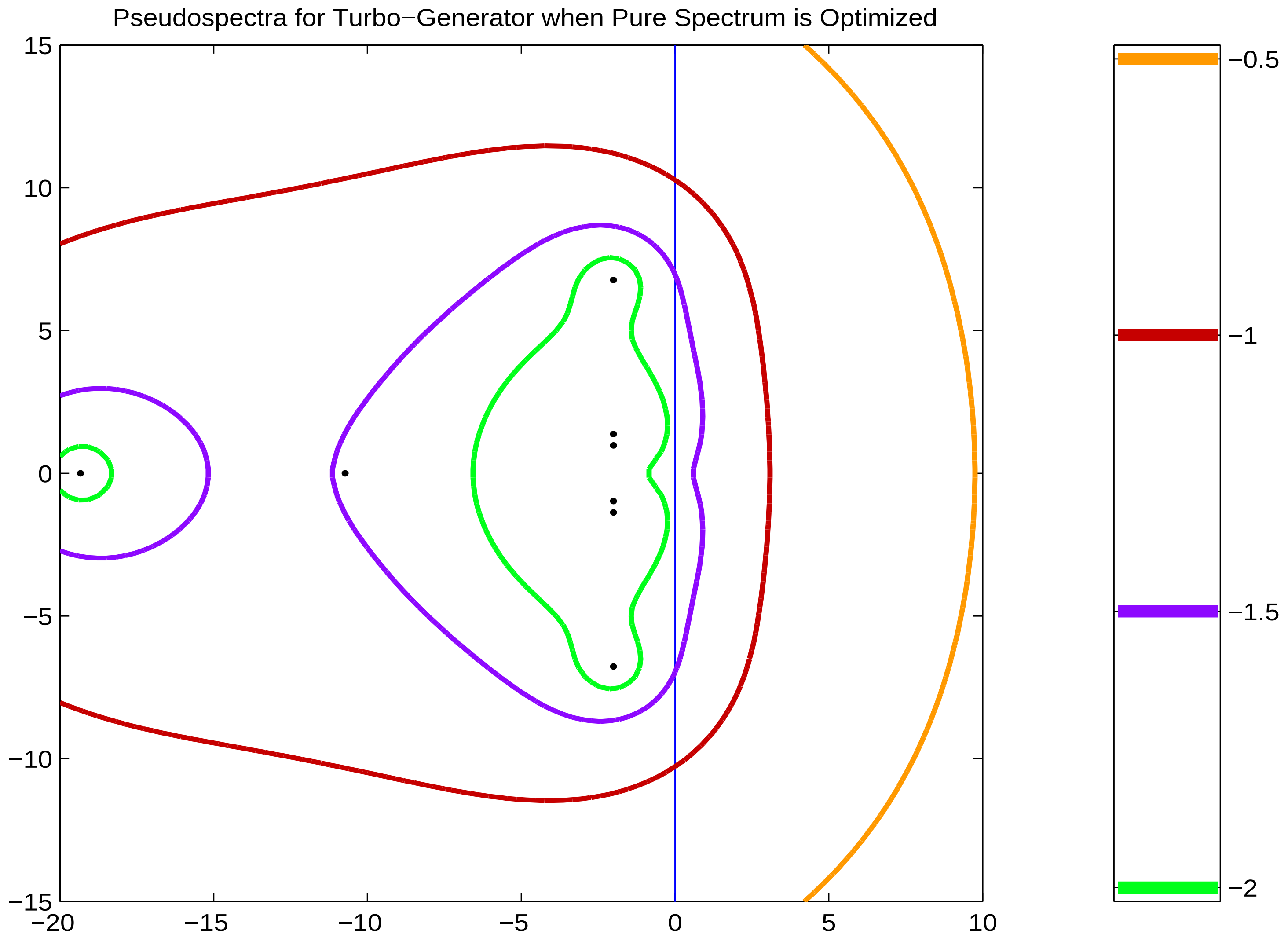
Example: $A(x) = F + GKH$ with $x = \mathrm{vec}(K)$, a static output feedback control design problem for a turbo generator with $n = 10$, $\ell = m = 2$, so controller $K \in \mathbf{R}^{2\times 2}$.

A Turbo Generator Control Problem



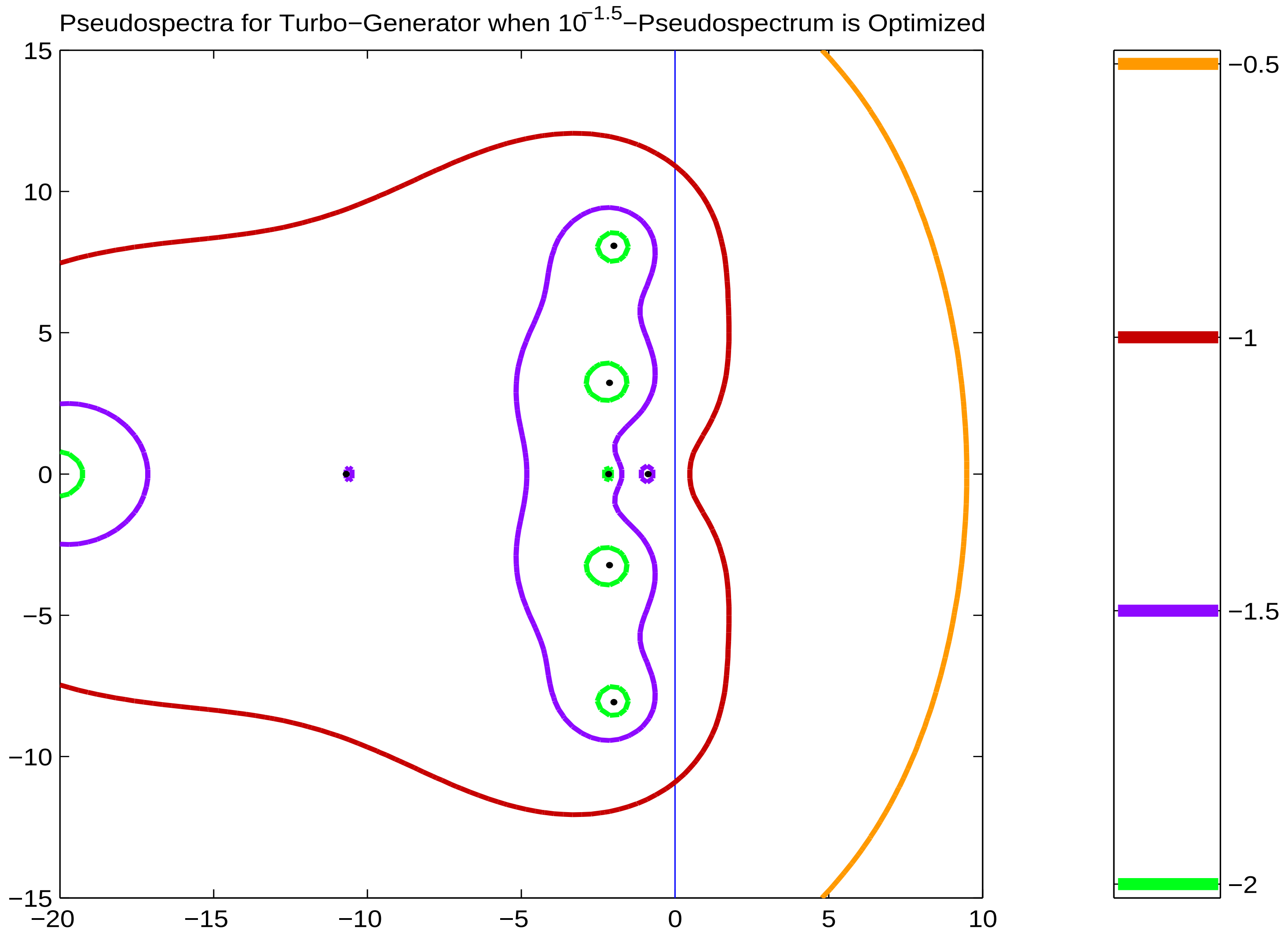
Pseudospectra for open-loop turbo generator plant with no feedback.

Turbo Generator with Optimized Eigenvalues



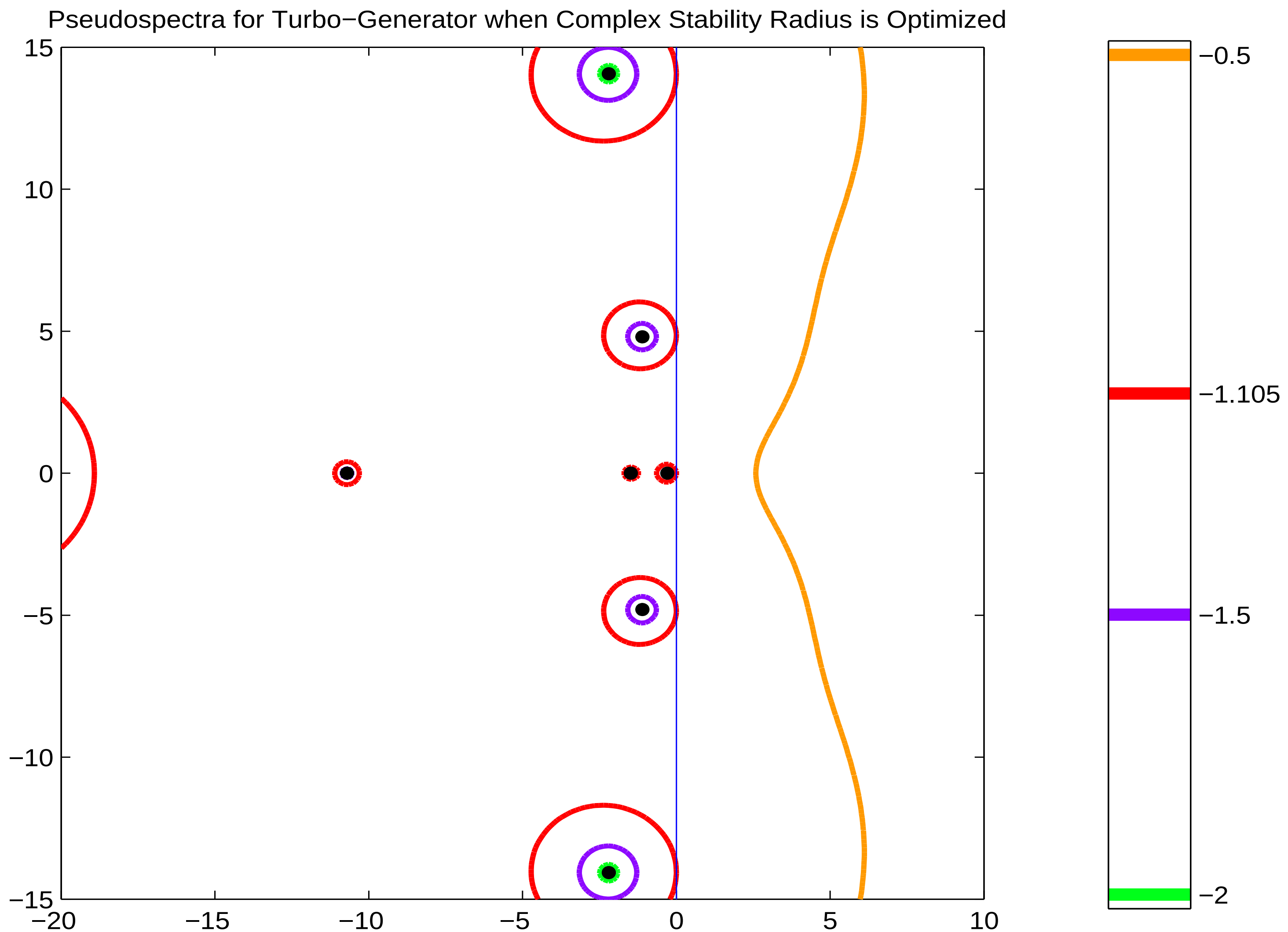
Pseudospectra for turbo generator plant with feedback computed by minimizing the spectral abscissa α

Turbo Generator with Optimized ϵ-Pseudospectrum



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

Turbo Generator with Optimized Dist. to Instability



Pseudospectra for turbo generator plant with feedback computed by maximizing the *distance to instability*: largest ϵ so that $\alpha_\epsilon(A(x)) \leq 0$.

References for Part III

Origins of Pseudospectra in 1980s:
Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, ?.

References for Part III

Origins of Pseudospectra in 1980s:
Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, ?.

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

References for Part III

Origins of Pseudospectra in 1980s:
Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, ?.

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

Optimization and Pseudospectra, with Applications to Robust Stability, J.V. Burke, A.S. Lewis and M.L. Overton, SIAM J. Matrix Anal. Appl. (2003).

A Nonsmooth, Nonconvex Optimization Approach to Robust Stabilization by Static Output Feedback and Low-Order Controllers, J.V. Burke, A.S. Lewis and M.L. Overton, Fourth IFAC Symposium on Robust Control Design, Milan (2003).

Fast algorithms for the approximation of the pseudospectral abscissa and pseudospectral radius of a matrix, N. Guglielmi and M.L. Overton, SIAM J. Matrix Anal. Appl. (2011)

Some Regularity Results for the Pseudospectral Abscissa and Pseudospectral Radius of a Matrix, M. Gürbüzbalaban and M.L. Overton, SIAM J. Optimization (2012)

References for Part III

Origins of Pseudospectra in 1980s:
Landau, Varah, Godunov, Demmel, Wilkinson, Trefethen, ?.

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

Optimization and Pseudospectra, with Applications to Robust Stability, J.V. Burke, A.S. Lewis and M.L. Overton, SIAM J. Matrix Anal. Appl. (2003).

A Nonsmooth, Nonconvex Optimization Approach to Robust Stabilization by Static Output Feedback and Low-Order Controllers, J.V. Burke, A.S. Lewis and M.L. Overton, Fourth IFAC Symposium on Robust Control Design, Milan (2003).

Fast algorithms for the approximation of the pseudospectral abscissa and pseudospectral radius of a matrix, N. Guglielmi and M.L. Overton, SIAM J. Matrix Anal. Appl. (2011)

Some Regularity Results for the Pseudospectral Abscissa and Pseudospectral Radius of a Matrix, M. Gürbüzbalaban and M.L. Overton, SIAM J. Optimization (2012)

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

Thanks a lot for your attention!