Optimization of Matrix Stability

Michael L. Overton
Courant Institute of Mathematical Sciences
New York University

Joint work with
Jim Burke, University of Washington
Adrian Lewis, Simon Fraser University

Outline

- Pure spectral functions
- Pseudo-spectral functions
- Distance to instability and H_∞ norm
- Variational properties
- Optimization over parameters
- A gradient sampling algorithm
- Applications, including static output feedback and low-order controllers

Pure Spectral Functions

Functions of the form: $\phi \circ \Lambda$

where Λ maps a matrix to its spectrum:

$$\Lambda(A) = \{z: \ \sigma_{min}(A - z\,I) = 0\}$$

Here σ_{min} means smallest singular value.

The elements of $\Lambda(A)$ are eigenvalues of A.

Important examples:

Spectral abscissa $\alpha = (\max \mathrm{Re}) \circ \Lambda$

Spectral radius $\rho = (\max \mathrm{mod}) \circ \Lambda$

Pure Spectral Modeling

- Pure spectral functions model asymptotic behavior
- The spectral abscissa models asymptotic growth/decay of the solutions of continuous time dynamical systems
- E.g: $z'(t) = A\, z(t),\; z(0) = z_0$
- Solution: $z(t) = \exp(t\, A)\, z_0$
- Norm: $\|z(t)\| / \exp(\gamma\, t) \to 0$ for all $\gamma > \alpha(A)$
- Stable if $\alpha(A) < 0$
- The spectral radius models asymptotic growth/decay of the solutions of discrete time dynamical systems

Pseudo-Spectral Functions

Functions of the form: $\phi \circ \Lambda_\varepsilon$

where Λ_ε maps a matrix to its pseudo-spectrum:

$$\Lambda_\varepsilon(A) = \{z: \sigma_{min}(A - zI) \leq \varepsilon\}$$

or equivalently (Trefethen)

$$\Lambda_\varepsilon(A) = \{z: \sigma_{min}(B - zI) = 0, \text{ for a B with } \|B-A\| \leq \varepsilon\}$$

The points in $\Lambda_\varepsilon(A)$ are pseudo-eigenvalues of A.

Important examples:

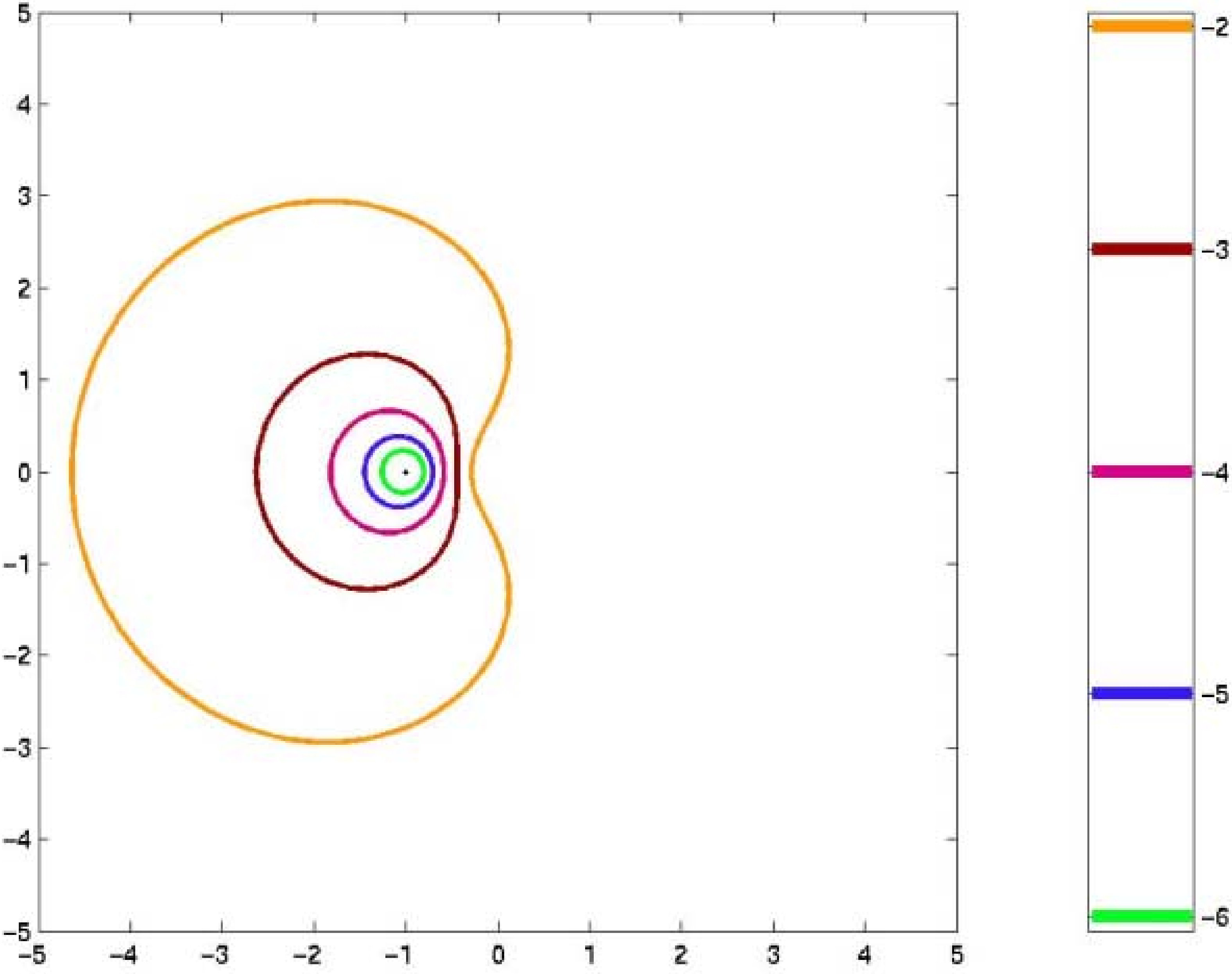
- Pseudo-spectral abscissa $\alpha_\varepsilon = (\max \mathrm{Re}) \circ \Lambda_\varepsilon$
- Pseudo-spectral radius $\rho_\varepsilon = (\max \mathrm{mod}) \circ \Lambda_\varepsilon$

Pseudo-Spectral Modeling

- Pseudo-spectral functions model worst-case asymptotic behavior in nearby matrices
- Equivalently, they model transient behavior (can be quantified by the Kreiss matrix theorem)

The Pseudo-Spectral GUI (T. Wright)

- Example due to Demmel: An upper triangular matrix with constant diagonal…



5
4
3
2
1
0
−1
−2
−3
−4
−5
−5 −4 −3 −2 −1 0 1 2 3 4 5
−2
−3
−4
−5
−6

Distance to Instability δ

- Let A be a stable matrix, so $\alpha(A) < 0$
- Its distance to instability $\delta(A)$ is the largest ε such that $\alpha_\varepsilon(A) \leq 0$
- Equivalently, is the largest ε such that $\alpha(A + E) \leq 0$ for all E with $\|E\| \leq \varepsilon$
- AKA "complex stability radius"
- AKA inverse of H_∞ norm of corresponding transfer matrix

Well known Equivalent Properties

- The pseudo-spectral abscissa $\alpha_\varepsilon(A) < 0$
- The spectral abscissa $\alpha(A) < 0$ and the distance to instability $\delta(A) < \varepsilon$ (that is, H_∞ norm > $1/\varepsilon$)
- The spectral abscissa $\alpha(A) < 0$ and the Hamiltonian matrix

$$\begin{matrix} -A^* & \varepsilon I \\ -\varepsilon I & A \end{matrix}$$

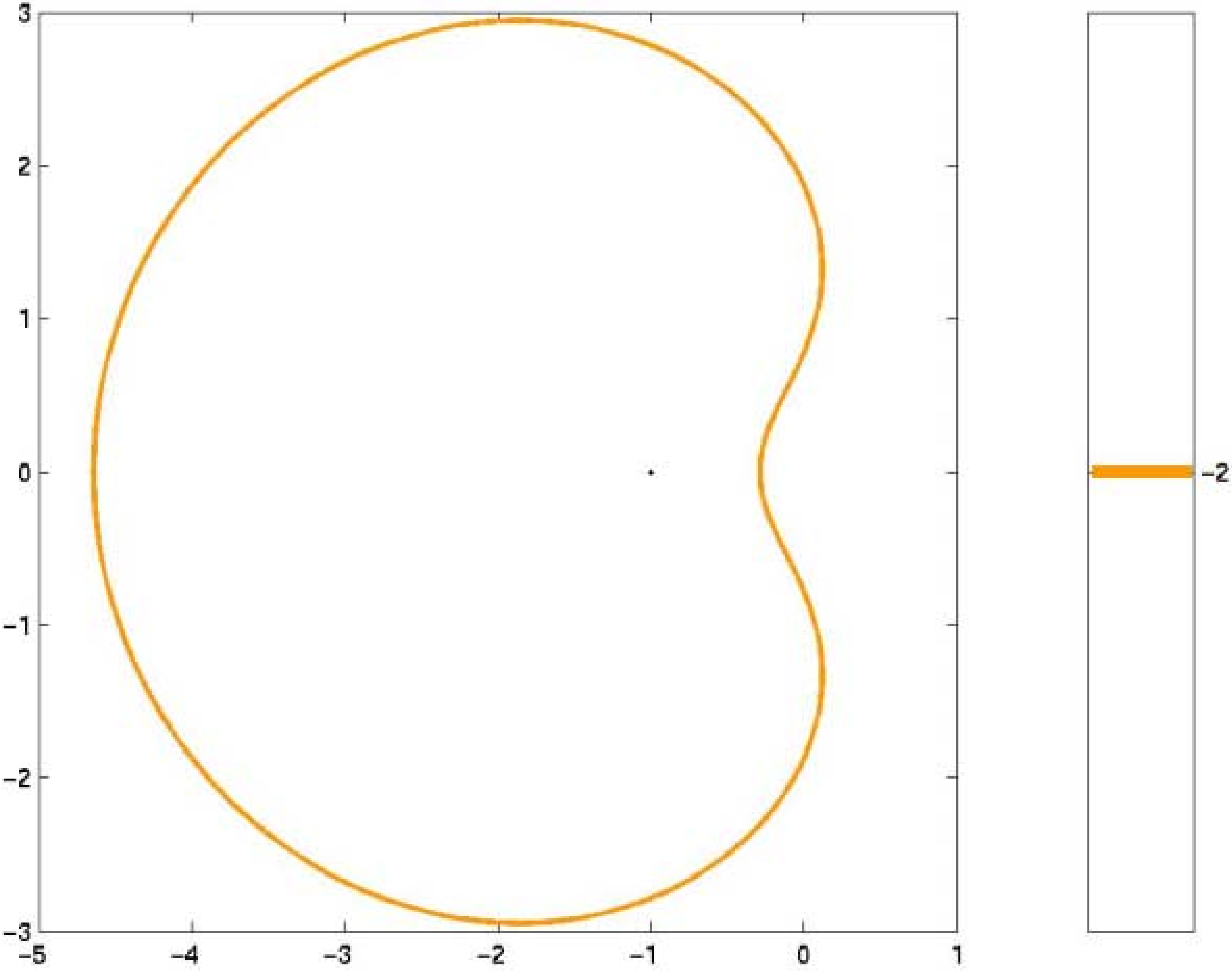
  has no imaginary eigenvalue
- There exist positive reals μ and λ and a positive definite Hermitian matrix P such that

$$\begin{matrix} AP + PA^* + (\lambda + \mu) I & \varepsilon P \\ \varepsilon P & -\lambda I \end{matrix}$$

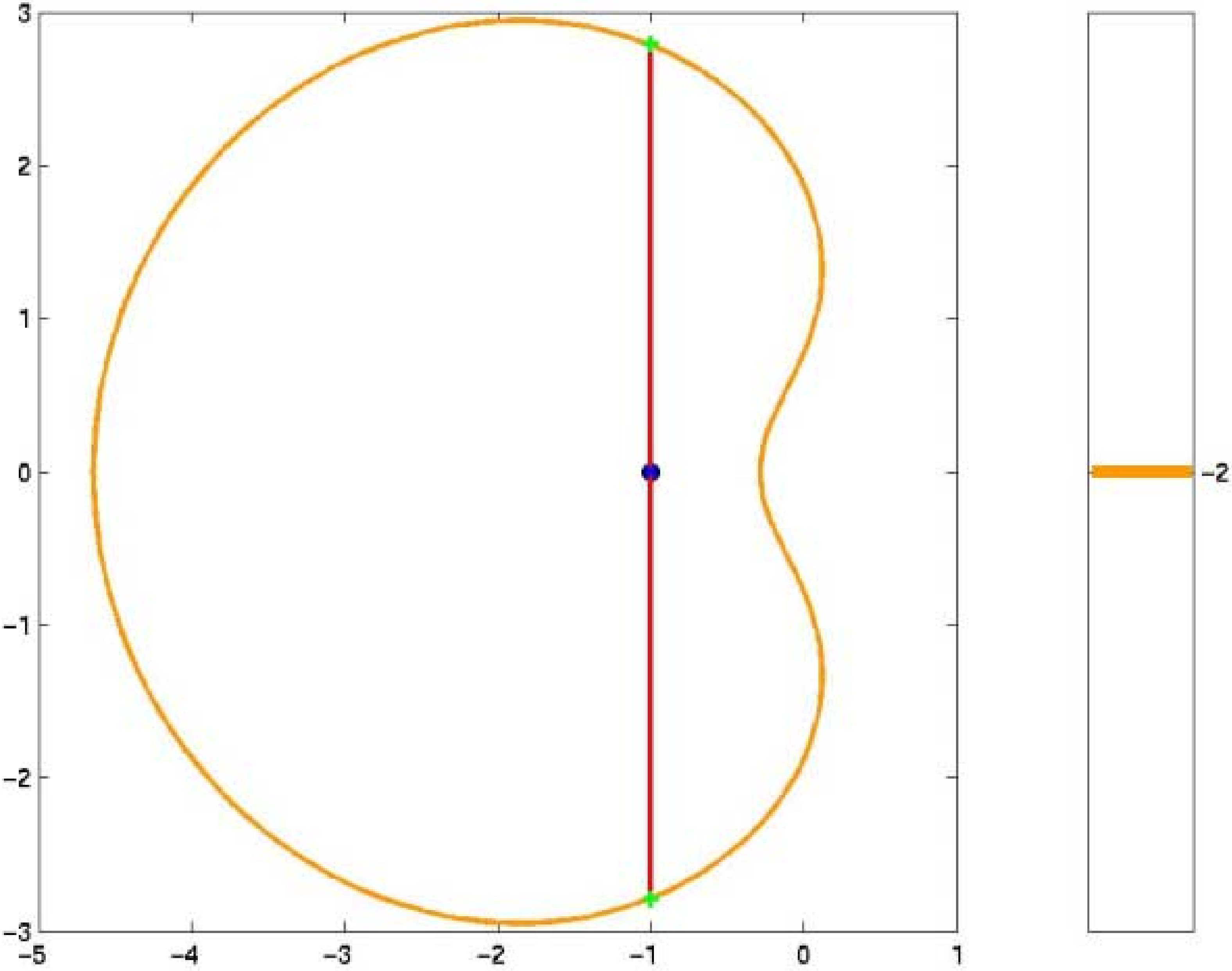
  is negative semidefinite (an LMI for fixed A)

Computing $\alpha_\varepsilon(A)$ or $\delta(A)$

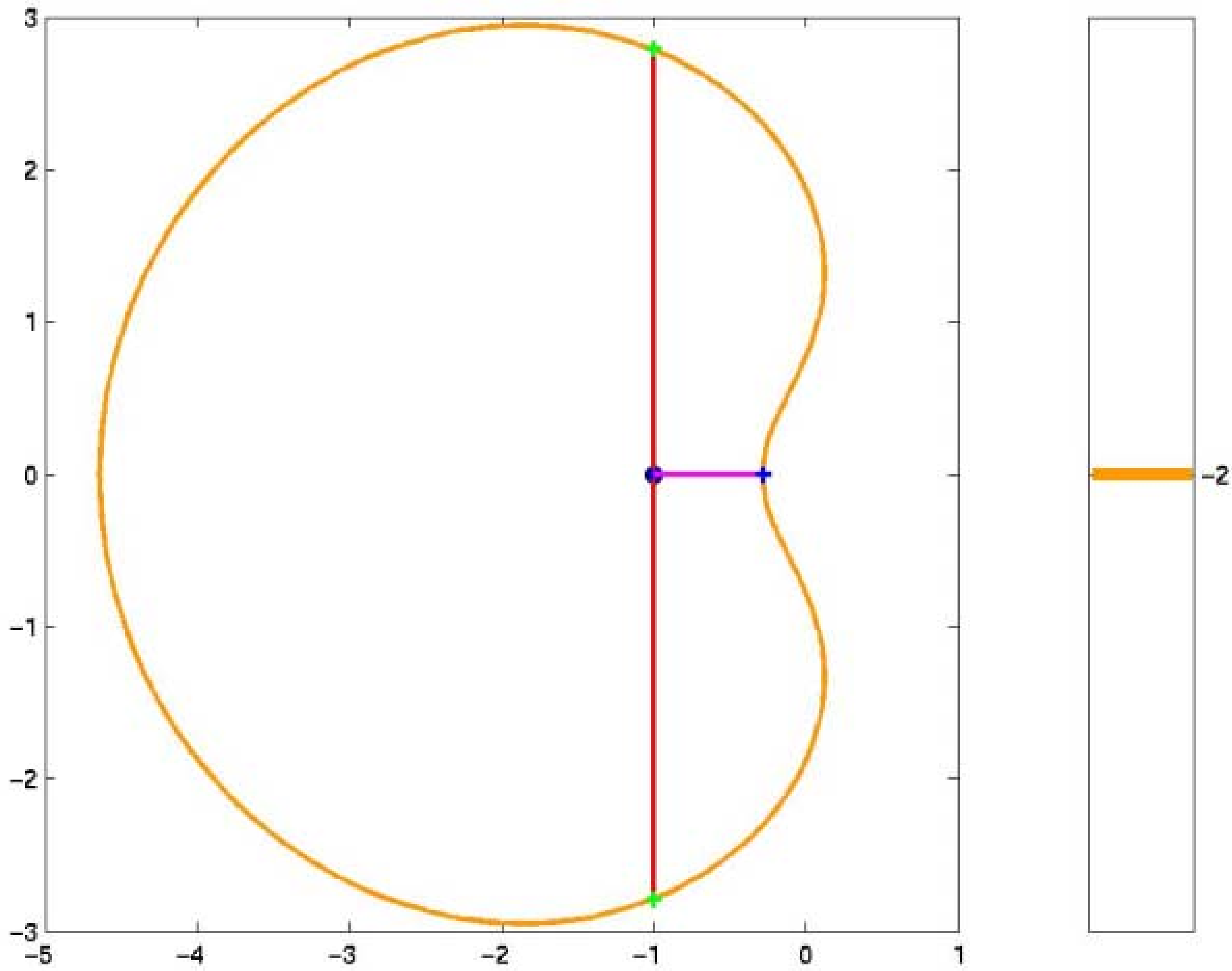
- Computing $\delta(A)$ (distance to instability, or H_∞ norm):
  - bisection algorithm: at each step, improve bound by checking if associated Hamiltonian matrix has any imaginary eigenvalues (Byers, Hinrichsen-Motscha-Linnemann)
  - quadratically convergent improvements (Boyd-Balakrishnan, van Dooren et al, etc)
- Computing $\alpha_\varepsilon(A)$ (pseudo-spectral abscissa)
  - bisection algorithm: essentially the same
  - quadratically convergent improvement: a bit different, requires Hamiltonian eigenvalue computation for horizontal as well as vertical "sweeps" in complex plane, and proof of convergence is tricky
  - Demmel matrix example….

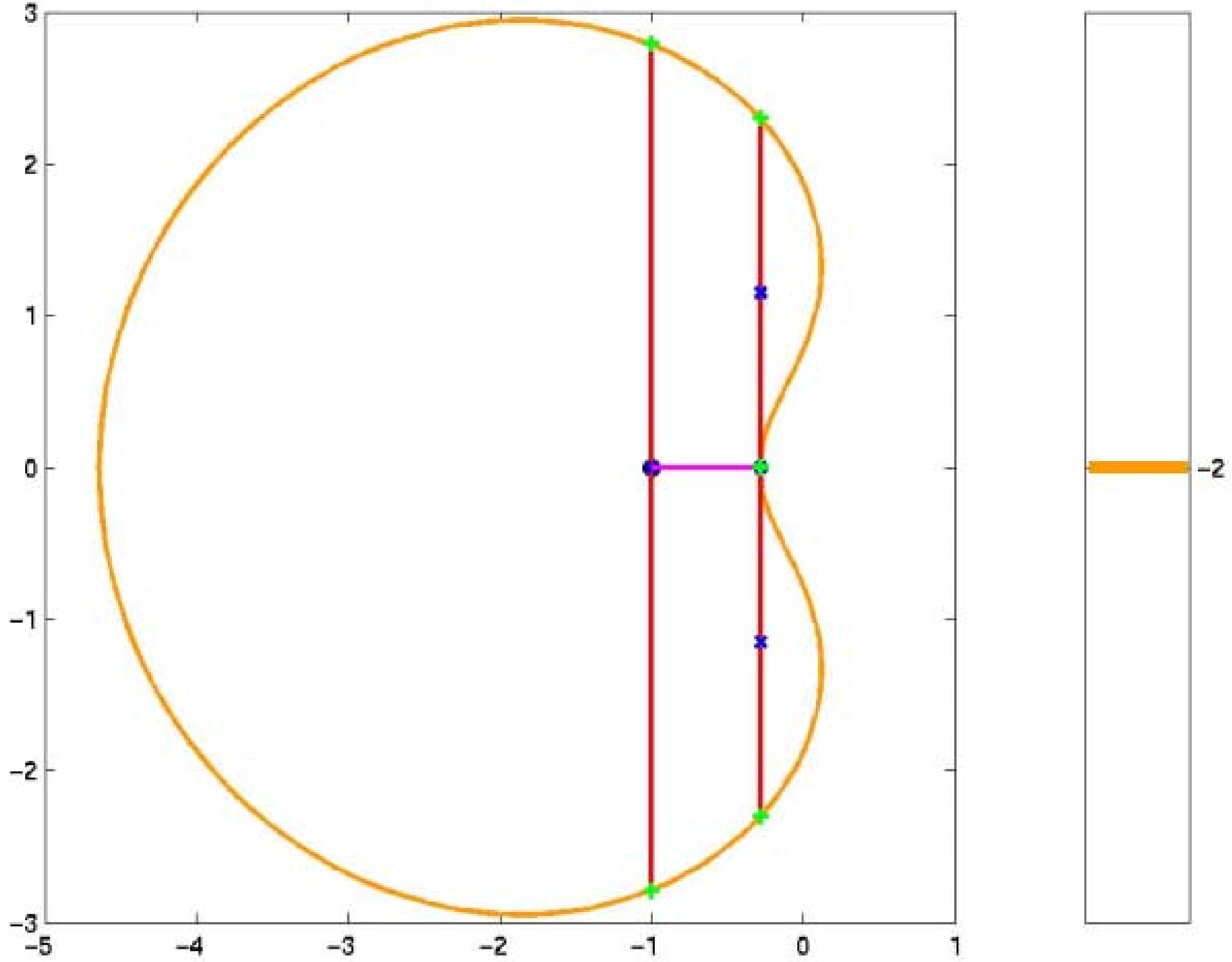


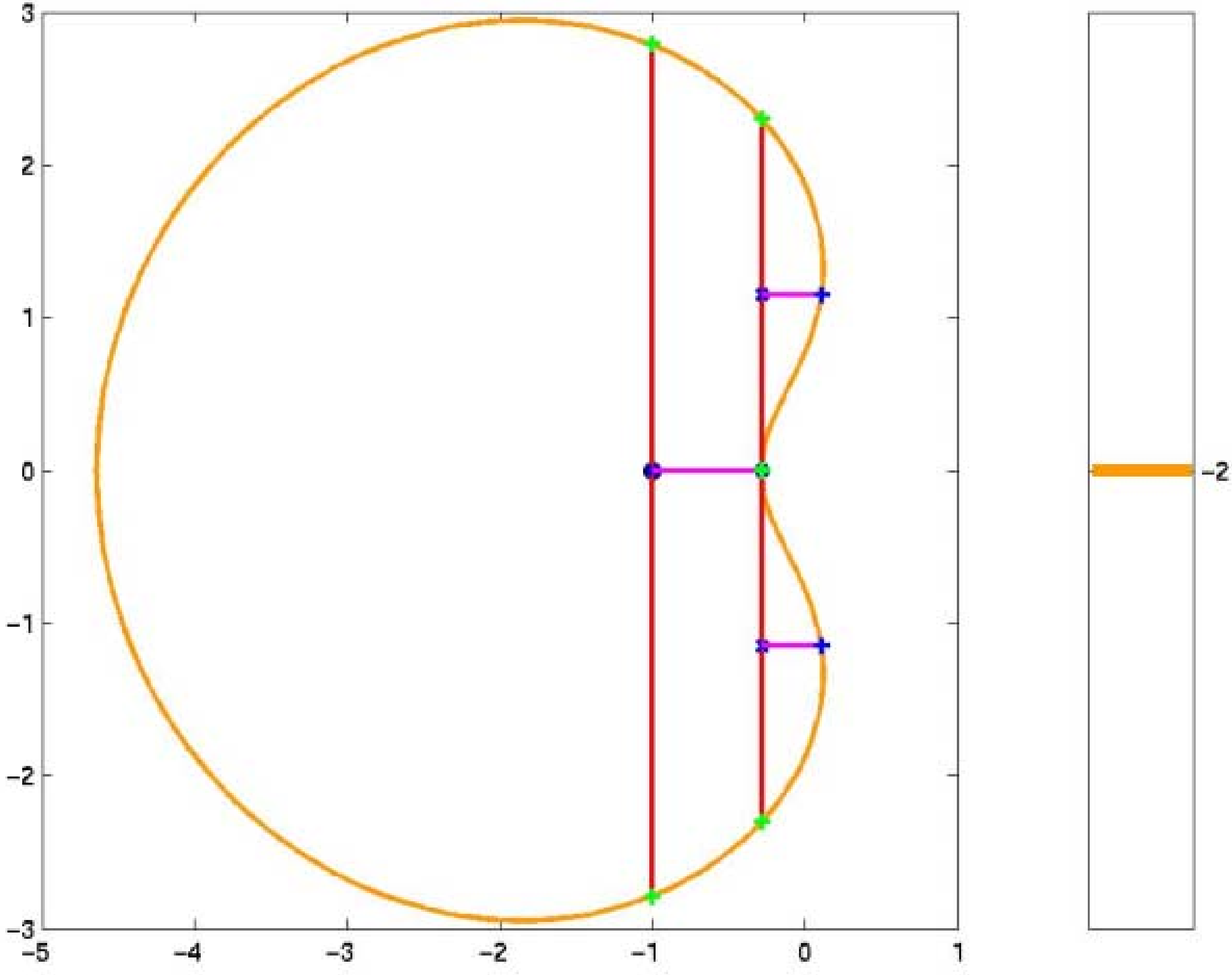
3
2
1
0
−1
−2
−3
−5
−4
−3
−2
−1
0
1
−2



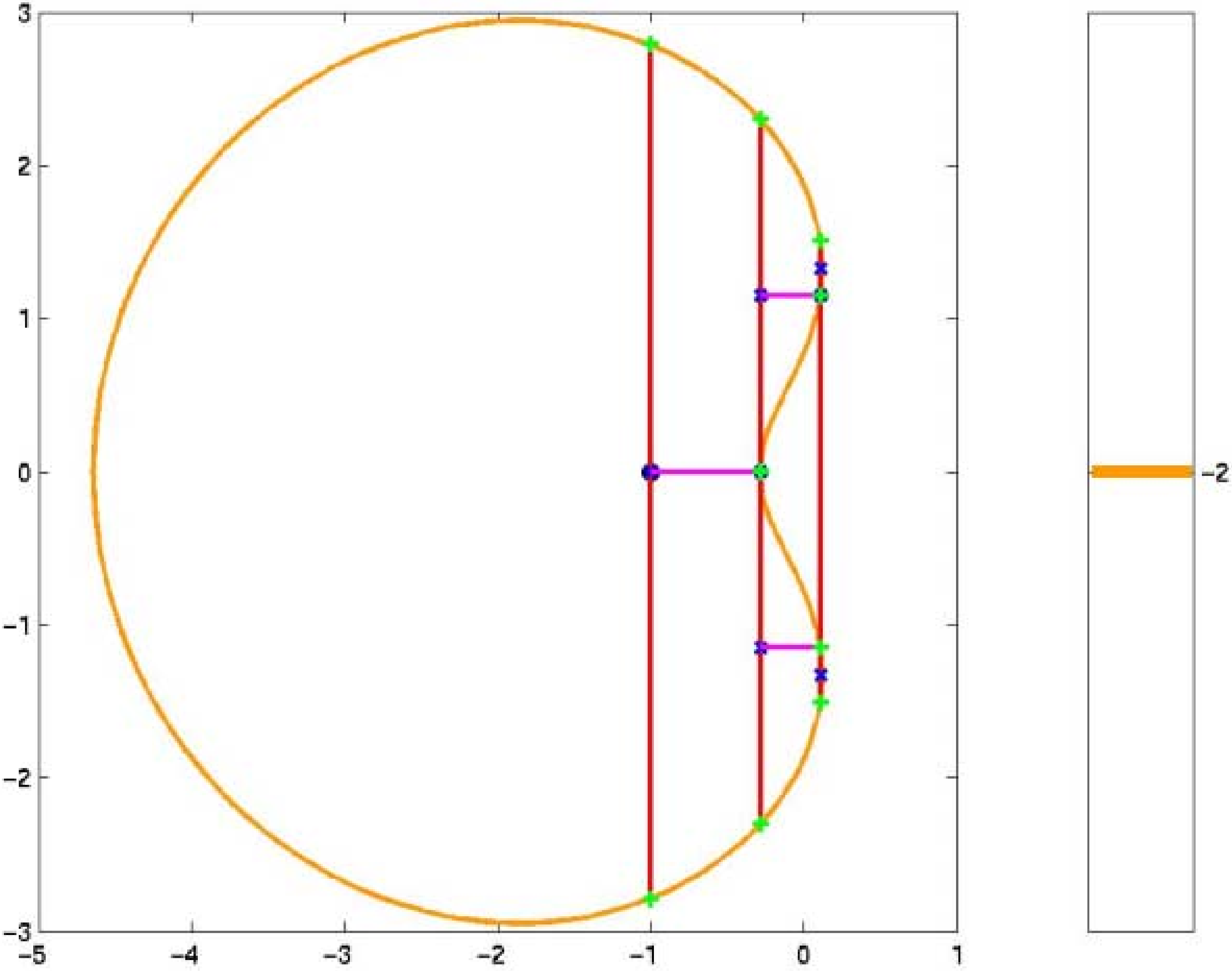
3
2
1
0
−1
−2
−3
−5
−4
−3
−2
−1
0
1
−2


3
2
1
0
−1
−2
−3
−5
−4
−3
−2
−1
0
1
−2


3
2
1
0
−1
−2
−3
−5
−4
−3
−2
−1
0
1
−2



3
2
1
0
−1
−2
−3
−5
−4
−3
−2
−1
0
1
−2



3
2
1
0
−1
−2
−3
−5
−4
−3
−2
−1
0
1
−2

Eigenvalues of Hamiltonian Matrices

- Are symmetric wrt imaginary axis (as well as real axis if matrix is real)
- If use algorithm that preserves Hamiltonian structure, no tolerance needed when testing eigenvalues for real part equals 0
- We use Benner's implentation of Van Loan's "square reduced" algorithm

Variational Properties

- Pure spectral functions, such as the spectral abscissa and spectral radius, are
  - not smooth
  - not convex
  - not Lipschitz
- Pseudo-spectral functions are
  - not smooth
  - not convex
  - Lipschitz

Nonsmooth Analysis

- Clarke (1973…), Mordukhovich (1976…), Ioffe (1981…) … Rockafellar and Wets (1998)
- Clarke regularity of a set S at a point x: implies (among other things) that S is locally closed at x and x isn't an "inward corner"
- The epigraph of a real valued function f on R^n (epi f): the subset of R^{n+1} lying on or above the graph of the function
- Subdifferential regularity: f is subdifferentially regular at x if epi f is Clarke regular at $(x, f(x))$
- Key point: regularity permits calculus (chain rule)

Nonsmooth Analysis of Spectral Functions

- Burke and Overton, *Math Programming*, 2001
- General results for subdifferential of $\phi \circ \Lambda$(a function on matrix space) in terms of subdifferential of ϕ(a function on C^n)
- Specific results for spectral abscissa α and spectral radius ρ
- Key result: the spectral abscissa α is subdifferentially regular at a matrix A iff all active eigenvalues of A (those whose real part equals $\alpha(A)$) are nonderogatory (have geometric multiplicity equal to 1)
- But it may not be Lipschitz (big Jordan blocks OK)

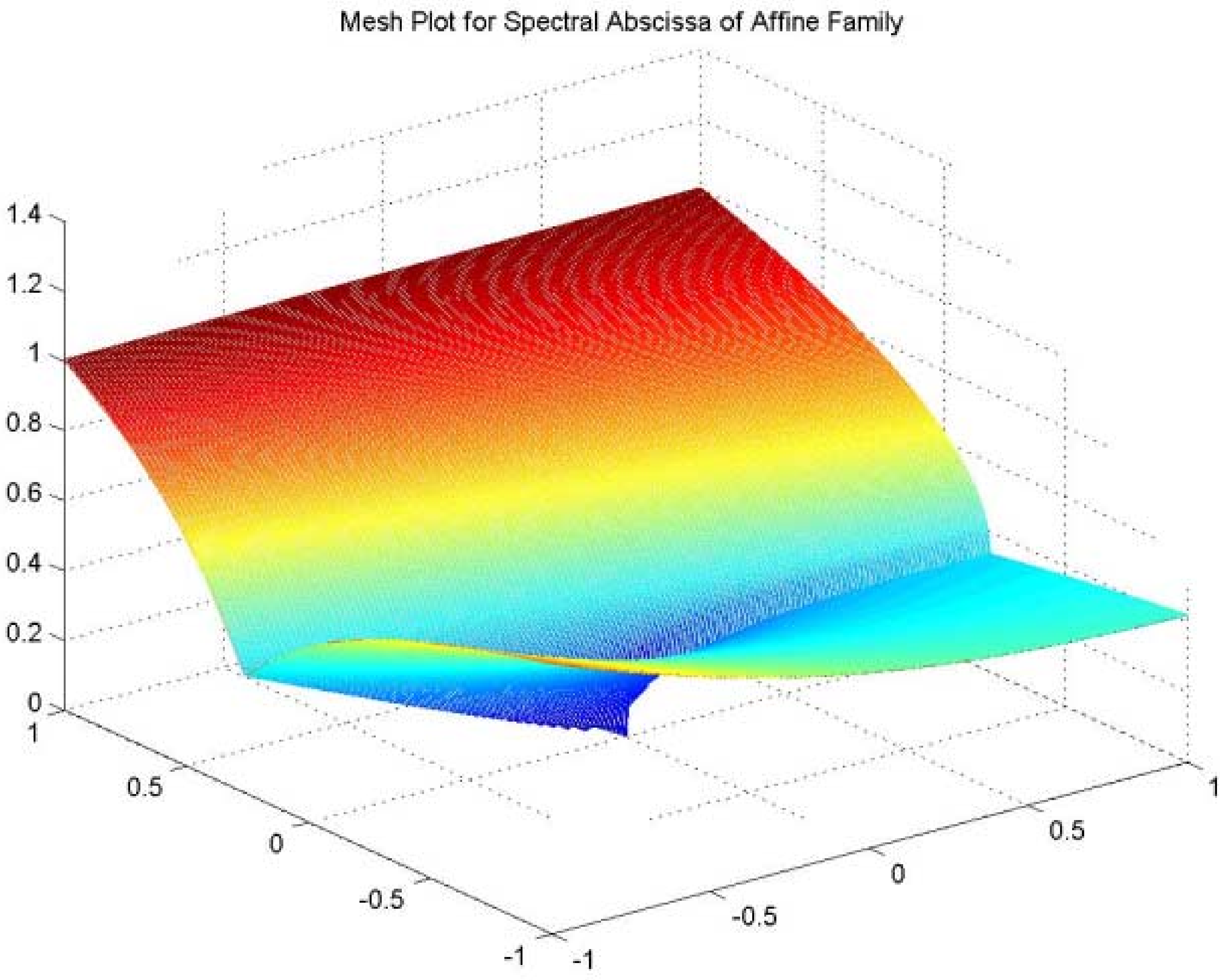
Nonsmooth Analysis of Pseudo-Spectral Functions

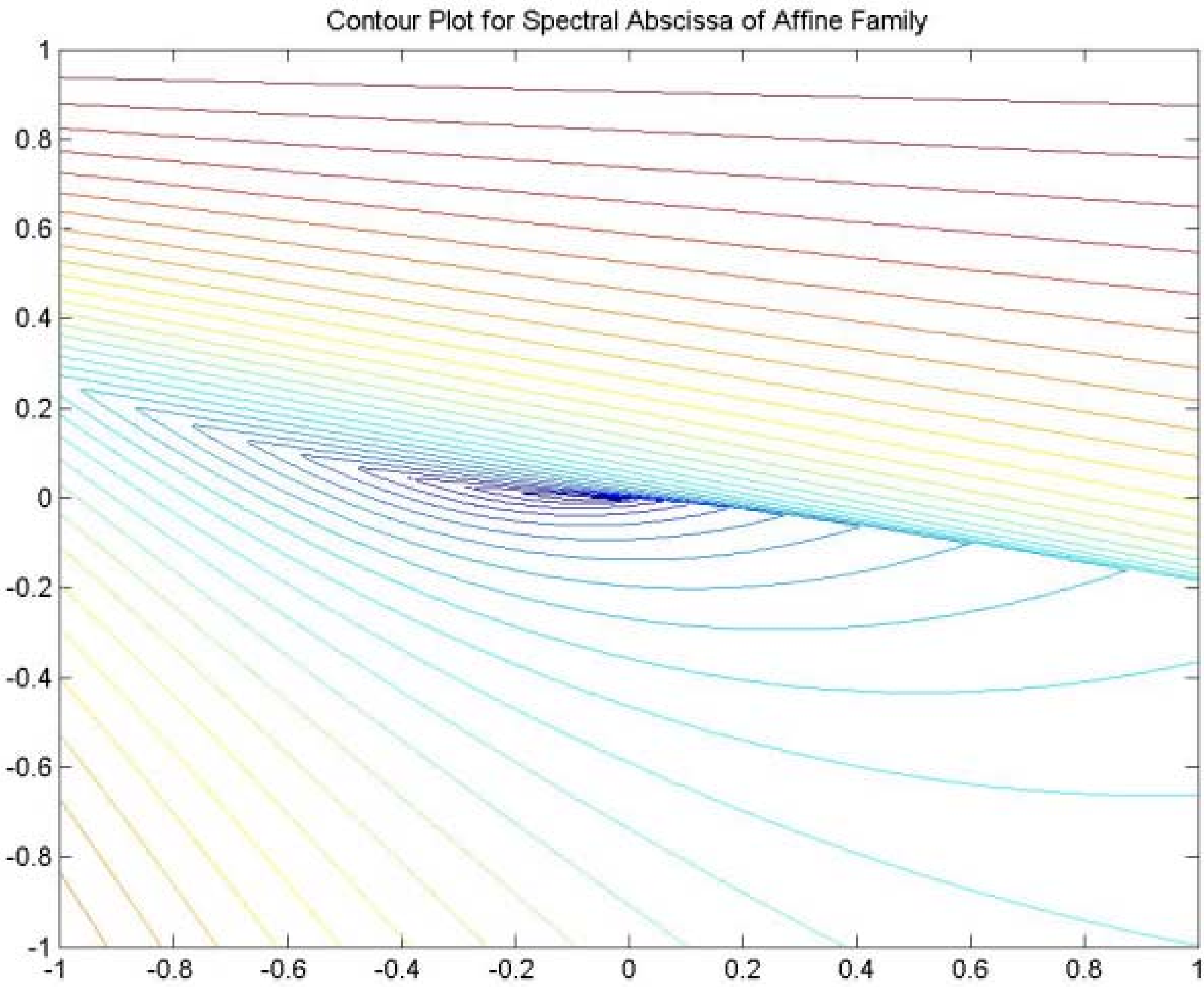
- Burke, Lewis, Overton, *SIMAX*, to appear
- Key result: at a matrix A whose active eigenvalues are nonderogatory, the pseudo-spectral abscissa α_ε is locally Lipschitz and subdifferentially regular for sufficiently small ε (in fact, it is locally the max of k smooth functions, where k is the number of active eigenvalues)

Optimization over Parameters

- Minimization of spectral abscissa over affine matrix family
- min $\alpha\,(A_0 + \Sigma\, x_k\, A_k)$
- example:

$$A_0 = \begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix} \qquad A_1 = \begin{matrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \qquad A_2 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{matrix}$$


Mesh Plot for Spectral Abscissa of Affine Family
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1



Contour Plot for Spectral Abscissa of Affine Family
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

Multiple Eigenvalues

- Typically, spectral abscissa minimizers are associated with eigenvalues with algebraic multiplicity > 1
- But with geometric multiplicity = 1 (with associated Jordan blocks)
- Such matrices are very sensitive to perturbation so even if $\alpha << 0$, distance to instability could be small (large H_∞ norm)
- There could be many different “active” multiple eigenvalues, all having same real part

Stabilization by Static Output Feedback

$$z'(t) = A_0\, z(t) + B_0\, u(t)$$

$y(t) = C_0\, z(t)$ measures state

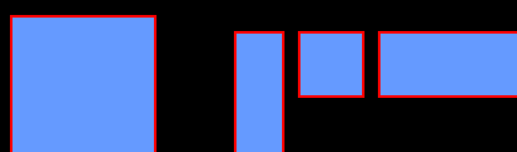
$u(t) = X\, y(t)$ control

Choose X so that solutions of

$$z'(t) = (A_0 + B_0\, X\, C_0)\, z(t)$$

are stable, i.e.

$$\alpha\,(A_0 + B_0\, X\, C_0) < 0$$



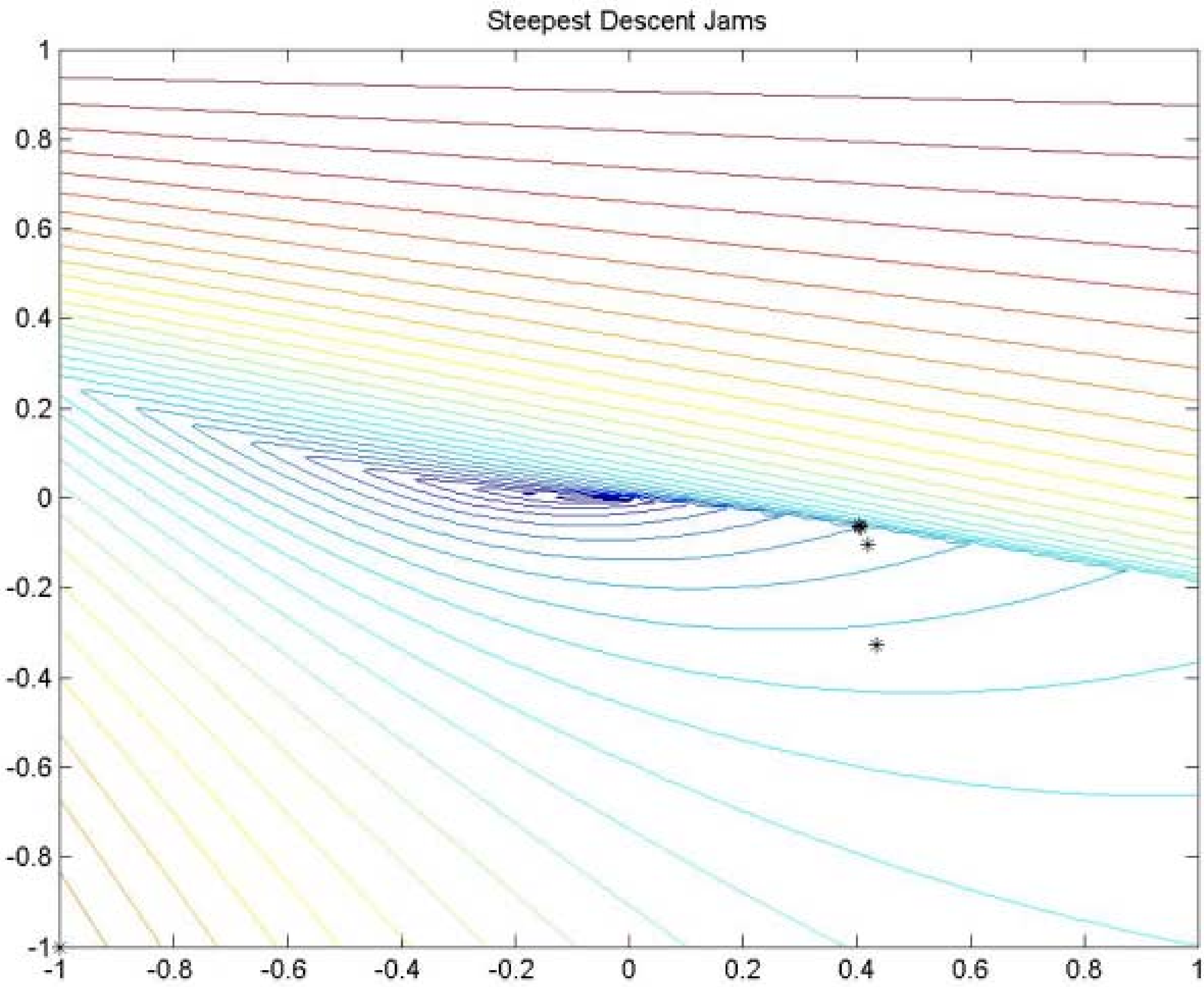
or better: "optimally stable".

What Should We Optimize?

- Spectral abscissa α:
  - cheap to compute
  - ideal asympotically
  - bad for transient behavior and robustness
- Pseudo-spectral abscissa α_ε:
  - good if we know what ε is tolerable
  - can balance asymptotic and transient considerations
- Distance to instability δ (equivalently, H_∞ norm):
  - good if want to tolerate biggest ε possible
  - bad if care about asymptotic rate
  - difficulty: feasible starting point often not available
  - solution: can be obtained by first minimizing α

Can we Optimize these Functions?

- Globally, no. Related problems are NP-hard (Blondell-Tsitsiklas, Nemirovski)
- Locally, yes
  - But not by standard methods for nonconvex, smooth optimization
  - Steepest descent, BFGS or nonlinear conjugate gradient will typically jam because of nonsmoothness


Steepest Descent Jams
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

Methods for Nonsmooth, Nonconvex Optimization

- Long history, but most methods are very complicated
- Typically they generalize bundle methods for nonsmooth, convex optimization (e.g. Kiwiel)
- Ad hoc methods, e.g. Nelder-Mead, are ineffective on nonsmooth functions with more than a few parameters, and local optimality cannot be verified
- We use a novel Gradient Sampling algorithm, requiring (in practice) only that
  - f is continuous
  - f is continuously differentiable almost everywhere
  - where defined, gradient of f is easily computed

Computing the Gradients

- Gradient of spectral abscissa α : when only one eigenvalue is active and it is simple, gradient of α in matrix space is: uv^* where u is left eigenvector and v is right eigenvector, with $u^*v = 1$
- Gradient of α_ε and δ : involves left and right singular vectors instead
- Chain rule gives gradient in parameter space

Gradient Sampling Algorithm:

Initialize η and x.

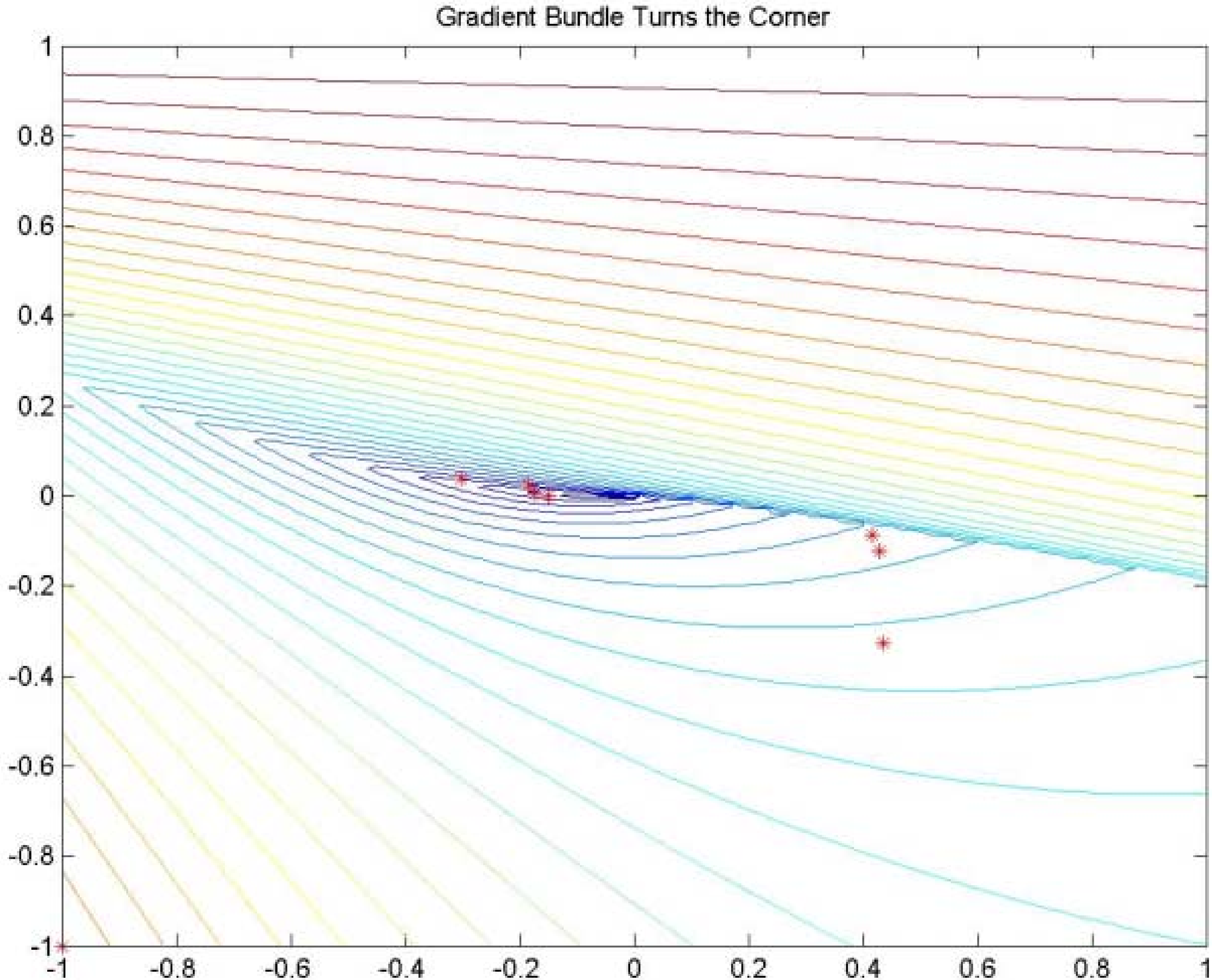
Repeat

- Get G, a set of gradients of function f evaluated at x and at points near x (sampling controlled by η)
- Let $d = \arg\min \{ \|d\|: d \in \mathrm{conv}\, G \}$
- Replace x by $x - t\, d$, such that

$$f(x - t\, d) < f(x) \quad \text{(if } d \text{ is not 0)}$$

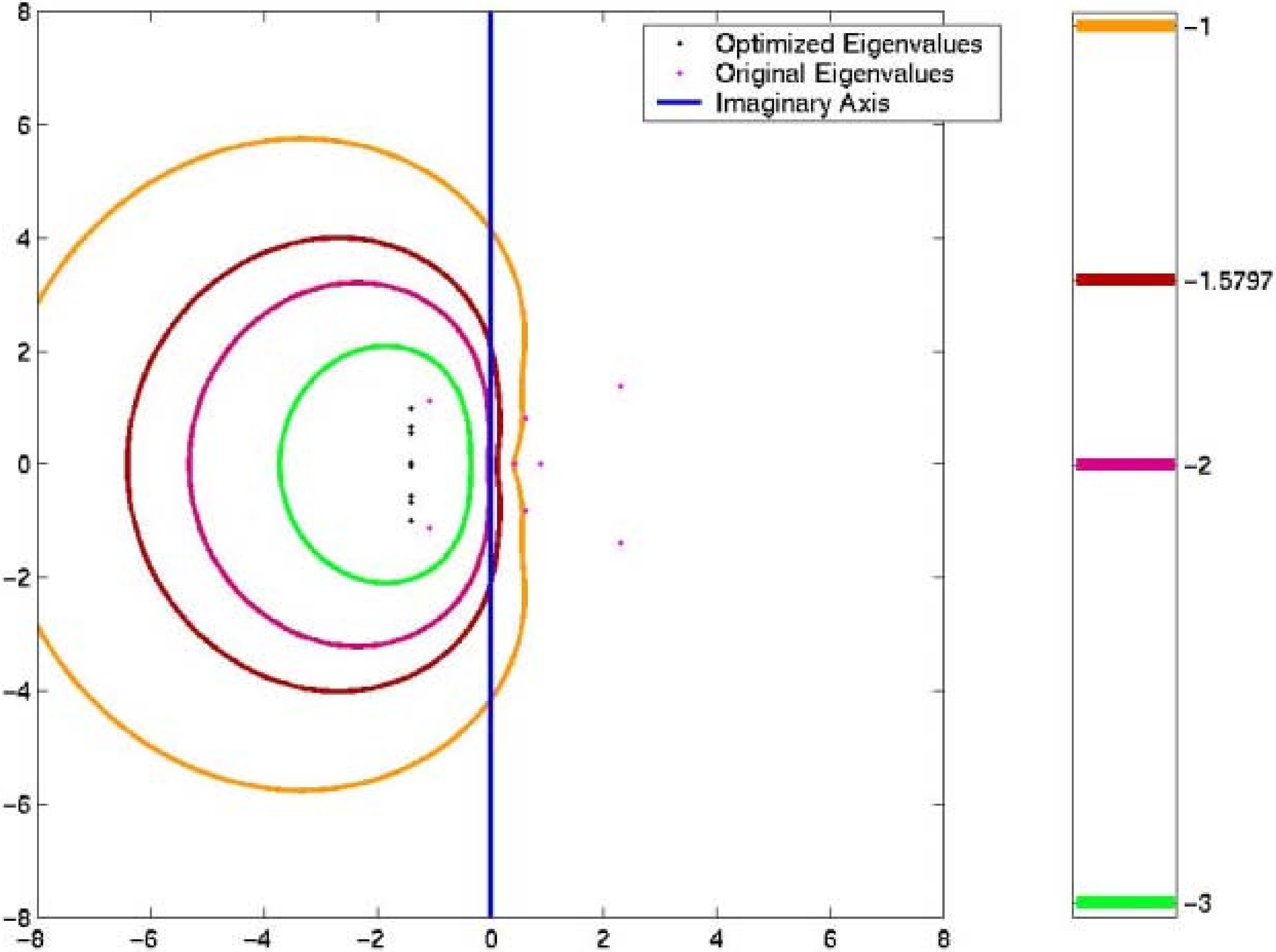
until $d = 0$.

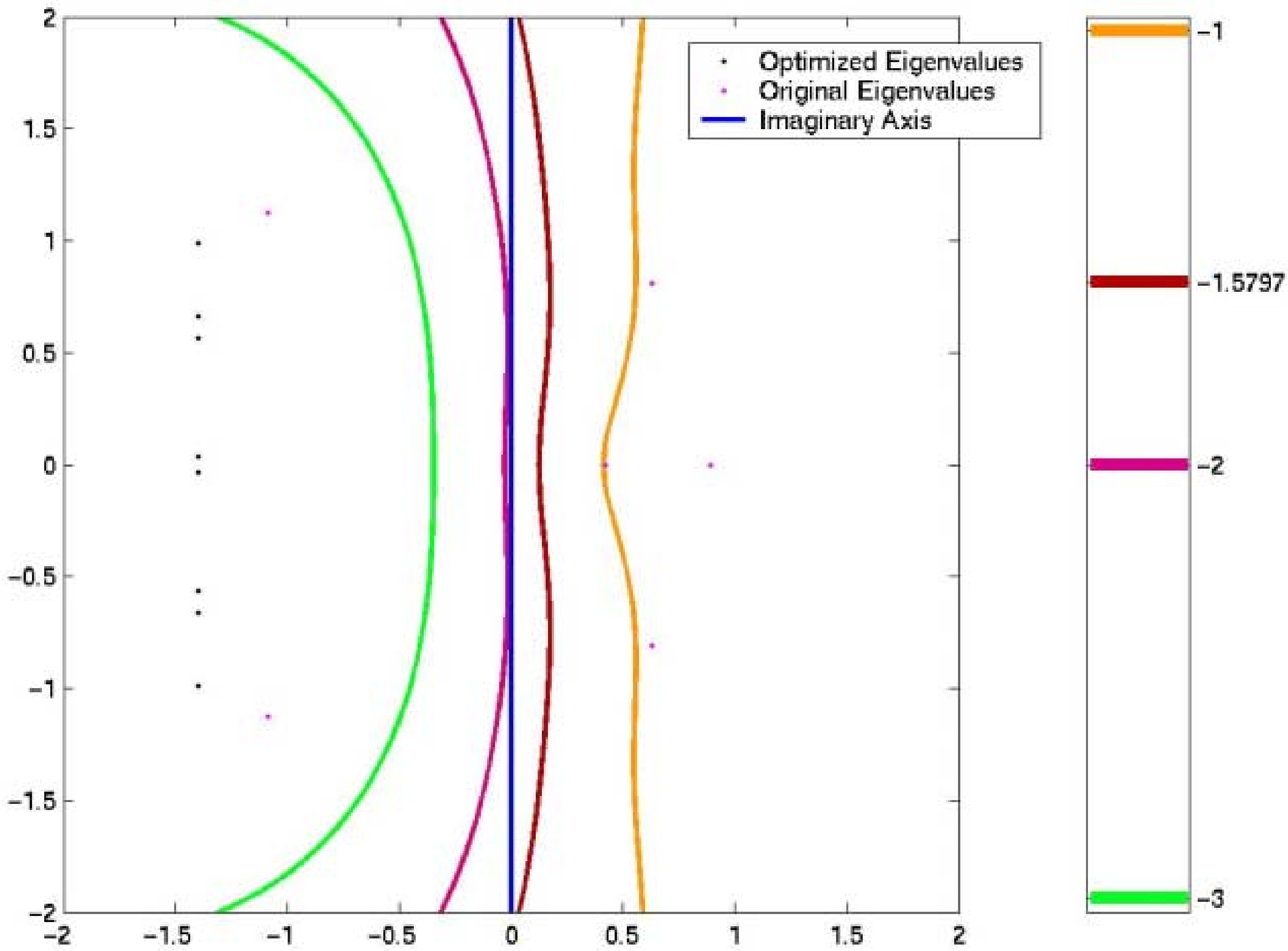
Then reduce η and repeat.


Gradient Bundle Turns the Corner
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

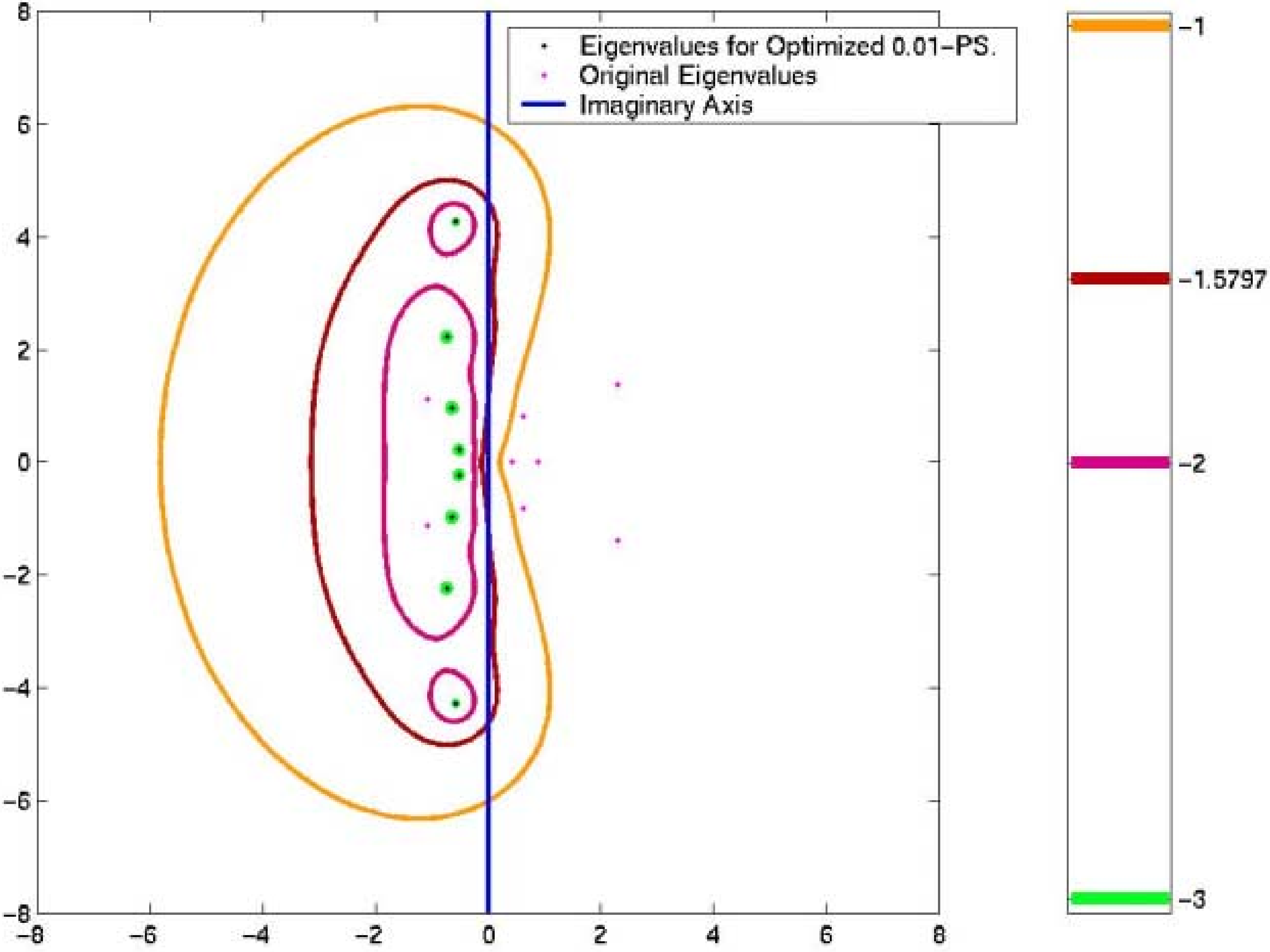
A Simple Static Output Feedback Example

- Wang (*Trans. Automatic Control*)
- Provided by F. Leibfritz ("Problem 39")
- Plots showing spectra and pseudo-spectra of the locally optimal solutions we found, minimizing
  - spectral abscissa α
  - pseudo-spectral abscissa α_ε
  - H_∞ norm (maximizing distance to instability δ) (use spectral abscissa minimizer to initialize)

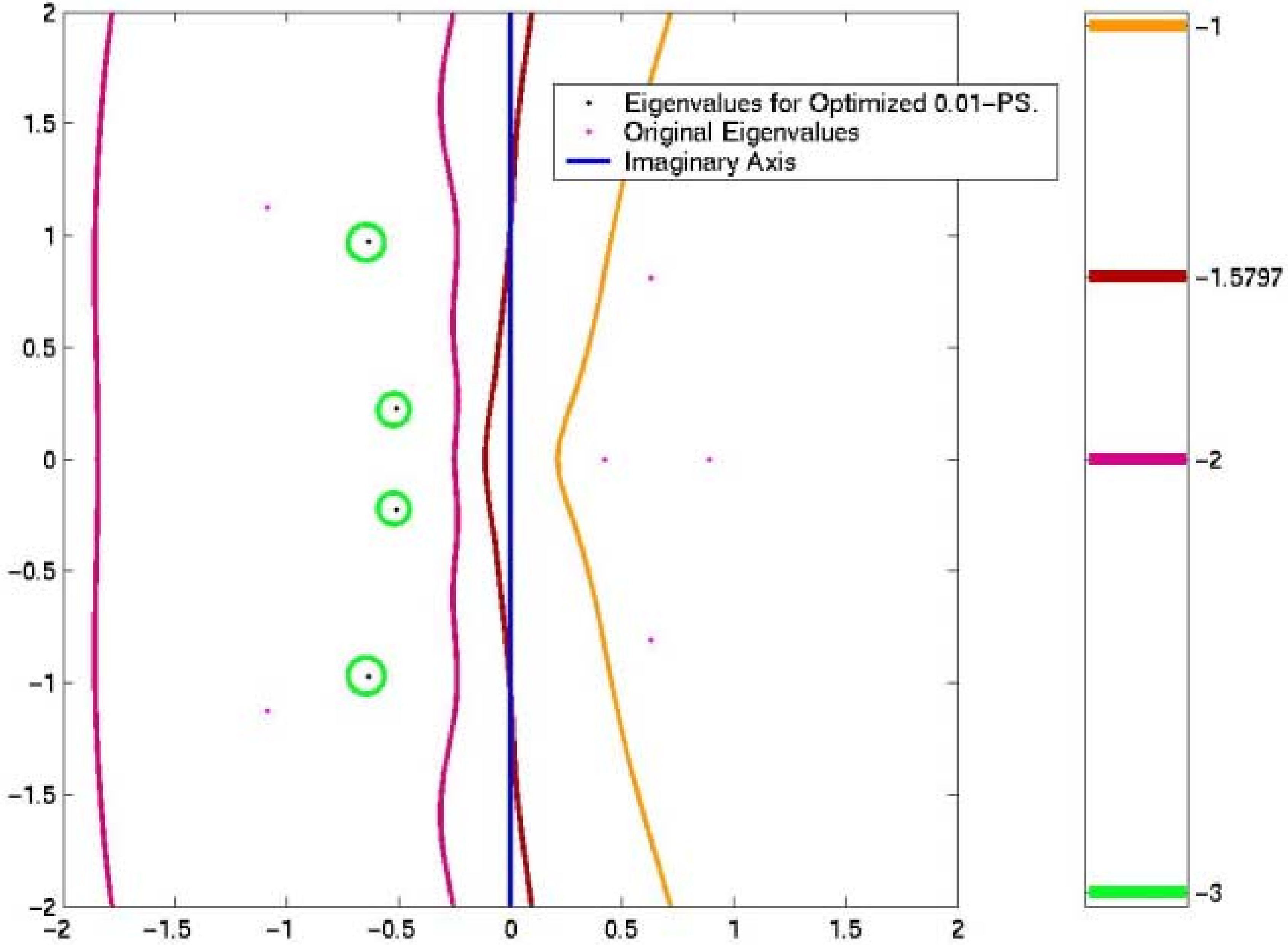

Optimized Eigenvalues
Original Eigenvalues
Imaginary Axis
8
6
4
2
0
−2
−4
−6
−8
−8
−6
−4
−2
0
2
4
6
8
−1
−1.5797
−2
−3

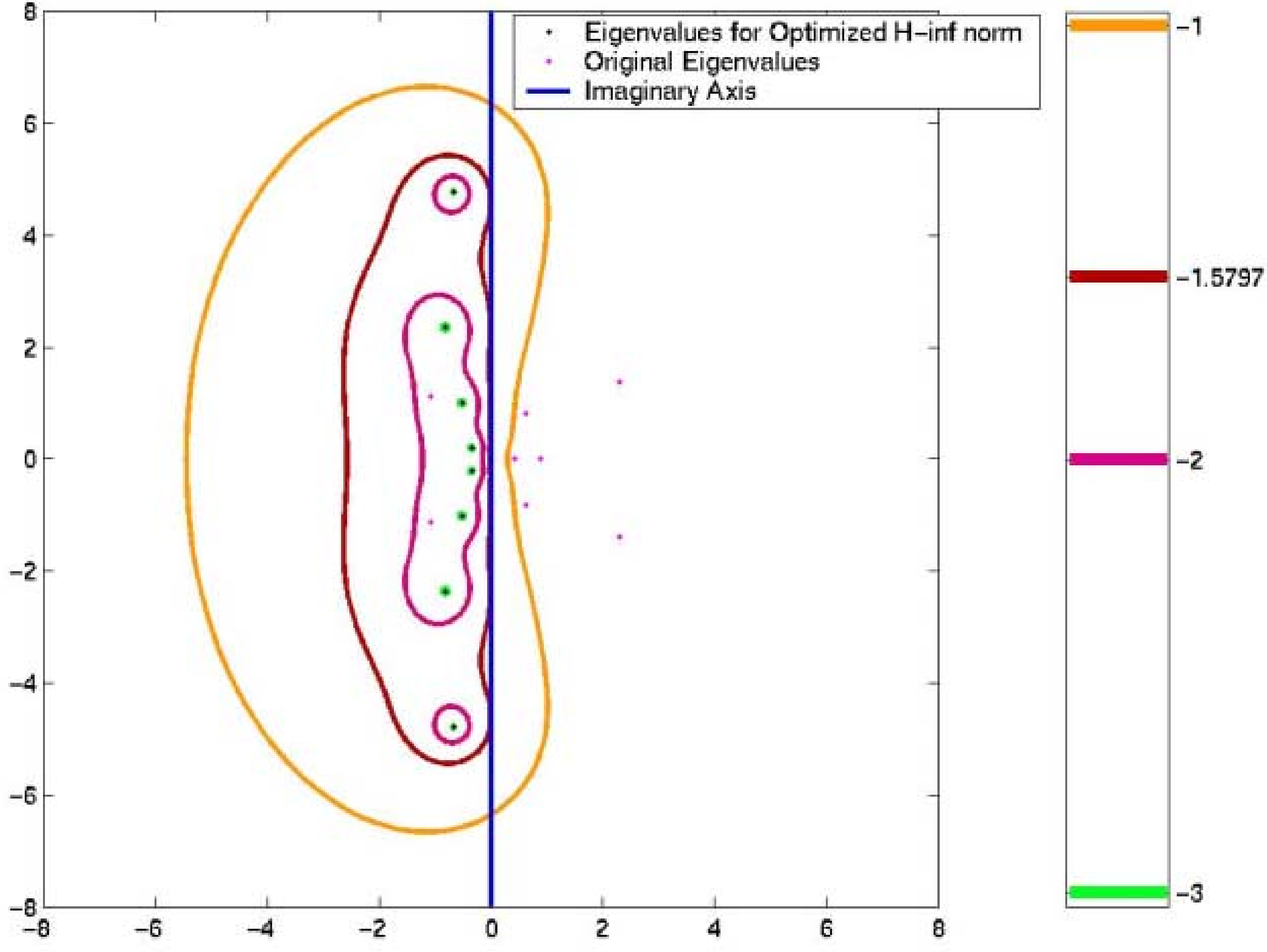


Optimized Eigenvalues
Original Eigenvalues
Imaginary Axis
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
−1
−1.5797
−2
−3

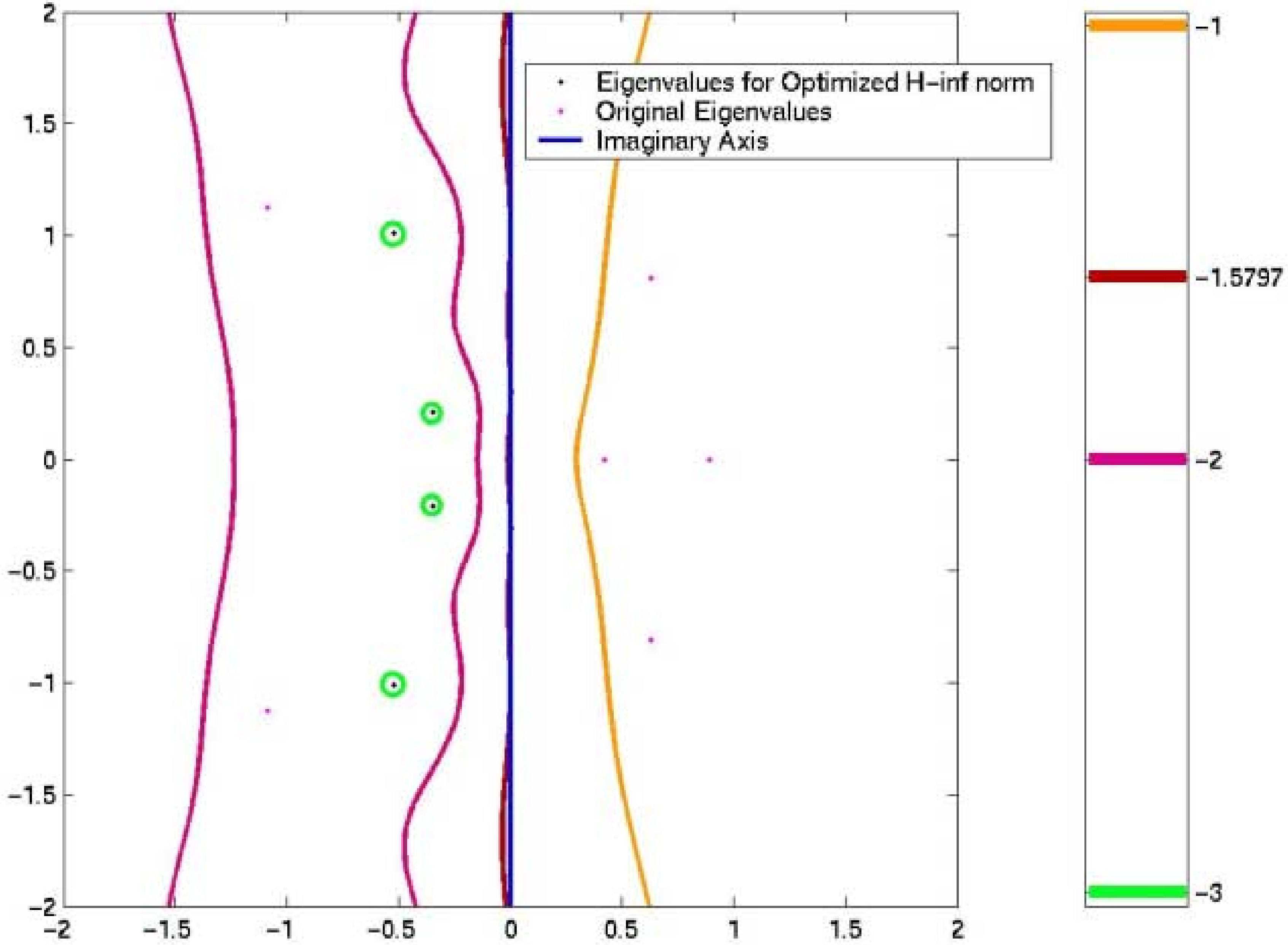


Eigenvalues for Optimized 0.01−PS.
Original Eigenvalues
Imaginary Axis
8
6
4
2
0
−2
−4
−6
−8
−8
−6
−4
−2
0
2
4
6
8
−1
−1.5797
−2
−3


Eigenvalues for Optimized 0.01−PS.
Original Eigenvalues
Imaginary Axis
−1
−1.5797
−2
−3
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2

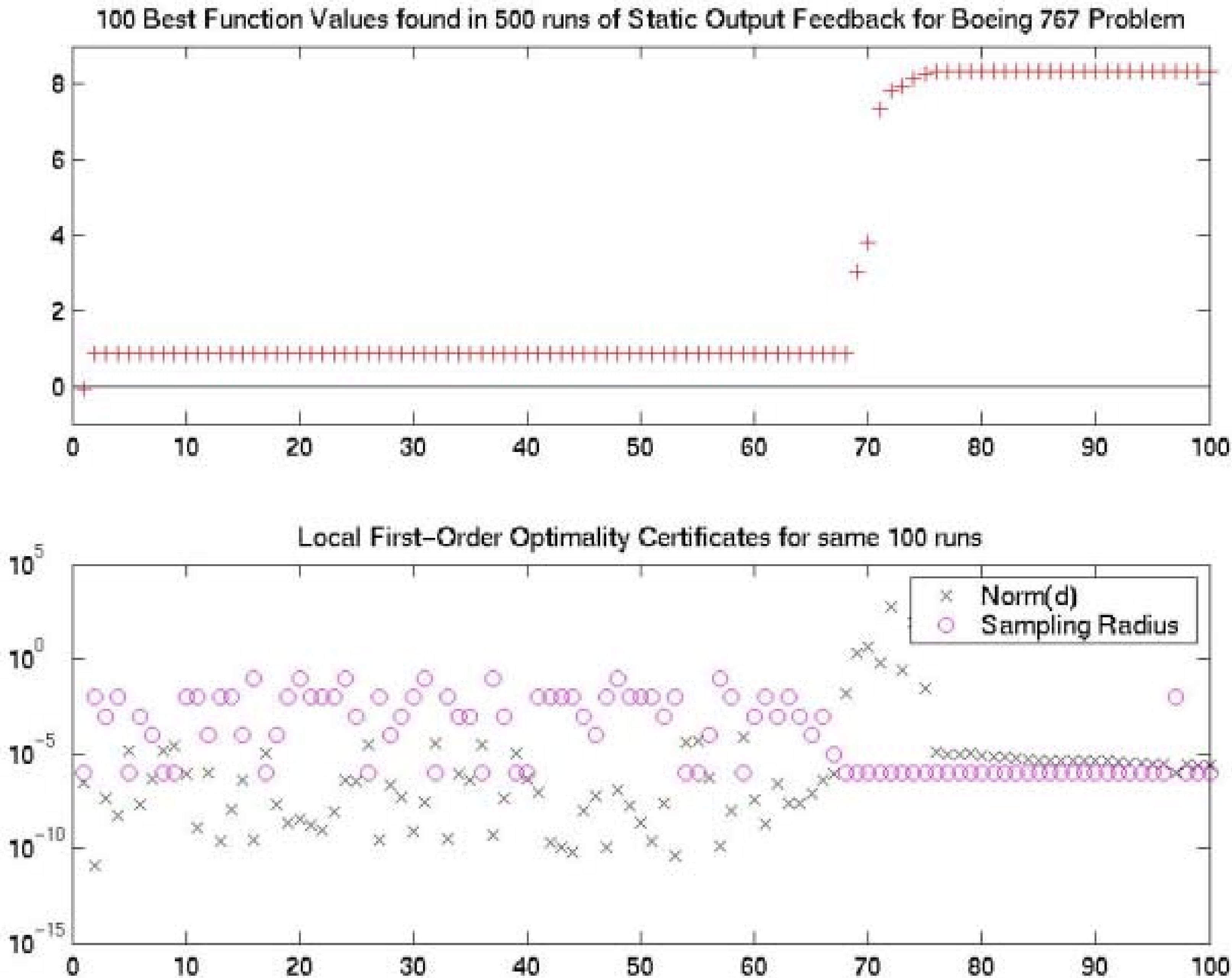


Eigenvalues for Optimized H−inf norm
Original Eigenvalues
Imaginary Axis
−1
−1.5797
−2
−3
8
6
4
2
0
−2
−4
−6
−8

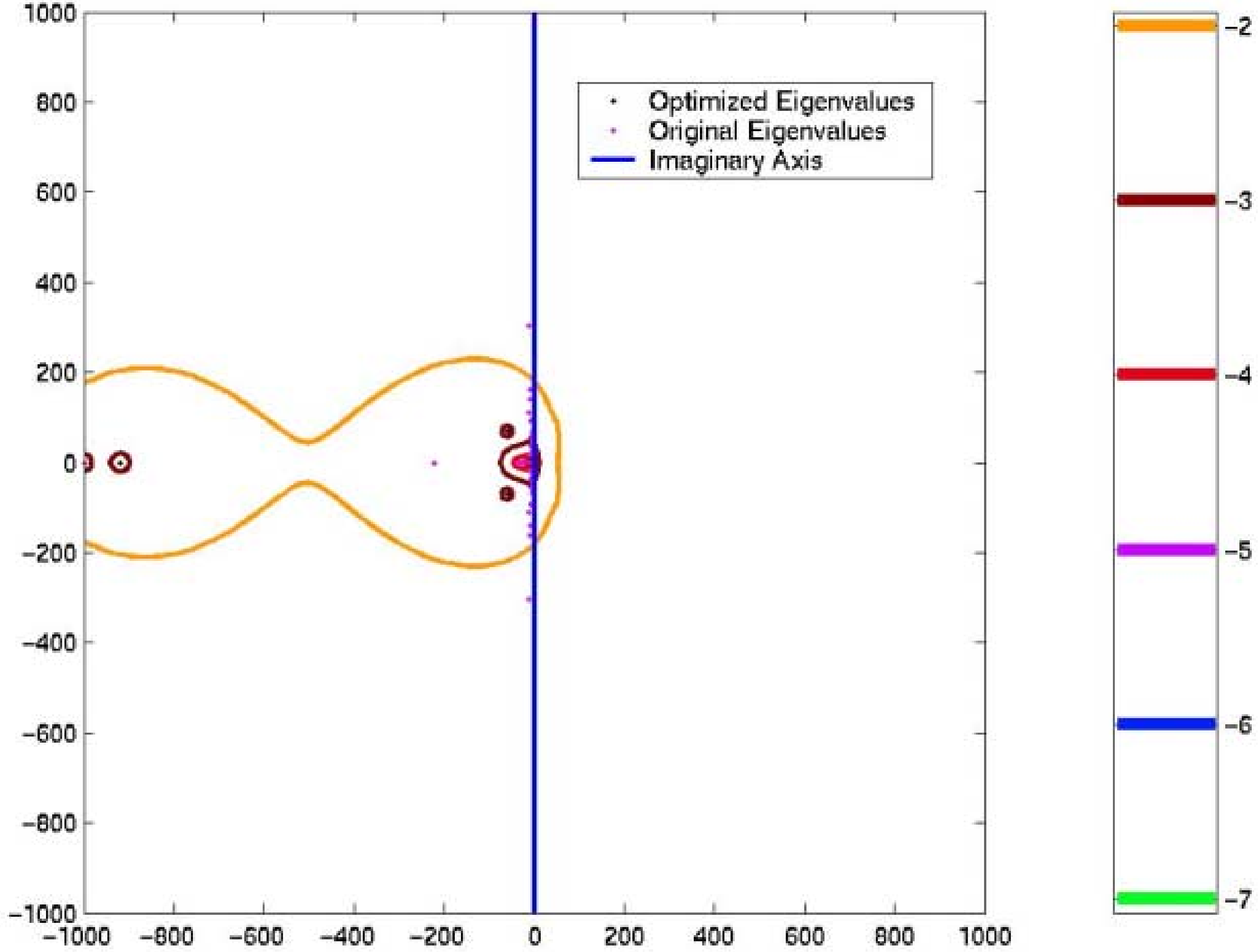

Eigenvalues for Optimized H−inf norm
Original Eigenvalues
Imaginary Axis
−1
−1.5797
−2
−3

The Boeing 767 Test Problem

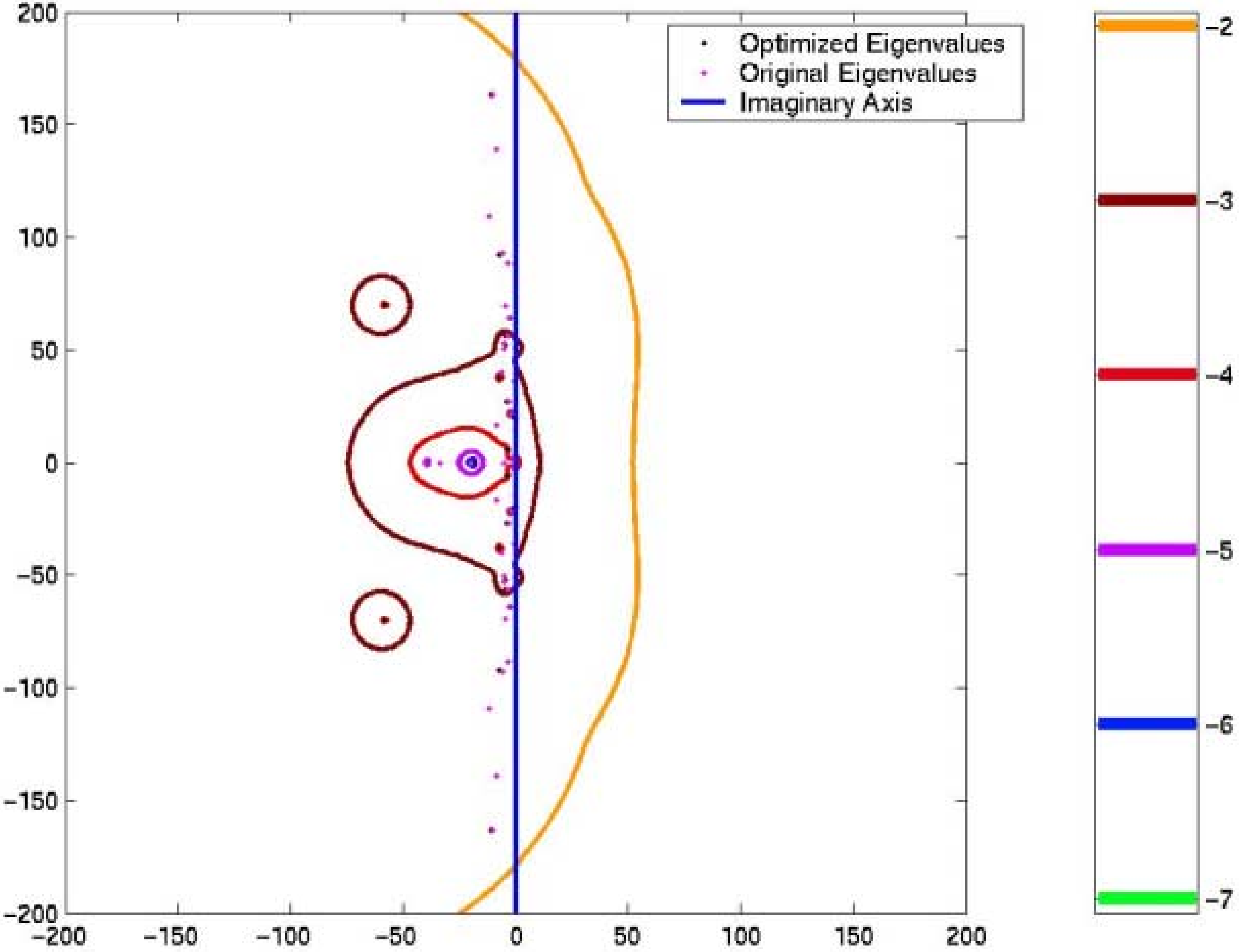
- Provided by F. Leibfritz (“Problem 37”), also on SLICOT web page
- Aeroelastic model of Boeing 767 at flutter condition
- Spectral abscissa minimization: min $\alpha(A_0 + B_0 X C_0)$
- A_0 is 55 by 55, X is 2 by 2
- Apparently no X making $\alpha(A_0 + B_0 X C_0) < 0$ was known

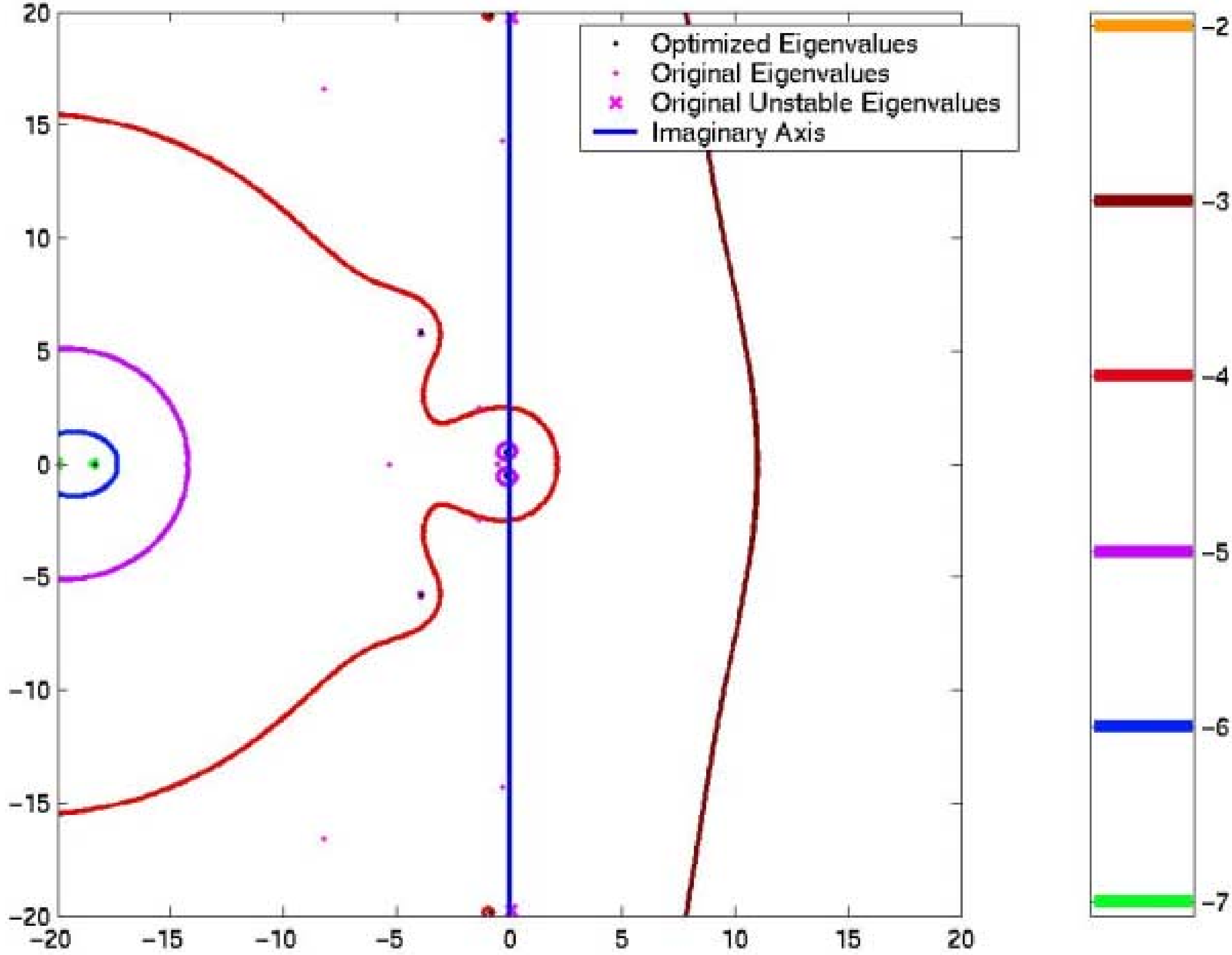


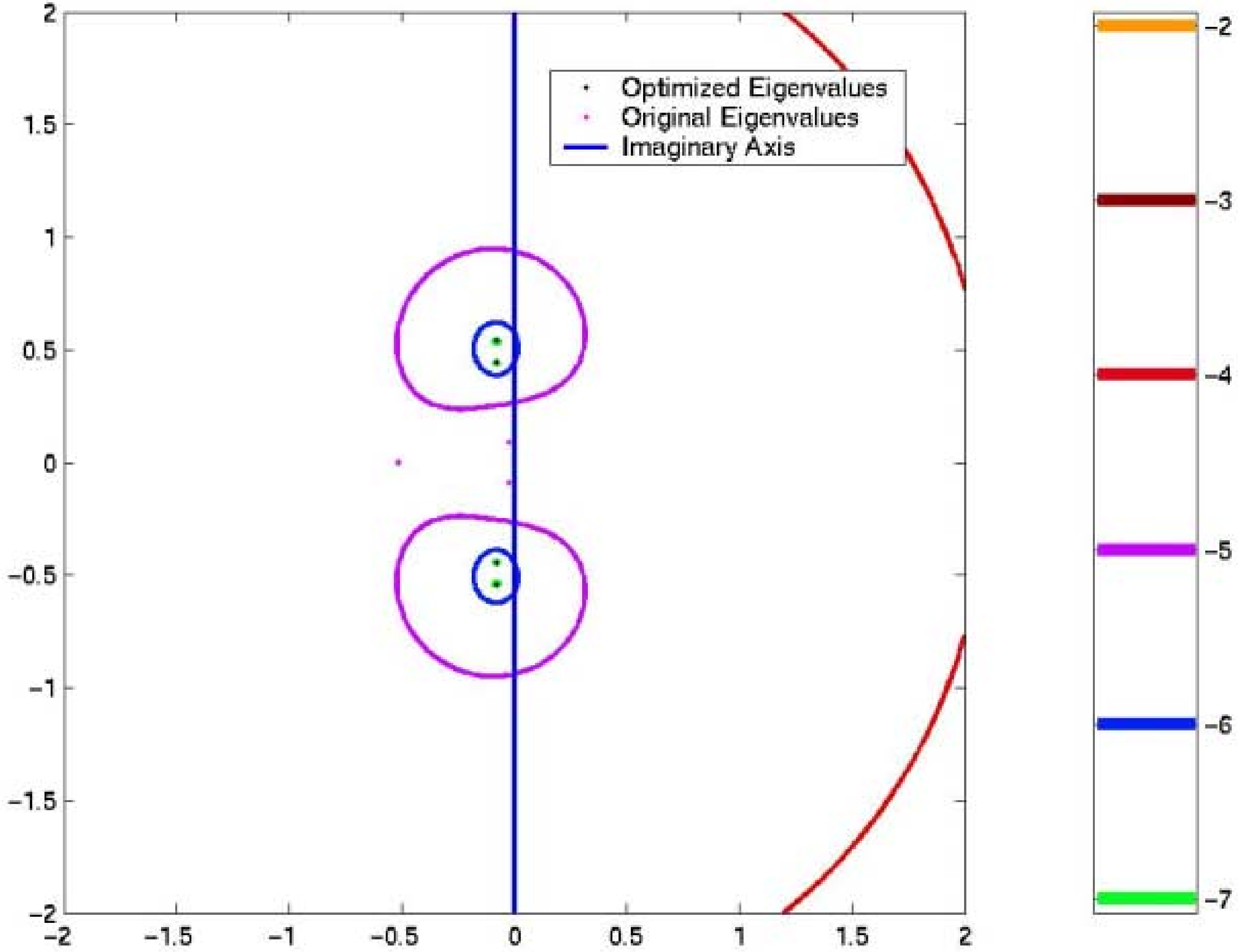
100 Best Function Values found in 500 runs of Static Output Feedback for Boeing 767 Problem
Local First–Order Optimality Certificates for same 100 runs
Norm(d)
Sampling Radius



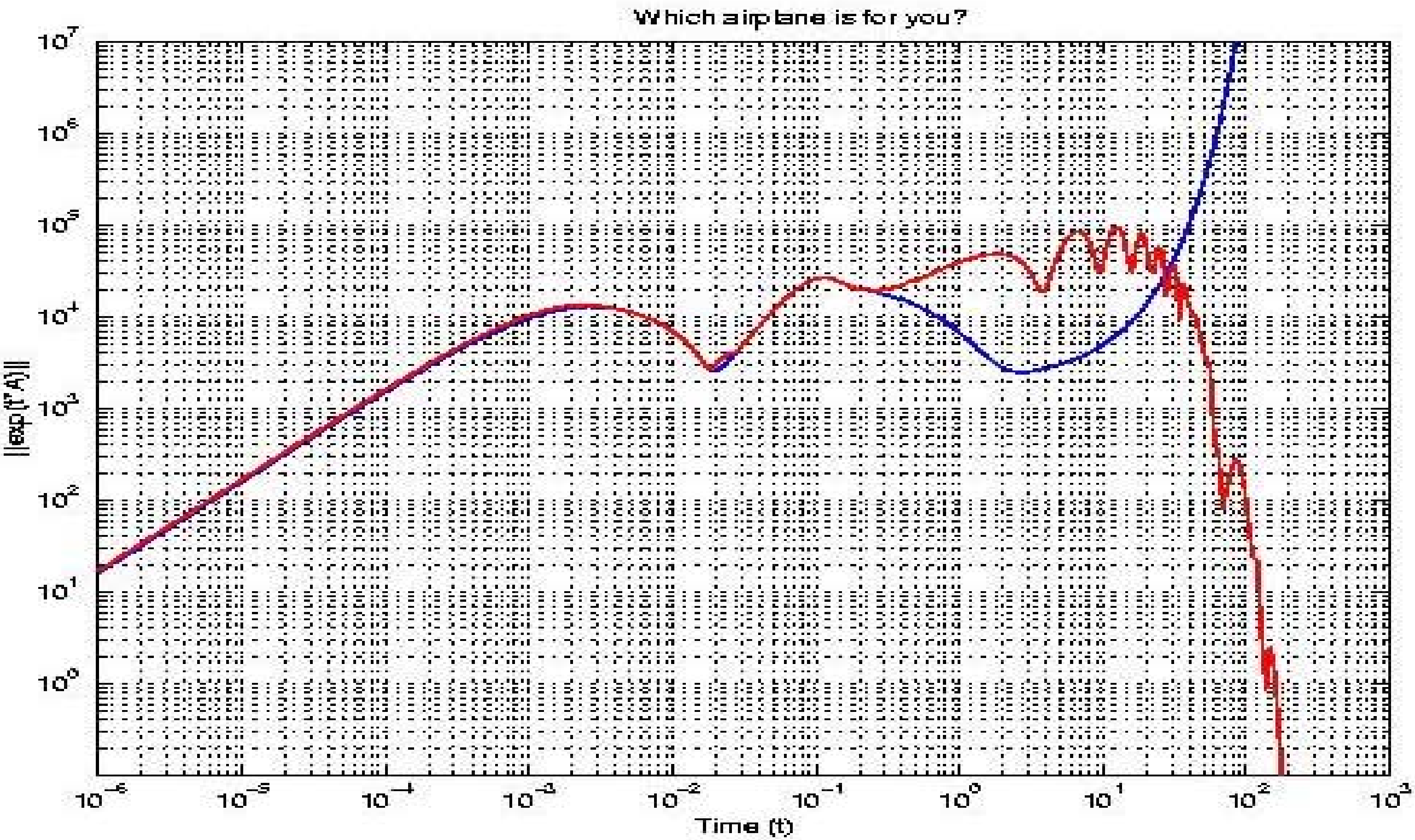
Optimized Eigenvalues
Original Eigenvalues
Imaginary Axis
1000
800
600
400
200
0
−200
−400
−600
−800
−1000
−1000
−800
−600
−400
−200
0
200
400
600
800
1000
−2
−3
−4
−5
−6
−7


Optimized Eigenvalues
Original Eigenvalues
Imaginary Axis
200
150
100
50
0
−50
−100
−150
−200
−200
−150
−100
−50
0
50
100
150
200
−2
−3
−4
−5
−6
−7


Optimized Eigenvalues
Original Eigenvalues
Original Unstable Eigenvalues
Imaginary Axis
20
15
10
5
0
−5
−10
−15
−20
−20
−15
−10
−5
0
5
10
15
20
−2
−3
−4
−5
−6
−7



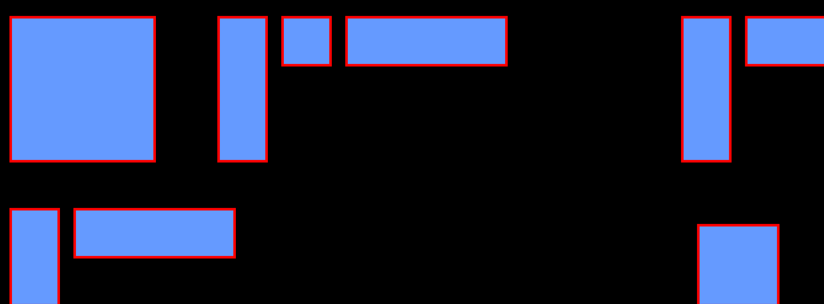
Optimized Eigenvalues
Original Eigenvalues
Imaginary Axis
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
−2
−3
−4
−5
−6
−7


Which airplane is for you?
||exp(t*A)||
Time (t)
10^{7}
10^{6}
10^{5}
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

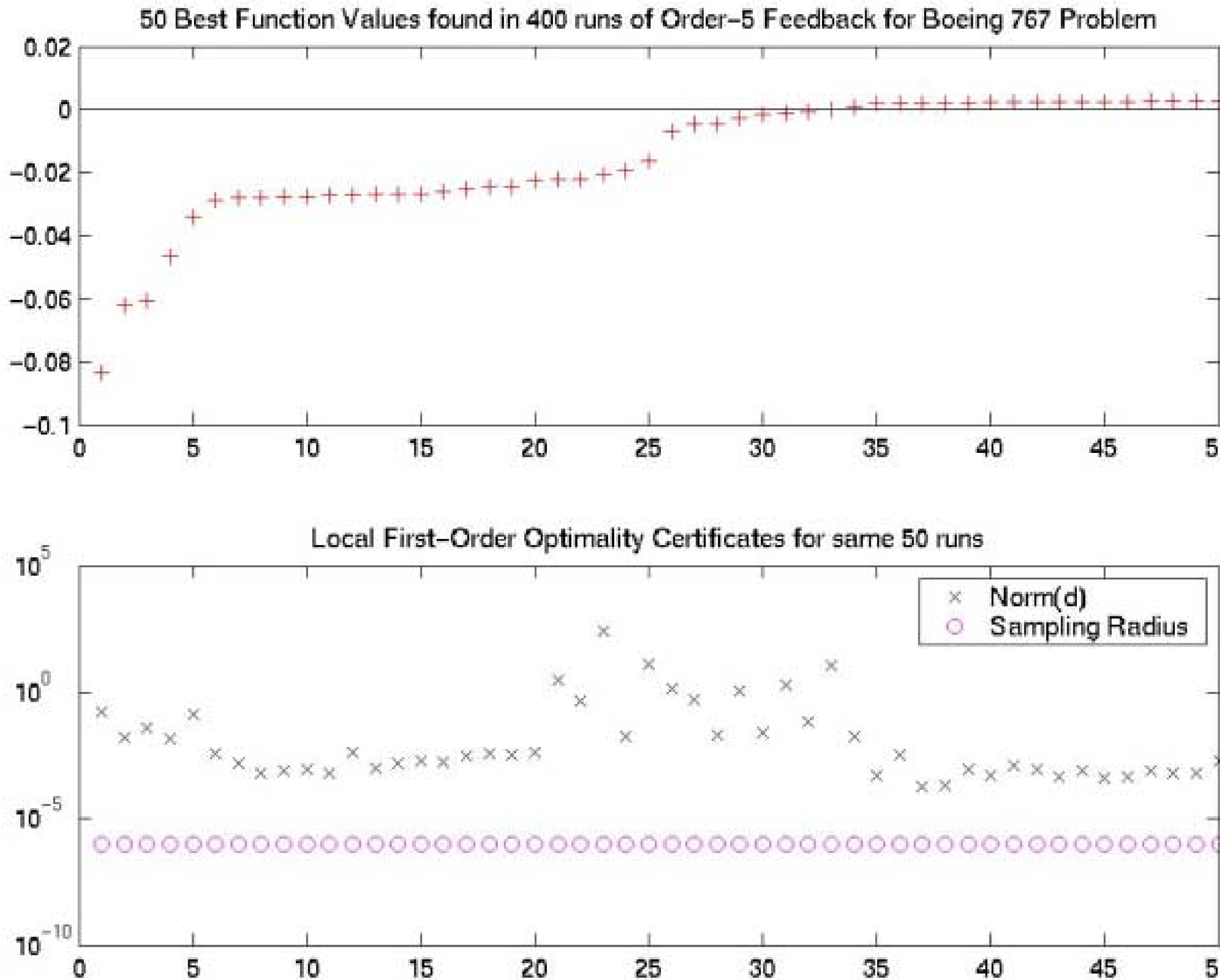
Low-Order Controller Design

- Stabilize the matrix

$$\begin{bmatrix} A_0 + B_0 X_1 C_0 & B_0 X_2 \\ X_3 C_0 & X_4 \end{bmatrix}$$



- Dimension of X_4 is order of controller
- Static output feedback is special case order = 0
- Still affine

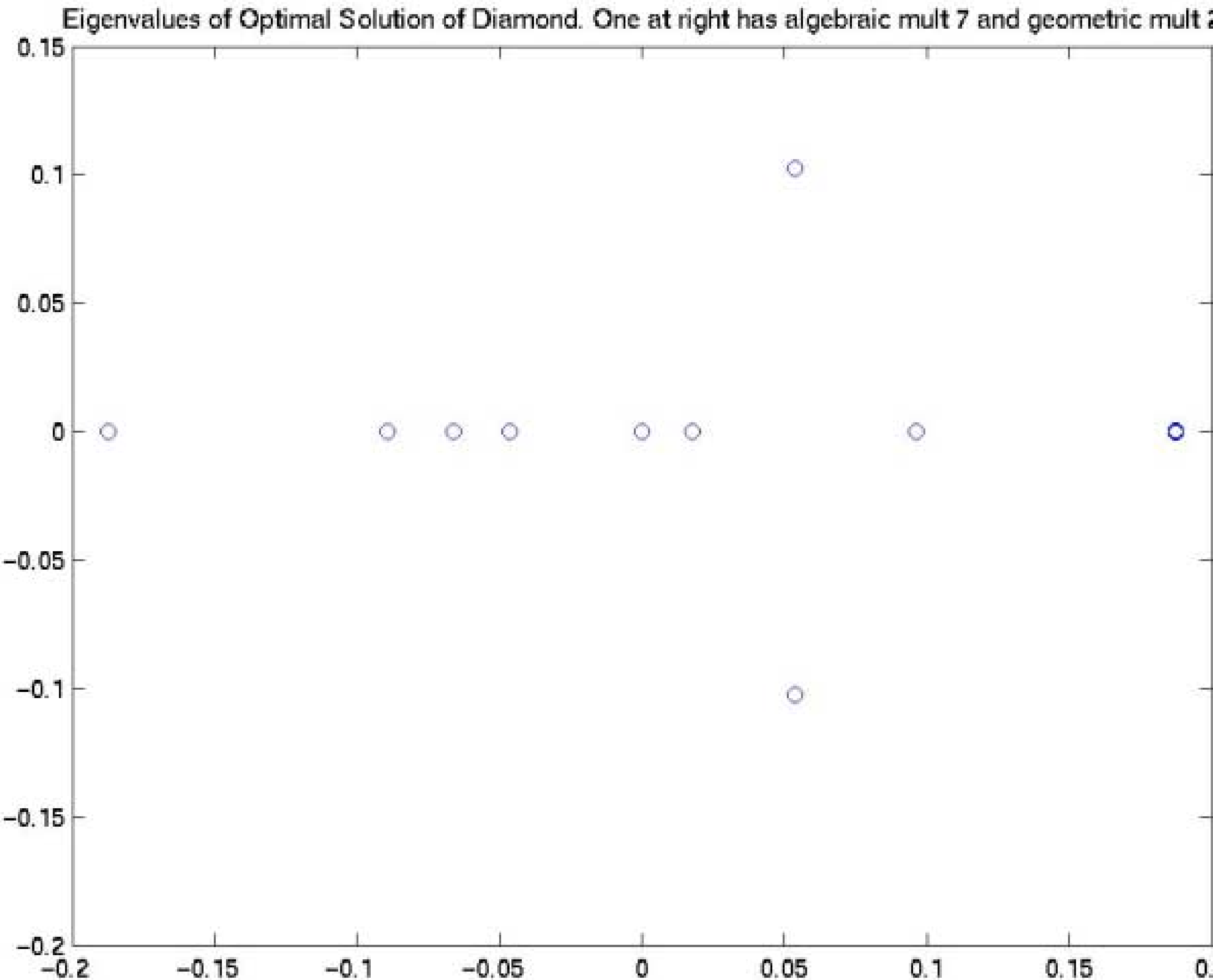

50 Best Function Values found in 400 runs of Order-5 Feedback for Boeing 767 Problem
0.02
0
−0.02
−0.04
−0.06
−0.08
−0.1
0
5
10
15
20
25
30
35
40
45
50
Local First-Order Optimality Certificates for same 50 runs
Norm(d)
Sampling Radius
10^5
10^0
10^−5
10^−10

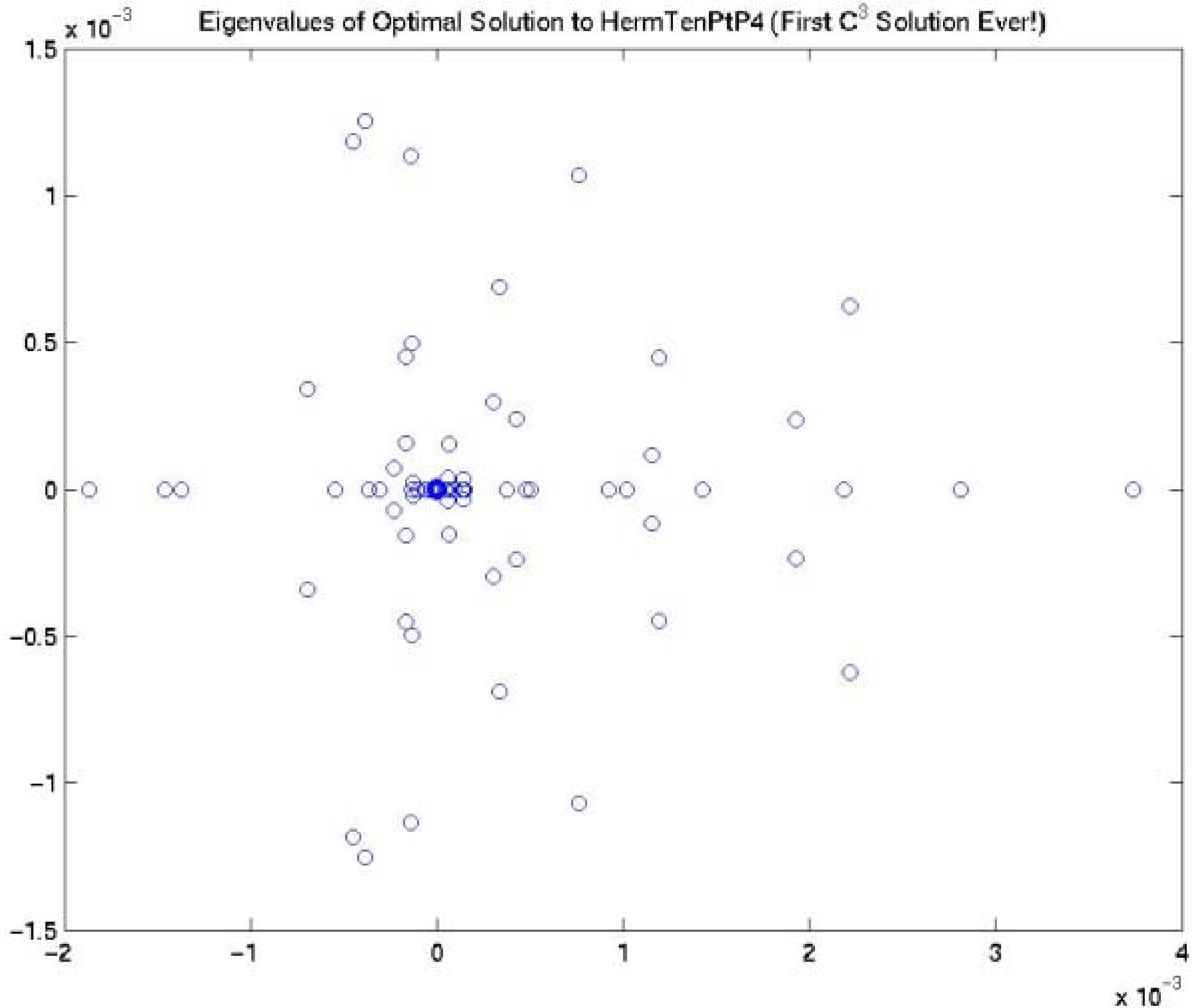
Convergence Theory for Gradient Sampling Method

- Suppose
  - f is locally Lipschitz and coercive
  - f is continuously differentiable on an open dense subset of its domain
  - number of gradients sampled near each iterate is greater than problem dimension
- Then, with probability one and for fixed sampling diameter η, algorithm generates a sequence of points with a cluster point x that is η-Clarke stationary
- If f has a unique Clarke stationary point x, then the set of all cluster points generated by the algorithm converges to x as η is reduced to zero

Subdivision Surface Design

- Thomas Yu, RPI
- Critical L^2 Sobolev smoothness of a refinable Hermite interpolant is given by spectral radius of a matrix dependent on the refinement mask
- Maximizing the smoothness amounts to minimizing the spectral radius


Eigenvalues of Optimal Solution of Diamond. One at right has algebraic mult 7 and geometric mult 2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
−0.2
−0.15
−0.1
−0.05
0
0.05
0.1
0.15
0.2


Eigenvalues of Optimal Solution to HermTenPtP4 (First C^3 Solution Ever!)
x 10^-3
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−1
0
1
2
3
4
x 10^-3

Beamforming Optimization

- Boche and Schubert
- Still at the email stage

Papers by J.V. Burke, A.S. Lewis and M.L. Overton (continued)

- Approximating Subdifferentials by Random Sampling of Gradients
  - *Math. Oper. Res.* 27 (2002), pp. 567-584
- Optimal Stability and Eigenvalue Multiplicity
  - *Foundations of Comp. Math*. 1 (2001), pp. 205-225
- Optimizing Matrix Stability
  - *Proc. Amer. Math. Soc.* 129 (2001), pp. 1635-1642

Papers by J.V. Burke and M.L. Overton

- Variational Analysis of Non-Lipschitz Spectral Functions
  - *Math. Programming* 90 (2001), pp. 317-352
- Variational Analysis of the Abscissa Mapping for Polynomials
  - *SIAM J. Control Optim.* 39 (2001), 1651-1676

- http://www.cs.nyu.edu/faculty/overton/

Papers by J.V. Burke, A.S. Lewis and M.L. Overton

- A Robust Gradient Sampling Algorithm for Nonsmooth, Nonconvex Optimization
  – In preparation, will be submitted to *SIAM J. Optim.*
- Robust Stability and a Criss-Cross Algorithm for Pseudospectra
  – To be submitted soon to *IMA J. Numer. Anal*.
- Optimization over Pseudospectra
  – To appear in *SIAM J. Matrix Anal. Appl.*
- Two Numerical Methods for Optimizing Matrix Stability
  – *Lin. Alg. Appl*. 351-352 (2002), pp.117-145