Convergence of Polynomial Restart Krylov Methods

$$\mathbf{Ax} = \mathbf{x}\lambda$$

D.C. Sorensen

Rice University

joint work with

C. Beattie and M. Embree

BIRS Workshop
Banff 2003

Outline – Polynomial Restart Convergence

Motivation: To Understand

Christopher Beattie, Mark Embree and John Rossi, *Convergence of restarted Krylov subspaces to invariant subspaces*, SIAM J. Matrix Anal. Appl. (to appear)

(Also: Saad, Jia, Stewart,...)

- The Question of Convergence
- Maximal Reachable Invariant Subspace
- Convergence Estimates with No Restarts
- Convergence of Restarted Krylov Subspaces
- Estimates from Complex Approximation Theory
- Examples

Polynomial Restart Spaces:

Goal: Approximate a selected invariant subspace of $\mathbf{A}$ with the Krylov space:

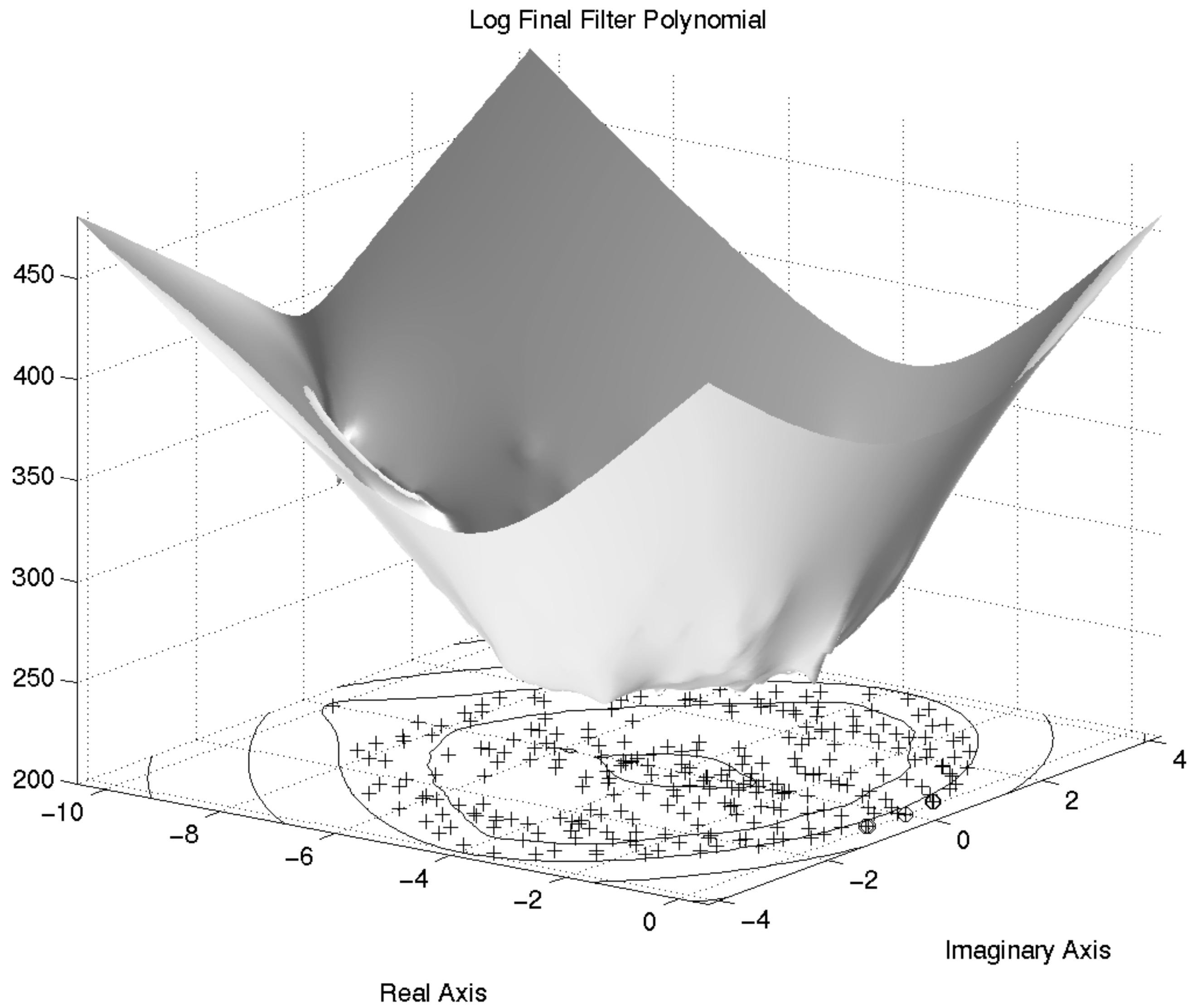
$$\mathcal{K}_k(\mathbf{A}, \mathbf{v_1}) := \operatorname{span}\{\mathbf{v_1}, \mathbf{A}\mathbf{v_1}, \ldots, \mathbf{A^{k-1}}\mathbf{v_1}\}.$$

Polynomial Restart Spaces: Project $\mathbf{A}$ onto $\mathcal{K}_m(\mathbf{A}, \mathbf{v_1^{(\ell)}})$

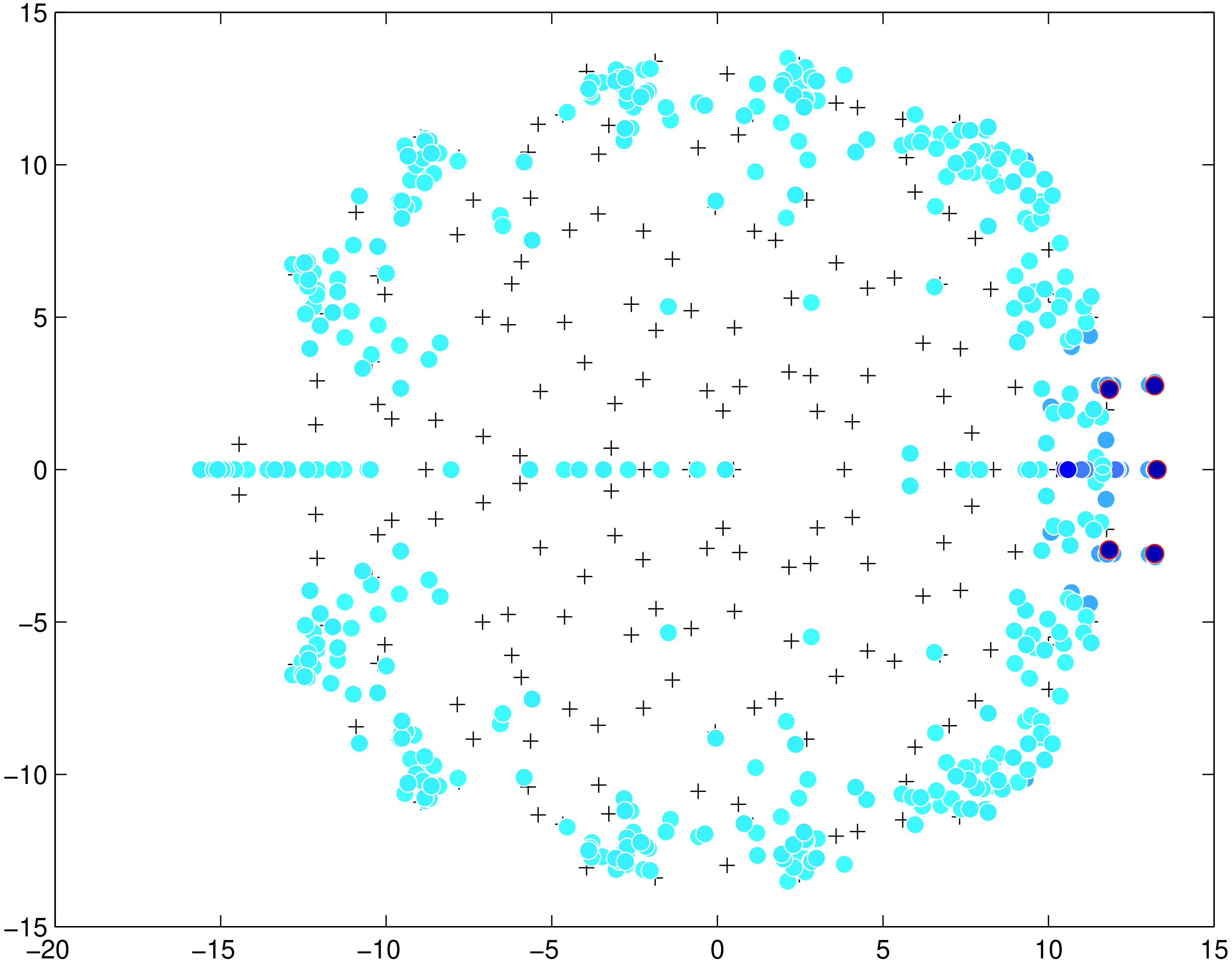
$$\mathbf{v_1^{(\ell)}} = \phi_\ell(\mathbf{A})\mathbf{v_1^{(\ell-1)}}, \qquad \mathbf{v_1^{(0)}} = \mathbf{v_1} \text{ and } \deg(\phi_\ell) < m.$$

Aggregate Filter Polynomial:

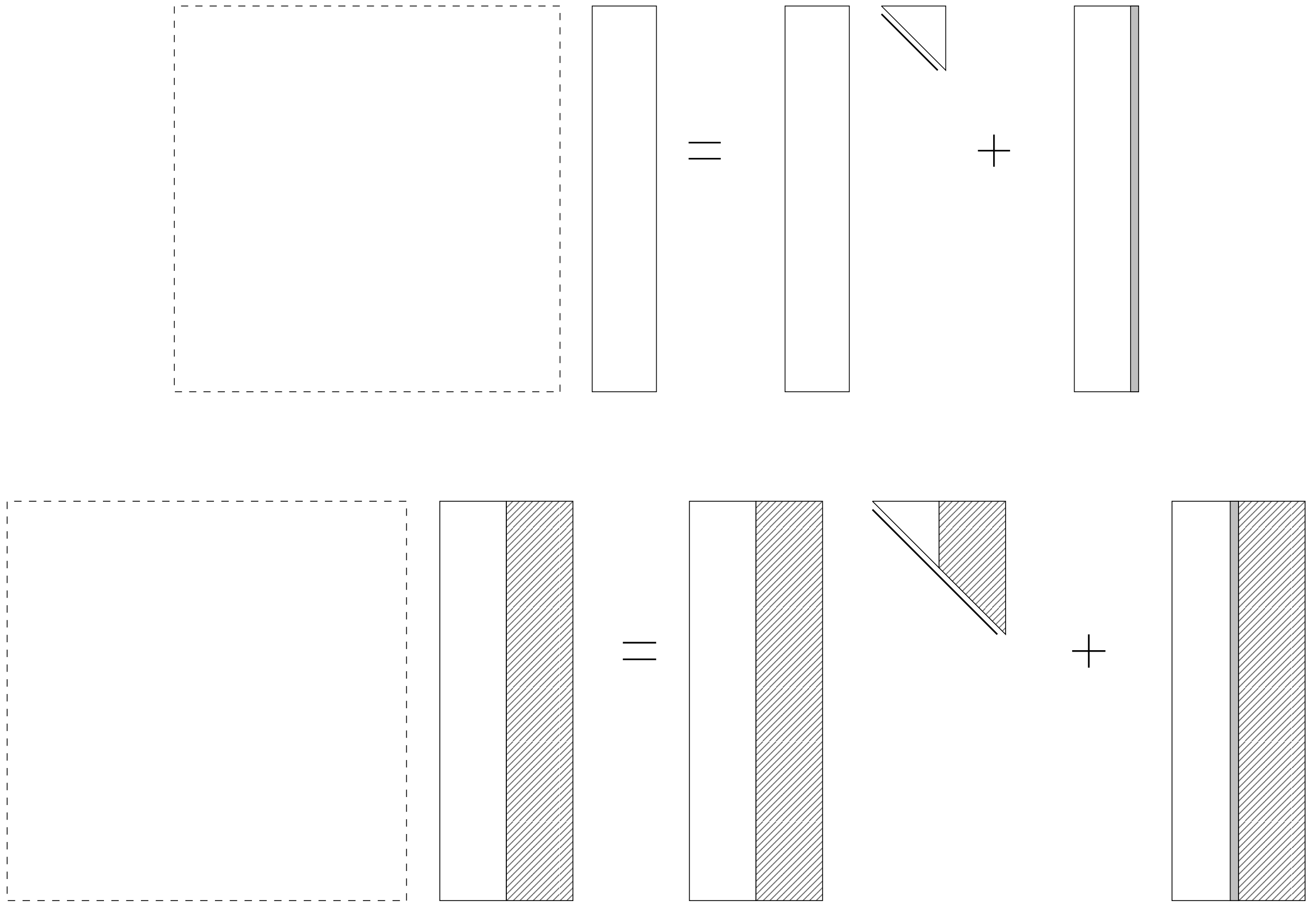
$$\Phi_\ell(z) = \prod_{j=1}^{\ell} \phi_j(z), \qquad \mathbf{v_1^{(\ell)}} = \boldsymbol{\Phi}_\ell(\mathbf{A})\mathbf{v_1};$$


Log Final Filter Polynomial
450
400
350
300
250
200
−10
−8
−6
−4
−2
0
−4
−2
0
2
4
Real Axis
Imaginary Axis

IRAM – Converging Ritz Values



Implicitly Restarted Arnoldi



Decomposition of $\mathbb{C}^n$ into $\mathcal{X}_g$ plus $\mathcal{X}_b$

Good and Bad Eigenvalues (Spaces)

$$\mathcal{X}_g := \bigoplus_{j=1}^{L} \mathrm{Ker}(\mathbf{A} - \lambda_j \mathbf{I})^{n_j} \quad \text{and} \quad \mathcal{X}_b := \bigoplus_{\ell=L+1}^{N} \mathrm{Ker}(\mathbf{A} - \lambda_\ell \mathbf{I})^{n_\ell}$$

Spectral Decomposition : $\qquad \mathbb{C}^n = \mathcal{X}_g \oplus \mathcal{X}_b$

Spectral Projectors $\mathbf{P}_g$ and $\mathbf{P}_b$:

$$\mathcal{X}_g = \mathrm{Range}(\mathbf{P}_g), \qquad \mathcal{X}_b = \mathrm{Range}(\mathbf{P}_b)$$

Krylov Subspace Decomposition

$$\mathcal{K}(\mathbf{A}, \mathbf{v_1}) = \mathcal{K}(\mathbf{A}, \mathbf{P_g v_1}) \oplus \mathcal{K}(\mathbf{A}, \mathbf{P_b v_1}).$$

Convergence in Gap - Distance to a Subspace

The *containment gap* between the subspaces $\mathcal{W}$ and $\mathcal{V}$:

$$\delta(\mathcal{W}, \mathcal{V}) := \max_{\mathbf{w} \in \mathcal{W}} \min_{\mathbf{v} \in \mathcal{V}} \frac{\|\mathbf{w} - \mathbf{v}\|}{\|\mathbf{w}\|}.$$

Note: $\delta(\mathcal{W}, \mathcal{V})$ is the sine of the largest canonical angle between $\mathcal{W}$ and the closest subspace of $\mathcal{V}$ with the same dimension as $\mathcal{W}$.

Convergence in Gap : We want

$$\delta(\mathcal{K}_m(\mathbf{A}, \mathbf{v}_{\mathbf{1}}^{(\ell)}), \mathcal{X}_{\mathbf{g}}) \to \mathbf{0},$$

where $\mathcal{X}_g$ is the desired invariant subspace for the “good” eigenvalues

Maximal Reachable Subspace

The maximal reachable invariant subspace

$$\mathcal{U}_g := \mathcal{K}(\mathbf{A}, \mathbf{P}_g \mathbf{v_1}) = \mathcal{K}(\mathbf{A}, \mathbf{v_1}) \cap \mathcal{X}_{\mathbf{g}}.$$

Largest invariant subspace of $\mathcal{X}_g$ to which our Krylov subspace can possibly converge.

- A derogatory good eigenvalue $\Rightarrow \quad \mathcal{U}_g \neq \mathcal{X}_g$!
- No single vector restart can fully capture $\mathcal{X}_g$.
- *If* $\mathcal{U}_g$ *is a proper subset of* $\mathcal{X}_g$, *then*

$$\delta(\mathcal{X}_g, \mathcal{K}(\mathbf{A}, \mathbf{v_1})) \geq \frac{\mathbf{1}}{\|\mathbf{P_g}\|}.$$

Convergence Questions

Two fundamental questions:

1) What is the gap $\delta(\mathcal{U}_g, \mathcal{K}_k(\mathbf{A}, \mathbf{v_1}))$ as k increases?

Key: Nonnormality of $\mathbf{A}$; Distribution of $\mathbf{v_1}$ w. r. t. $\mathcal{U}_g$.

2) How does $\delta(\mathcal{U}_g, \mathcal{K}_m(\mathbf{A}, \widehat{\mathbf{v}}_1))$ depend on $\widehat{\mathbf{v}}_1 = \Phi(\mathbf{A})\mathbf{v_1}$, and how can we optimize the asymptotic behavior?

Goal of restarting: Attain the unrestarted iteration performance, but within *restricted subspace dimensions.*

Some Definitions

Minimal Polynomial for $\mathcal{X}_g$:

$$\alpha_g := \textit{minimal polynomial of } \mathbf{A} \textit{ with respect to } \mathbf{P}_g \mathbf{v_1},$$

Monic polynomial of least degree s.t. $\alpha_g(\mathbf{A})\mathbf{P}_g\mathbf{v_1} = \mathbf{0}$.

Non-normality constant $\kappa(\Omega)$:

Smallest positive number s.t.

$$\|f(\mathbf{A})\,\mathbf{\Pi}_{\mathcal{U}}\| \le \kappa(\Omega) \max_{z\in\Omega} |f(z)|$$

uniformly for all functions f analytic on Ω.

ε-pseudospectrum of $\mathbf{A}$

$$\Lambda_\varepsilon(\mathbf{A}) := \{z \in \mathbb{C} : \|(z\mathbf{I} - \mathbf{A})^{-1}\| \ge \varepsilon^{-1}\}$$

Convergence with No Restarts

$$\delta(\mathcal{U}_g, \mathcal{K}_\ell(\mathbf{A}, \mathbf{v_1})) \le \mathbf{C_o C_b} \min_{\mathbf{p} \in \mathcal{P}_{\ell-\mathbf{2m}}} \max_{\mathbf{z} \in \boldsymbol{\Omega}_\mathbf{b}} \left| \mathbf{1} - \alpha_\mathbf{g}(\mathbf{z}) \mathbf{p}(\mathbf{z}) \right| \tag{BES}$$

$$\delta(\mathcal{U}_g, \mathcal{K}_\ell(\mathbf{A}, \mathbf{v_1})) \le \sqrt{\mathbf{2}} \mathbf{C_o C_b C_g} \min_{\mathbf{q} \in \mathcal{P}_{\ell-\mathbf{m}}} \frac{\max\{|\mathbf{q}(\mathbf{z})| : \mathbf{z} \in \boldsymbol{\Omega}_\mathbf{b}\}}{\min\{|\mathbf{q}(\mathbf{z})| : \mathbf{z} \in \boldsymbol{\Omega}_\mathbf{g}\}} \tag{BER}$$

The compact set $\Omega_g \subseteq \mathbb{C} \setminus \Omega_b$ contains all the good eigenvalues.

$C_o := \max_{\psi \in \mathcal{P}_{m-1}} \frac{\|\psi(\mathbf{A}) \mathbf{P}_b \mathbf{v_1}\|}{\|\psi(\mathbf{A}) \mathbf{P}_g \mathbf{v_1}\|}$, $\quad C_b := \kappa(\Omega_b)$, $\quad C_g := \kappa(\Omega_g)$.

Estimates from Complex Approximation Theory

Construct conformal map $\mathcal{G}$ taking the exterior of Ω_b to the exterior of the unit disk with $\mathcal{G}(\infty) = \infty$ and $\mathcal{G}'(\infty) > 0$.

Define

$$\rho := \Big(\min_{j=1,\ldots,L} |\mathcal{G}(\lambda_j)| \Big)^{-1}.$$

Then (Gaier, Walsh)

$$\limsup_{k\to\infty} \min_{p\in\mathcal{P}_k} \max_{z\in\Omega_b} \left| \frac{1}{\alpha_g(z)} - p(z) \right|^{1/k} = \rho.$$

Image of $\{|z| = \rho^{-1}\}$ is a curve $\mathcal{C} := \mathcal{G}^{-1}(\{|z| = \rho^{-1}\})$ around Ω_b.

This critical curve passes through a good eigenvalue “closest to” Ω_b

Convergence with Restarts

Restarts with Asymptotically Optimal Points

Restart Shifts: Fejér, Fekete, or Leja points for Ω_b

(Leja points are "practical")

Let Ψ_M interpolate $1/\alpha_g(z)$ at the M restart shifts:

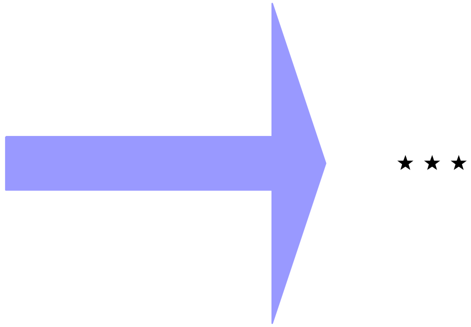
$$\boxed{\delta(\mathcal{U}_g, \mathcal{K}_\ell(\mathbf{A}, \widehat{\mathbf{v}}_1)) \le C_o C_g \max_{z \in \Omega_b} |1 - \Psi_M(z)\alpha_g(z)| \le C_o\, C_g\, C_r\, r^M}$$

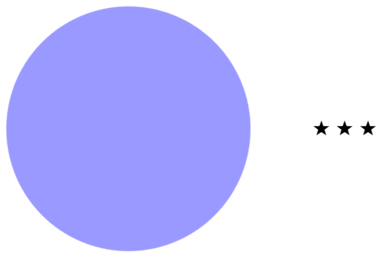
for any $r > \rho$. (Gaier §II.3, Fischer and Reichel 89)

$\widehat{\mathbf{v}}_1 = \Phi(\mathbf{A})\mathbf{v}_\mathbf{1}$,

Subspace dim: $\ell = 2m$, Restart deg: m, Agg. deg: $M = \nu m$

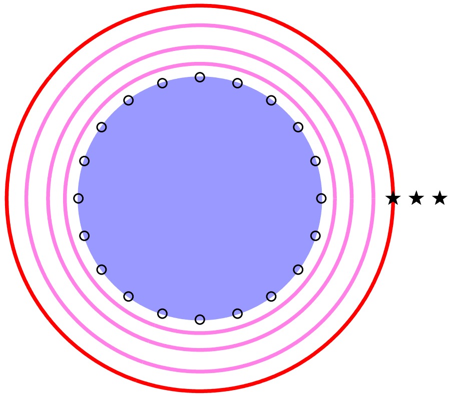
Map Bounding Region Ω_b to Unit Disk

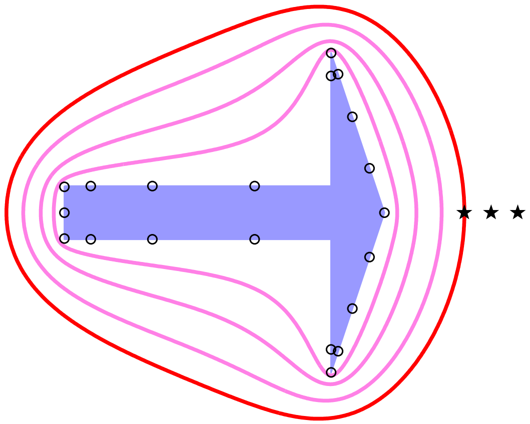




Convergence Rate

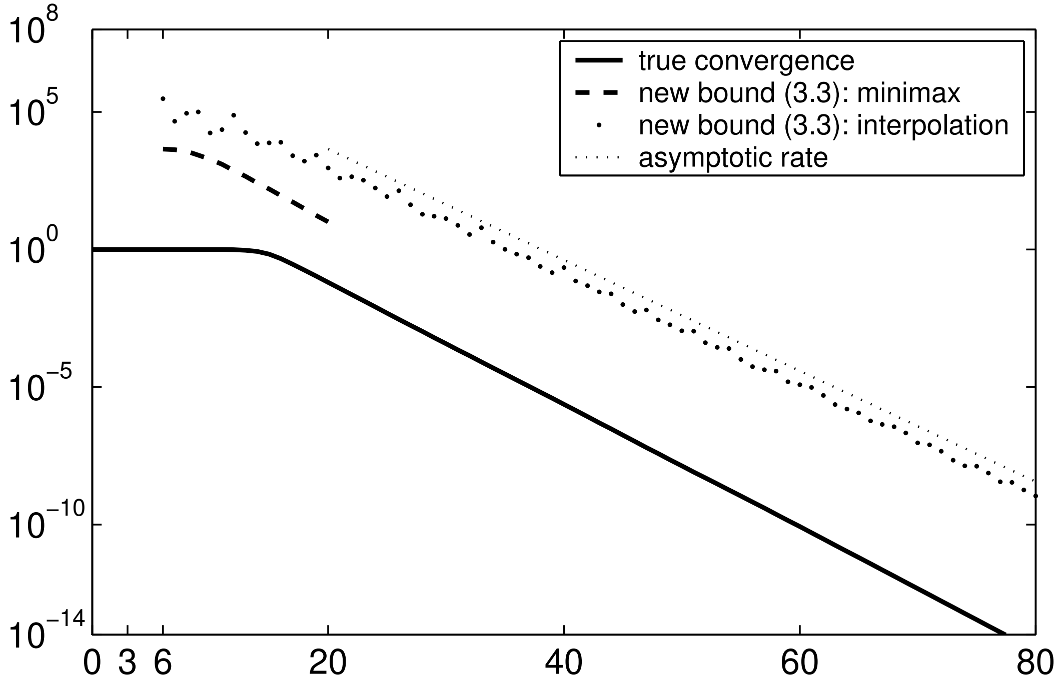
Asymptotic Convergence Rate = 1/radius





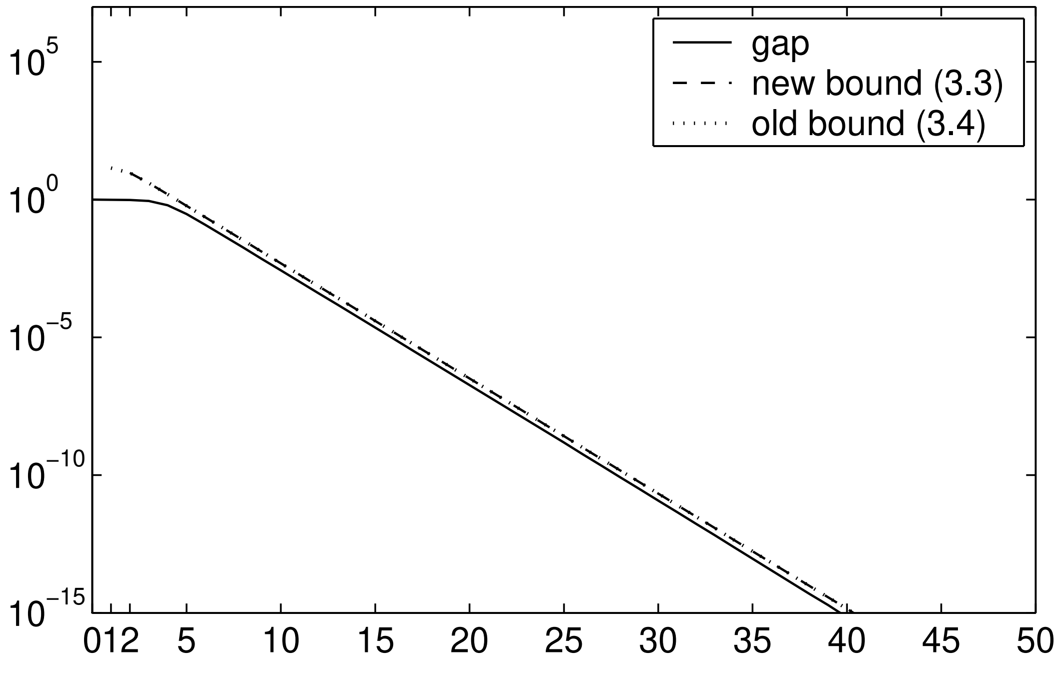
Inverting the map to original domain gives contours of equivalent convergence rates and asymptotically optimal interpolation points.

Gap Convergence $\delta(\mathcal{U}_g, \mathcal{K}_\ell(\mathbf{A}, \mathbf{v_1}))$ at iteration ℓ



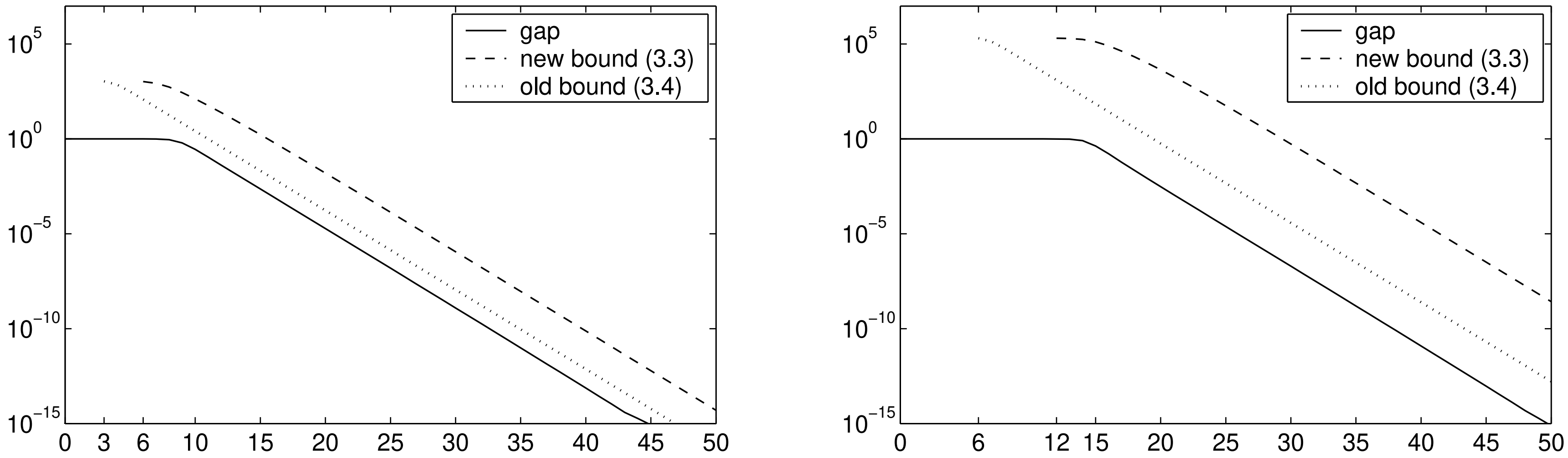
The better bound solves the minimax approximation problem directly, while the lesser bound approximates the optimal polynomial by interpolating $1/\alpha_g$ at Fejér points. This approximation degrades the bound but captures the asymptotic convergence rate.

Hermitian: Gap Convergence $\delta(\mathcal{U}_g, \mathcal{K}_\ell(\mathbf{A}, \mathbf{v_1}))$



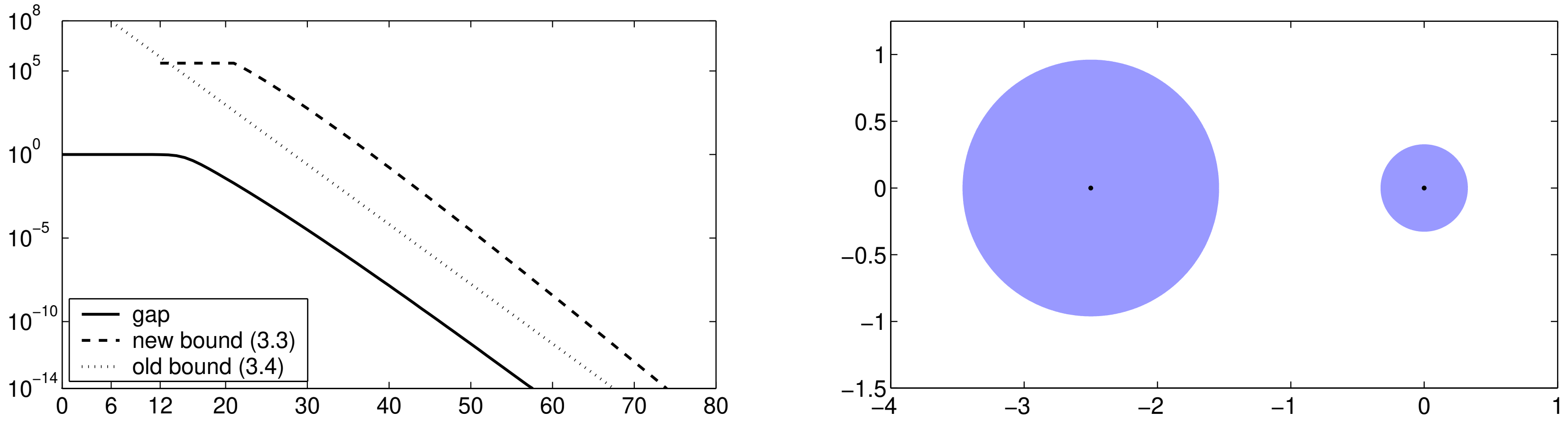
Comparison of convergence bounds for the Hermitian example with a single good eigenvalue. In this special case, the old and new bounds are identical.

Convergence $\delta(\mathcal{U}_g, \mathcal{K}_\ell(\mathbf{A}, \mathbf{v_1}))$: more wanted evals



Comparison of bounds: Hermitian, three good eigenvalues (left) and six good eigenvalues (right). The new bound degrades v.s. the old, but predicted asymptotic rate remains accurate.

Gap Convergence $\delta(\mathcal{U}_g, \mathcal{K}_\ell(\mathbf{A}, \mathbf{v_1}))$ for Defective A

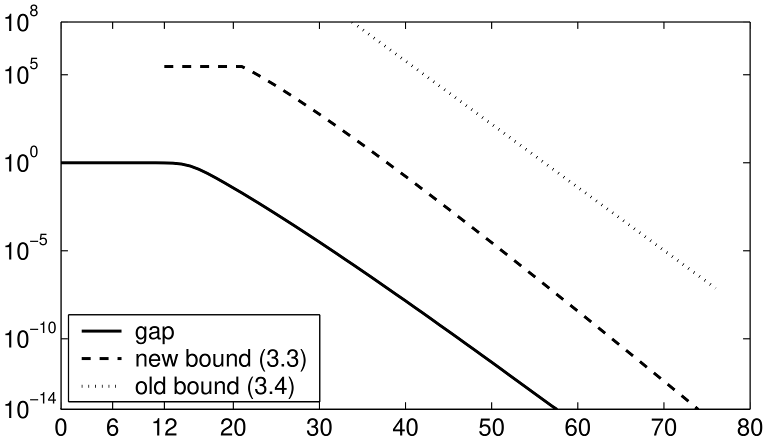


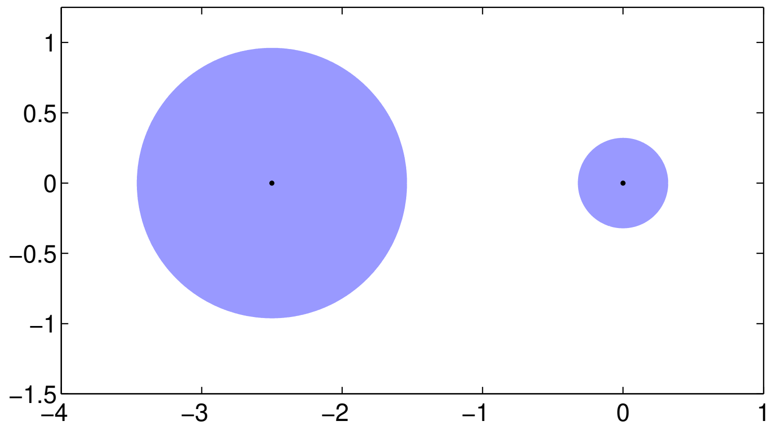
Comparison of convergence bounds for a nondiagonalizable example:

$\Omega_g = \Lambda_\varepsilon(\mathbf{J}_6(-\frac{5}{2}, 1))$ *and* $\Omega_b = \Lambda_\varepsilon(\mathbf{J}_{100}(0, 1))$ *for* $\varepsilon = 10^{-3}$.

New bound: approximate optimal polynomial with interpolant to $1/\alpha_g$ *at Fejér points.*

Gap Convergence for Increasing Sensitivity

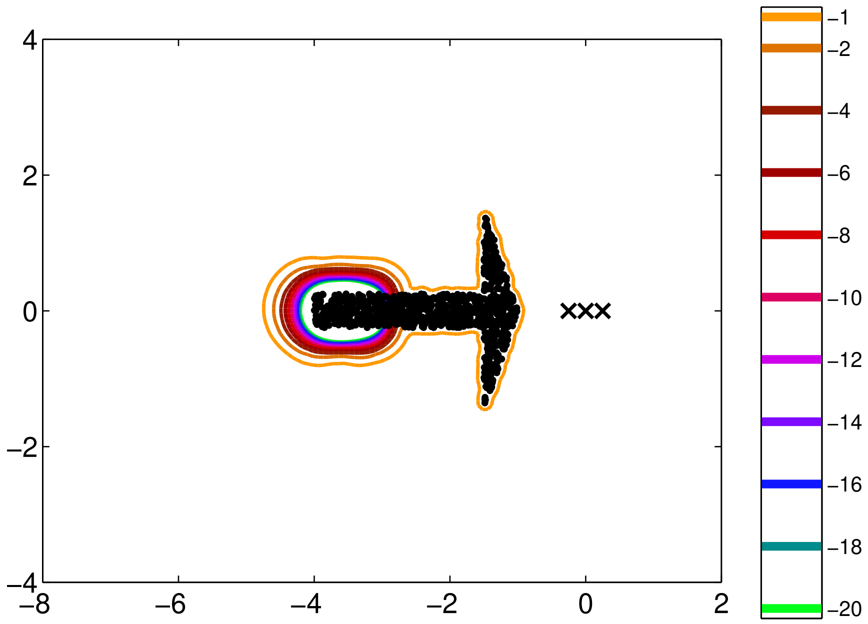




Good Jordan block $\mathbf{J}_6(-\frac{5}{2}, 1))$ *is replaced by* $\mathbf{J}_6(-\frac{5}{2}, 100))$.

This increases the sensitivity of the good eigenvalues, and now the new bound is superior to the old one.

Convergence for Restarting with High Non-normality

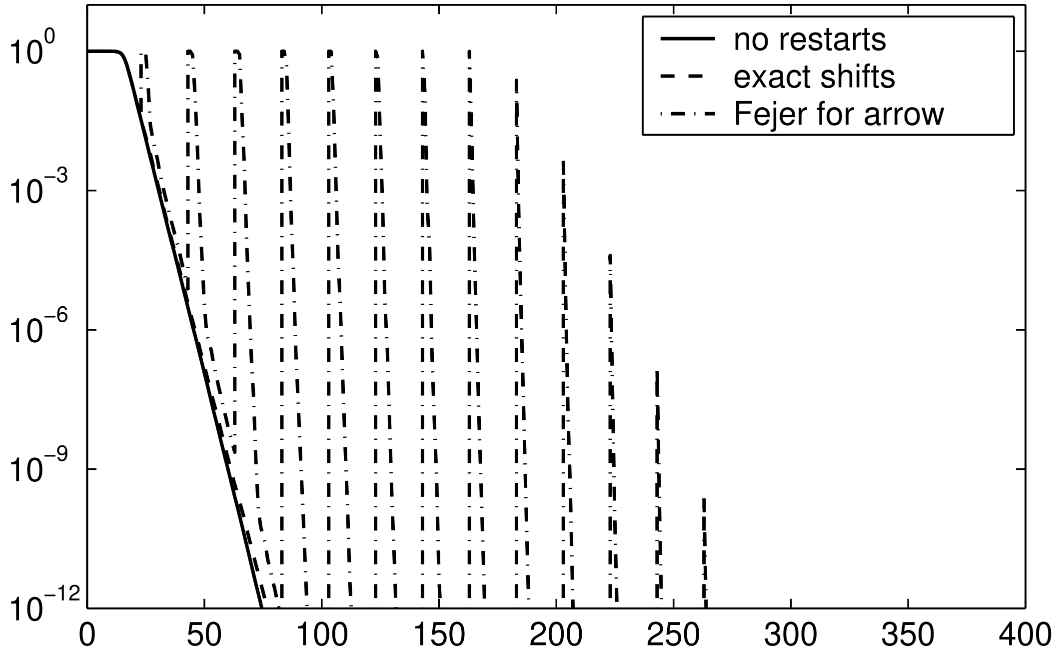


The ε*-pseudospectrum of* $\mathbf{A}_{\rm bad}$*, with*

$$\varepsilon = 10^{-1}, 10^{-2}, 10^{-4}, \ldots, 10^{-20}.$$

Bad eigenvalues fill arrow-shaped region; the three good eigenvalues are denoted by $\times$*.*

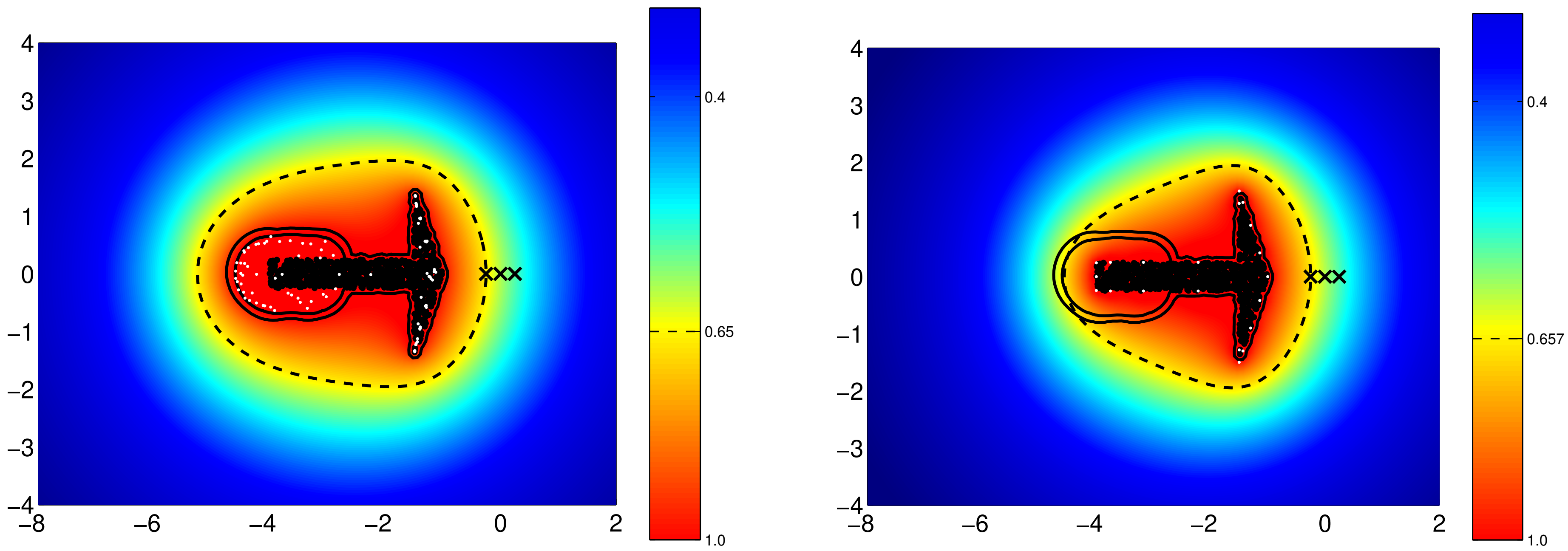
Gap Convergence for Restarting



Gap $\delta(\mathcal{U}_g, \text{approximating subspace})$ vs Number of Matvecs

Restarting with exact shifts gives near optimal (no restart) convergence. Restarting with shifts at the degree-20 Fejér points for the arrow-shaped Ω_b gives slower convergence (saw-tooth history).

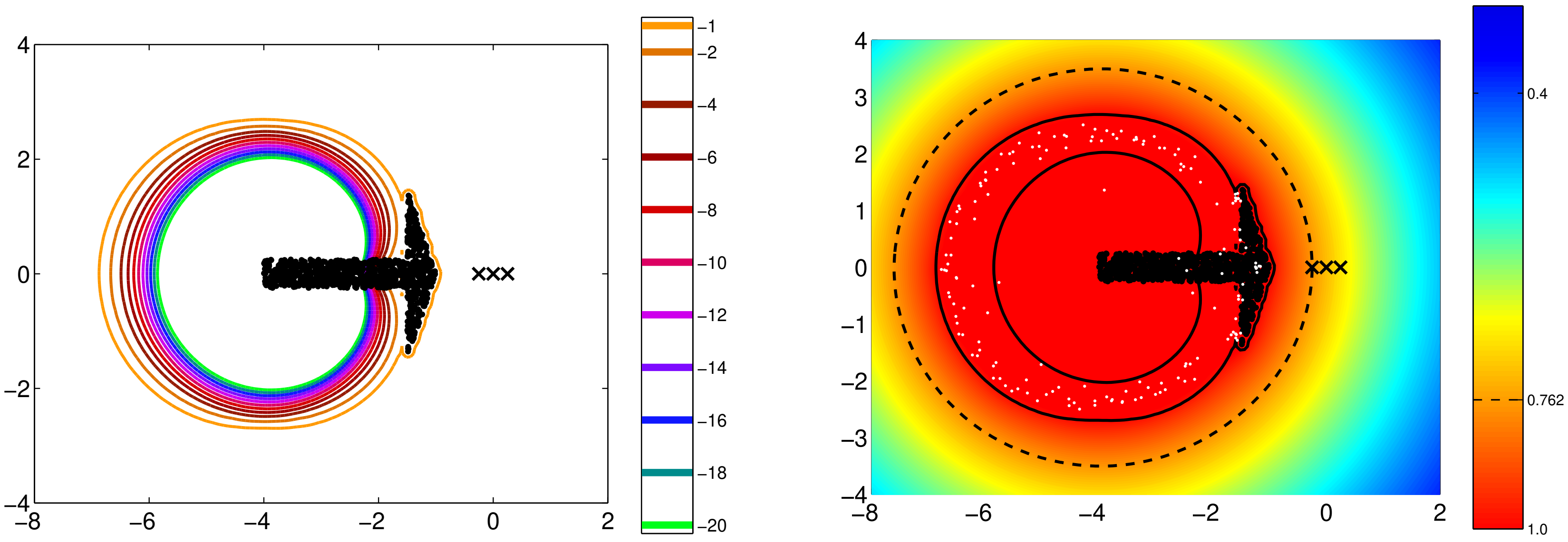
Convergence for Restarting with High Non-normality



These plots show the magnitude of the aggregate exact shift polynomial (left) and the degree-20 Fejér polynomial (right).

Roots of these polynomials are shown as white dots. Areas with the same color yield the same convergence rate.

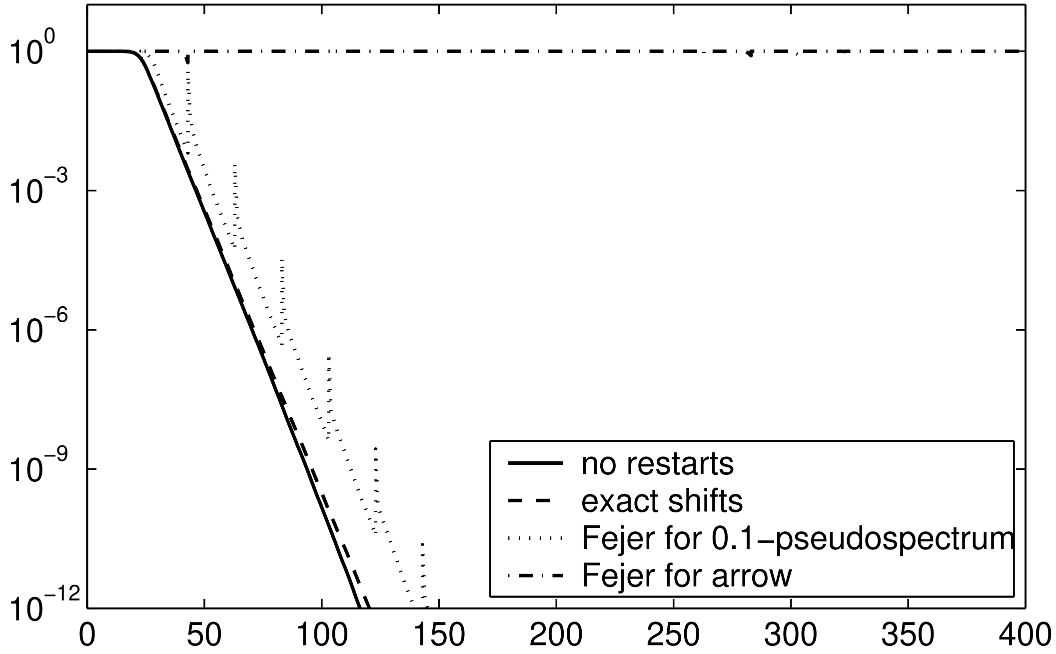
Restarting with Extreme Non-normality



(Left) The ε-pseudospectrum of $\mathbf{A}_{\mathrm{bad}}$, with $\varepsilon = 10^{-1}, 10^{-2}, \ldots, 10^{-20}$. Bad eigenvalues fill arrow-shaped region; the three good eigenvalues are denoted by $\times$.

(Right) This plot shows the magnitude of the aggregate exact shift polynomial.

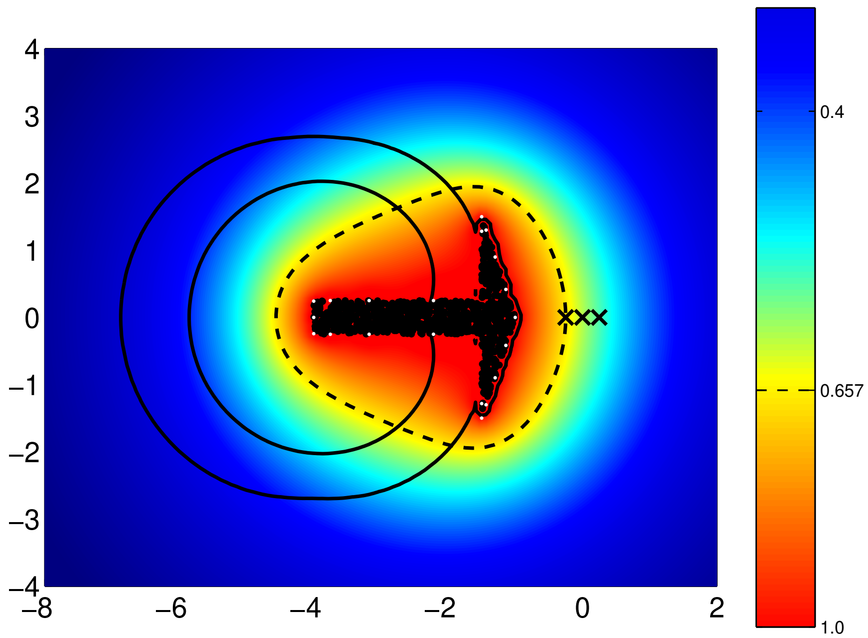
Restarting with Extreme Non-normality

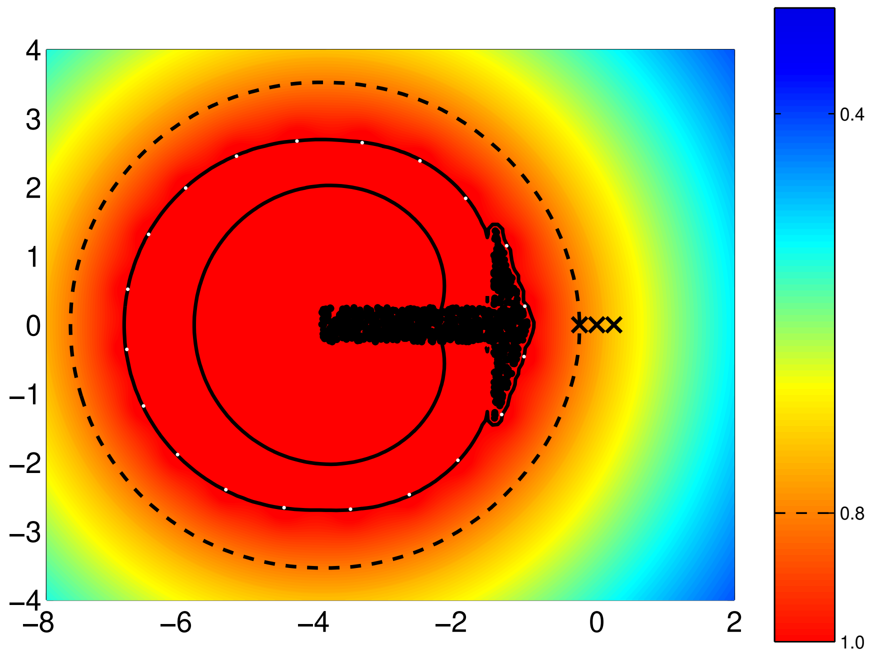


Gap $\delta(\mathcal{U}_g, \text{approximating subspace})$ vs Number of Matvecs

Restarting with exact shifts gives near optimal convergence. Fejér points for the $\varepsilon = 10^{-1}$ *pseudospectrum give good convergence. Fejér points for the arrow-shaped set* Ω_b *yield little convergence.*

Restarting with Extreme Non-normality





These plots show the magnitude of the degree-20 Fejér polynomial for the arrow (left), and the degree-20 Fejér polynomial for the 10^{-1}*-pseudospectrum (right). The solid lines now denote the boundaries of the* 10^{-1}*- and* 10^{-20}*-pseudospectra of* $\mathbf{A}_{\text{bad}}$.

Summary

Report CAAM TR 03-08 available at

`http:\\www.caam.rice.edu`

- Theory is Complete and Straightforward
- No Qualifying Assumptions (e.g. non-defective)
- Correct Asymptotic Rate Always Predicted
- Restarting Can Achieve Optimal Convergence Rate
- Bounds are Not Quantitative
- Exact Shift Behavior Observed – Not Proven