Convergence Theory for Iterative Eigensolvers

Mark Embree
Virginia Tech

with Chris Beattie, Russell Carden, John Rossi, Dan Sorensen

RandNLA Workshop · Simons Institute · September 2018

setting for the talk

- Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a large square matrix, potentially non-Hermitian ($\mathbf{A} \neq \mathbf{A}^*$).
- Computing all eigenvalues of $\mathbf{A}$ is too expensive (and usually not needed).
- Thus we seek $m \ll n$ distinguished eigenvalues relevant to our application (*largest, smallest, rightmost, etc.*)

setting for the talk

- Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a large square matrix, potentially non-Hermitian ($\mathbf{A} \neq \mathbf{A}^*$).
- Computing all eigenvalues of $\mathbf{A}$ is too expensive (and usually not needed).
- Thus we seek $m \ll n$ distinguished eigenvalues relevant to our application (*largest, smallest, rightmost, etc.*)
- **Projection Methods**
  $\mathcal{V} \subset \mathbb{C}^n = k$-dimensional subspace of $\mathbb{C}^n$, *the projection subspace for* $\mathbf{A}$
  The columns of $\mathbf{V} \in \mathbb{C}^{n\times k}$ for an orthonormal basis for $\mathcal{V}$:

$$\mathbf{V}^*\mathbf{V} = \mathbf{I}, \qquad \mathbf{V}^*\mathbf{A}\mathbf{V} \in \mathbb{C}^{k\times k}$$

We hope *some* eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ $\quad \sigma(\mathbf{V}^*\mathbf{A}\mathbf{V}) = \{\theta_1, \ldots, \theta_k\}$

approximate *some* eigenvalues of $\mathbf{A}$. $\quad \sigma(\mathbf{A}) = \{\lambda_1, \ldots, \lambda_n\}$

For example, $\quad \theta_1 \approx \lambda_1, \quad \ldots \quad \theta_m \approx \lambda_m \quad$ for some $1 \leq m \leq k$.

setting for the talk

- Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a large square matrix, potentially non-Hermitian ($\mathbf{A} \neq \mathbf{A}^*$).
- Computing all eigenvalues of $\mathbf{A}$ is too expensive (and usually not needed).
- Thus we seek $m \ll n$ distinguished eigenvalues relevant to our application (*largest, smallest, rightmost, etc.*)
- **Projection Methods**
  $\mathcal{V} \subset \mathbb{C}^n = k$-dimensional subspace of $\mathbb{C}^n$, *the projection subspace for* $\mathbf{A}$
  The columns of $\mathbf{V} \in \mathbb{C}^{n \times k}$ for an orthonormal basis for $\mathcal{V}$:

$$\mathbf{V}^*\mathbf{V} = \mathbf{I}, \qquad \mathbf{V}^*\mathbf{A}\mathbf{V} \in \mathbb{C}^{k \times k}$$

  We hope *some* eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ $\quad \sigma(\mathbf{V}^*\mathbf{A}\mathbf{V}) = \{\theta_1, \ldots, \theta_k\}$

  approximate *some* eigenvalues of $\mathbf{A}$. $\quad \sigma(\mathbf{A}) = \{\lambda_1, \ldots, \lambda_n\}$

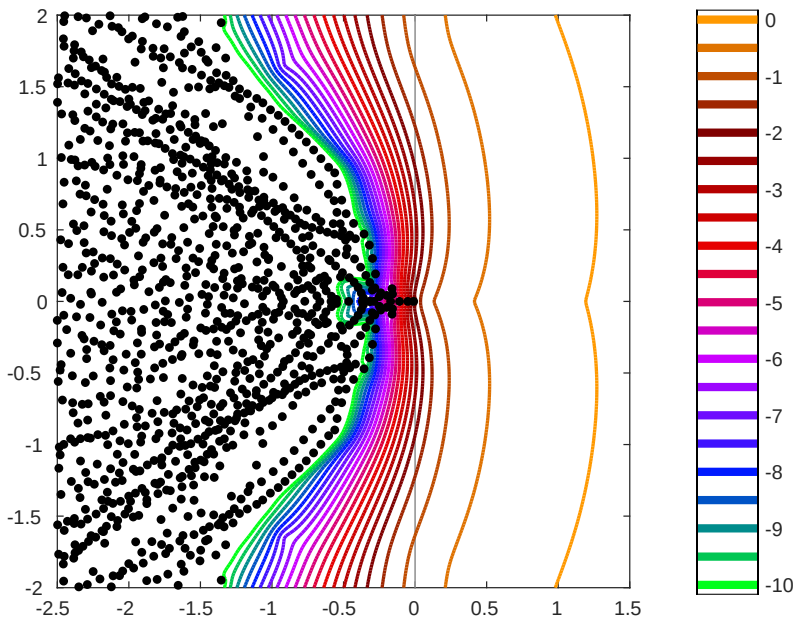
  For example, $\quad \theta_1 \approx \lambda_1, \quad \ldots \quad \theta_m \approx \lambda_m \quad$ for some $1 \leq m \leq k$.

- *This talk mainly describes established results for the deterministic case, with some thoughts from a RandNLA perspective along the way.*

why consider $A \neq A^*$?

While Hermitian problems are common (SVD, quantum mechanics, etc.), many important applications lead to non-Hermitian problems – and subtler issues of spectral perturbation theory. Many examples: [Trefethen, E. 2005].

- atmospheric science — Farrell . . .
- fluid flow stability — Trefethen; Schmid & Henningson; . . .
- damped mechanical systems — Cox & Zuazua, . . .
- control theory — Hinrichsen & Pritchard, . . .
- data-driven modeling — Antoulas, Beattie, Gugercin, . . .
- lasers — Landau, Siegman, . . .
- ecology — May, Caswell, . . .
- Markov chains — Diaconis, . . .



computed rightmost eigenvalues for stability of 2d flow over a backward facing step, $n = 381{,}539$
[E., Keeler 2017]

An Overview of Projection-Based Eigensolvers

projection-based eigensolvers

$\mathcal{V} \subset \mathbb{C}^n$ = k-dimensional subspace of $\mathbb{C}^n$, *the projection subspace for* **A**.

The columns of $\mathbf{V} \in \mathbb{C}^{n \times k}$ form an orthonormal basis for $\mathcal{V}$:

$$\mathbf{V}^*\mathbf{V} = \mathbf{I}, \qquad \mathbf{V}^*\mathbf{A}\mathbf{V} \in \mathbb{C}^{k \times k}$$

We hope *some* eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ $\quad \sigma(\mathbf{V}^*\mathbf{A}\mathbf{V}) = \{\theta_1, \ldots, \theta_k\}$

approximate *some* eigenvalues of **A**. $\quad \sigma(\mathbf{A}) = \{\lambda_1, \ldots, \lambda_n\}$

For example, $\quad \theta_1 \approx \lambda_1, \quad \ldots \quad \theta_m \approx \lambda_m \quad$ for some $1 \leq m \leq k$.

projection-based eigensolvers

$\mathcal{V} \subset \mathbb{C}^n$ = k-dimensional subspace of $\mathbb{C}^n$, *the projection subspace for* $\mathbf{A}$.

The columns of $\mathbf{V} \in \mathbb{C}^{n \times k}$ form an orthonormal basis for $\mathcal{V}$:

$$\mathbf{V}^*\mathbf{V} = \mathbf{I}, \qquad \mathbf{V}^*\mathbf{A}\mathbf{V} \in \mathbb{C}^{k \times k}$$

We hope *some* eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ $\quad \sigma(\mathbf{V}^*\mathbf{A}\mathbf{V}) = \{\theta_1, \ldots, \theta_k\}$

approximate *some* eigenvalues of $\mathbf{A}$. $\quad \sigma(\mathbf{A}) = \{\lambda_1, \ldots, \lambda_n\}$

For example, $\quad \theta_1 \approx \lambda_1, \quad \ldots \quad \theta_m \approx \lambda_m \quad$ for some $1 \leq m \leq k$.

Power method (minimal storage, easy to implement, can be slow)

$$\mathcal{V} = \text{span}\{\mathbf{A}^p\mathbf{x}\}$$

Subspace iteration (more storage, subtler to implement, computes repeated eigs)

$$\mathcal{V} = \text{Range}(\mathbf{A}^p\mathbf{X}) \text{ for } \mathbf{X} \in \mathbb{C}^{n \times k}$$

[Halko, Martinsson, Tropp 2011] et al.

projection-based eigensolvers: krylov methods

Power method (minimal storage, easy to implement, can be slow)

$$\mathcal{V} = \text{span}\{\mathbf{A}^p \mathbf{x}\}$$

Subspace iteration (more storage, subtler to implement, computes multiple eigs)

$$\mathcal{V} = \text{Range}(\mathbf{A}^p \mathbf{X}) \text{ for } \mathbf{X} \in \mathbb{C}^{n \times b}$$

Krylov subspace methods (growing subspace dimension; higher powers of $\mathbf{A}$)

$$\mathcal{V} = \text{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}$$

Block Krylov methods (subspace dimension grows quickly: $\dim(\mathcal{V}) \leq kb$)

$$\mathcal{V} = \text{Range}([\mathbf{X} \;\; \mathbf{A}\mathbf{X} \;\; \mathbf{A}^2\mathbf{X} \;\; \cdots \;\; \mathbf{A}^{k-1}\mathbf{X}]) \text{ for } \mathbf{X} \in \mathbb{C}^{n \times b}$$

SVD: [Musco & Musco 2015], [Drineas et al. 2018]

Must balance benefit of large k with block size b, storage.

projection-based eigensolvers: krylov methods (extensions)

Krylov subspace methods (growing subspace dimension; higher powers of $\mathbf{A}$)

$$\mathcal{V} = \text{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}$$

Restarted Krylov (used in `eigs`: filter ψ improves starting vector)

$$\mathcal{V} = \text{span}\{\psi(\mathbf{A})\mathbf{x}, \mathbf{A}\psi(\mathbf{A})\mathbf{x}, \mathbf{A}^2\psi(\mathbf{A})\mathbf{x}, \ldots, \mathbf{A}^{k-1}\psi(\mathbf{A})\mathbf{x}\}$$

Polynomial Preconditioned Krylov (very high degree polys, care needed)

$$\mathcal{V} = \text{span}\{\mathbf{x}, \pi(\mathbf{A})\mathbf{x}, \pi(\mathbf{A})^2\mathbf{x}, \ldots, \pi(\mathbf{A})^{k-1}\mathbf{x}\}$$

Shift-Invert Krylov (used in `eigs`: ideal for eigenvalues near μ)

$$\mathcal{V} = \text{span}\{\mathbf{x}, (\mathbf{A} - \mu\mathbf{I})^{-1})\mathbf{x}, (\mathbf{A} - \mu\mathbf{I})^{-2}\mathbf{x}, \ldots, (\mathbf{A} - \mu\mathbf{I})^{-(k-1)}\mathbf{x}\}$$

Rational Krylov (helps for finding eigenvalues in a region)

$$\mathcal{V} = \text{span}\{\mathbf{x}, (\mathbf{A} - \mu_1)^{-1}\mathbf{x}, (\mathbf{A} - \mu_2)^{-1}\mathbf{x}, \ldots, (\mathbf{A} - \mu_{k-1}\mathbf{I})^{-1}\mathbf{x}\}$$

preliminaries: spectral structure of A

- Distinct eigenvalues of $\mathbf{A}$: $\lambda_1, \lambda_2, \ldots, \lambda_{\widehat{n}}$
- Spectral projectors $\mathbf{P}_j$ and invariant subspaces $\mathcal{U}_j$:

$$\mathbf{P}_j := \frac{1}{2\pi i} \int_{\Gamma_j} (z\mathbf{I} - \mathbf{A})^{-1} \, dz, \qquad \mathcal{U}_j := \text{Range}(\mathbf{P}_j),$$

  Γ_j is a contour in $\mathbb{C}$ containing λ_j but no other distinct eigenvalues.
- If $\mathbf{A} = \mathbf{A}^*$ and λ_j is simple with unit eigenvector $\mathbf{u}_j$, then $\mathbf{P}_j = \mathbf{u}_j \mathbf{u}_j^*$.
- $\mathbf{P}_j$ is a projector onto the invariant subspace $\mathcal{U}_j$,
  but $\mathbf{P}_j$ need not be an orthogonal projector when $\mathbf{A} \neq \mathbf{A}^*$.
- The spectral projectors give a *resolution of the identity*: $\sum_{j=1}^{\widehat{n}} \mathbf{P}_j = \mathbf{I}$.

preliminaries: spectral structure of A

- Distinct eigenvalues of $\mathbf{A}$: $\lambda_1, \lambda_2, \ldots, \lambda_{\widehat{n}}$
- Spectral projectors $\mathbf{P}_j$ and invariant subspaces $\mathcal{U}_j$:

$$\mathbf{P}_j := \frac{1}{2\pi i} \int_{\Gamma_j} (z\mathbf{I} - \mathbf{A})^{-1} \, dz, \qquad \mathcal{U}_j := \text{Range}(\mathbf{P}_j),$$

  Γ_j is a contour in $\mathbb{C}$ containing λ_j but no other distinct eigenvalues.
- If $\mathbf{A} = \mathbf{A}^*$ and λ_j is simple with unit eigenvector $\mathbf{u}_j$, then $\mathbf{P}_j = \mathbf{u}_j \mathbf{u}_j^*$.
- $\mathbf{P}_j$ is a projector onto the invariant subspace $\mathcal{U}_j$,
  but $\mathbf{P}_j$ *need not be an orthogonal projector* when $\mathbf{A} \neq \mathbf{A}^*$.
- The spectral projectors give a *resolution of the identity*: $\displaystyle\sum_{j=1}^{\widehat{n}} \mathbf{P}_j = \mathbf{I}$.
- $\mathbf{P}_{\text{g}} := \mathbf{P}_1 + \cdots + \mathbf{P}_{\widehat{m}}, \qquad \mathcal{U}_{\text{g}} := \text{Range}(\mathbf{P}_{\text{g}}), \qquad m = \dim(\mathcal{U}_{\text{g}}).$
- $\mathbf{P}_{\text{b}} := \mathbf{I} - \mathbf{P}_{\text{g}}, \qquad \mathcal{U}_{\text{b}} := \text{Range}(\mathbf{P}_{\text{b}}), \qquad \dim(\mathcal{U}_{\text{b}}) = n - m.$

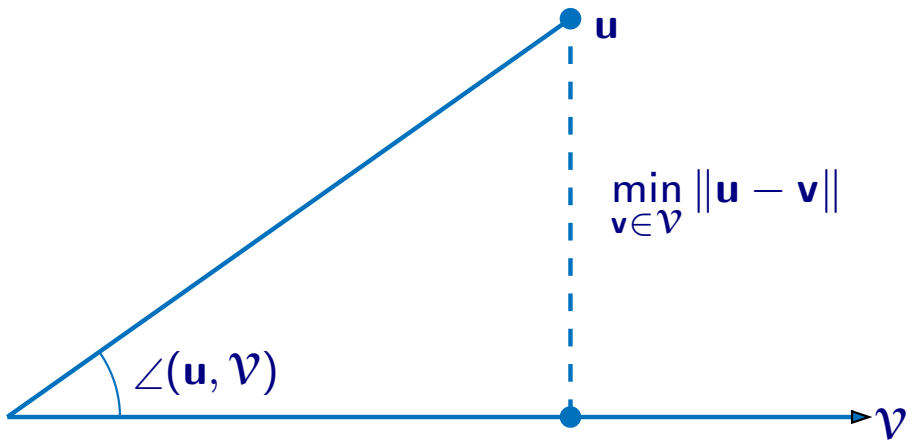
preliminaries: angles between subspaces

- $\mathcal{V}$ = approximating subspace.
  For our problems, $\mathcal{V} = \mathcal{K}_k(\mathbf{A}, \mathbf{x}) := \text{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}$.

- $\mathcal{U}_g$ = desired invariant subspace

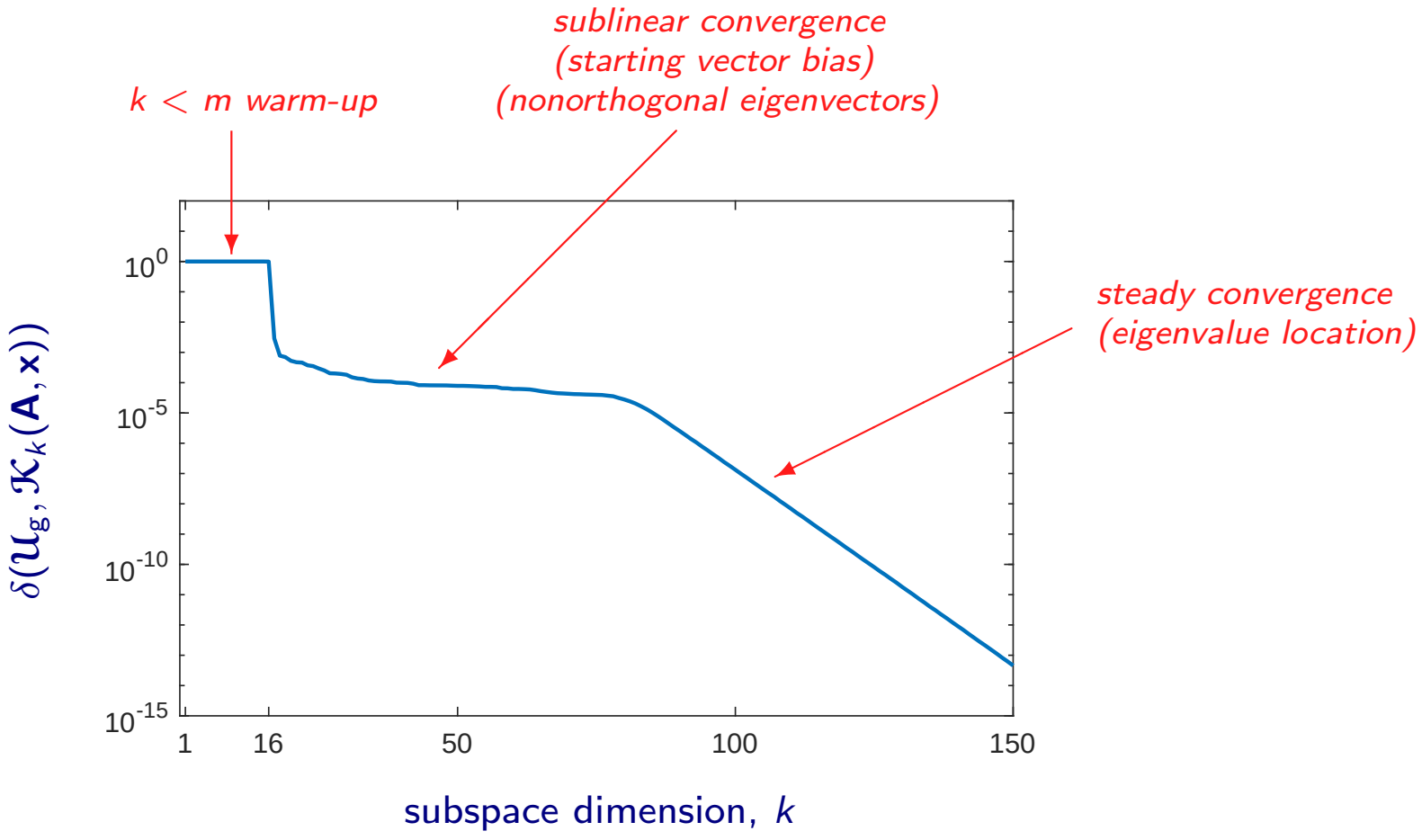
- Measure convergence via the *containment gap*:

$$\delta(\mathcal{U}_g, \mathcal{V}) = \max_{\mathbf{u} \in \mathcal{U}_g} \sin \angle(\mathbf{u}, \mathcal{V}) = \max_{\mathbf{u} \in \mathcal{U}_g} \min_{\mathbf{v} \in \mathcal{V}} \frac{\|\mathbf{u} - \mathbf{v}\|}{\|\mathbf{u}\|}.$$



- *We will monitor how $\delta(\mathcal{U}_g, \mathcal{K}_k(\mathbf{A}, \mathbf{x}))$ develops as k increases.*

example convergence behavior, $\mathbf{A} \neq \mathbf{A}^*$



cf. GMRES convergence model of [Nevanlinna 1993]

example adapted from [Beattie, E., Rossi 2004]

basic convergence model

Building on [Saad 1980, 1983], [Jia 1995], [Sorensen 2002], [Beattie, E., Rossi 2004], and others....

Theorem [Beattie, E., Sorensen 2005].

Suppose $\mathcal{U}_g$ is *reachable* from the Krylov space $\mathcal{K}_k(\mathbf{A}, \mathbf{x})$.
Then for $k \geq 2m$,

$$\delta(\mathcal{U}_g, \mathcal{K}_k(\mathbf{A}, \mathbf{x})) \;\leq\; C_1\, C_2 \min_{\phi \in P_{k-2m}} \max_{z \in \Omega_b} |1 - \alpha(z)\phi(z)|.$$

- $C_1 = C_1(\mathbf{A}, \mathbf{x}) =$ measure of starting vector bias.
- $C_2 = C_2(\mathbf{A}, \Omega_b) =$ measure of eigenvector departure from orthogonality.
- $P_{k-2m} =$ set of polynomials of degree $k - 2m$ or less.
- $\Omega_b \subset \mathbb{C}$ contains the undesired eigenvalues.
- $\alpha(z) = (z - \lambda_1) \cdots (z - \lambda_m)$.

Invariant Subspaces

reachable by

Krylov Subspaces

reachable invariant subspaces

- If $\mathbf{x} \in \mathbb{C}^n$ lacks a component in the desired eigenvector, e.g.,

$$\mathbf{P}_1 \mathbf{x} = \mathbf{0},$$

the desired eigenvalue/eigenvector is *invisible* to Krylov methods (in exact arithmetic). For example, in the power method

$$\mathbf{A}^p \mathbf{x} = \sum_{j=1}^{n} \lambda_j^p \mathbf{P}_j \mathbf{x} = \mathbf{0} + \sum_{j=2}^{n} \lambda_j^p \mathbf{P}_j \mathbf{x},$$

the eigenvalue λ_1 *has no influence*. (We will address this more later.)

- A different problem arises when $\mathbf{A}$ has *repeated eigenvalues* with *linearly indpendent eigenvectors* (derogatory eigenvalues).

A simple example: $\mathbf{A} = \mathbf{I}$ (identity matrix).

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

The Krylov method converges in one step (*happy breakdown*), exactly finding one copy of the eigenvalue $\lambda = 1$ and eigenvector $\mathbf{x}$.

reachable invariant subspaces

A more perplexing example from Chris Beattie [Beattie, E., Rossi 2004]:

$$\mathbf{A} = \begin{bmatrix} 1 & & & & \\ 1 & 1 & & & \\ & & 1 & & \\ & & 1 & 1 & \\ & & & 1 & 1 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} 1 \\ c \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

By taking c large, we bias $\mathbf{x}$ toward the eigenvector $[0, 1, 0, 0, 0]^T$.

For any c the Krylov method breaks down (happily) at iteration $k = 3$, discovering the Jordan block and invariant subspace

$$\mathbf{V}^*\mathbf{A}\mathbf{V} = \mathbf{S} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{S}^{-1}, \qquad \text{Range} \left(\begin{bmatrix} 1 & 0 & 0 \\ c & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \right).$$

The Krylov method finds the 3×3 Jordan block with eigenvector $[0, 0, 0, 0, 1]^T$.

Only a set of measure zero $\mathbf{x}$ *can discover* $[0, 1, 0, 0, 0]^T$.

reachable invariant subspaces

- Unlucky choices of $\mathbf{x} \in \mathbb{C}^n$ can (in principle) prevent the Krylov method from seeing a desired (simple) eigenvector.

  This behavior is fragile to numerical computations, since *infinitesimal perturbations to* $\mathbf{x}$ *will add a small component in the desired eigenvector*.

- Single-vector Krylov methods can (in principle) *find one Jordan block associated with each eigenvalue.*

  This behavior is fragile to numerical computations, since *infinitesimal perturbations split multiple eigenvalues*.

- Block Krylov methods (with block size b, $\mathbf{X} \in \mathbb{C}^{n \times b}$)

$$\mathcal{K}_k(\mathbf{A}, \mathbf{X}) = \text{Range}([\mathbf{X} \;\; \mathbf{AX} \;\; \mathbf{A}^2\mathbf{X} \;\; \cdots \;\; \mathbf{A}^{k-1}\mathbf{X}]),$$

  can find b linearly independent eigenvectors for a single eigenvalue.

How does the

starting vector x

affect convergence ?

effect of starting vector on convergence

Henceforth assume the desired invariant subspace $\mathcal{U}$ is *reachable* from $\mathbf{x}$:

$$\mathcal{U}_g \subset \mathcal{K}_n(\mathbf{A}, \mathbf{x}).$$

How does $\mathbf{x}$ *influence convergence?*

For a single eigenpair $(\lambda_1, \mathbf{u}_1)$ with spectral projector $\mathbf{P}_1$, Saad [1980] gives

$$\sin \angle(\mathbf{u}_1, \mathcal{K}_k(\mathbf{A}, \mathbf{x})) \leq \frac{1}{\|\mathbf{P}_1 \mathbf{x}\|} \min_{\substack{\phi \in P_{k-1} \\ \phi(\lambda_1)=1}} \|(\mathbf{I} - \mathbf{P}_1)\psi(\mathbf{A})\|,$$

The leading constant grows as the orientation of $\mathbf{x}$ *toward* $\mathbf{u}_1$ *diminishes.*

effect of starting vector on convergence

Henceforth assume the desired invariant subspace $\mathcal{U}$ is *reachable* from $\mathbf{x}$:

$$\mathcal{U}_g \subset \mathcal{K}_n(\mathbf{A}, \mathbf{x}).$$

How does $\mathbf{x}$ *influence convergence?*

For a single eigenpair $(\lambda_1, \mathbf{u}_1)$ with spectral projector $\mathbf{P}_1$, Saad [1980] gives

$$\sin \angle(\mathbf{u}_1, \mathcal{K}_k(\mathbf{A}, \mathbf{x})) \leq \frac{1}{\|\mathbf{P}_1 \mathbf{x}\|} \min_{\substack{\phi \in P_{k-1} \\ \phi(\lambda_1)=1}} \|(\mathbf{I} - \mathbf{P}_1)\psi(\mathbf{A})\|,$$

The leading constant grows as the orientation of $\mathbf{x}$ *toward* $\mathbf{u}_1$ *diminishes.*

For m-dimensional invariant subspaces $\mathcal{U}_g$, our bounds replace $1/\|\mathbf{P}_1 \mathbf{x}\|$ with

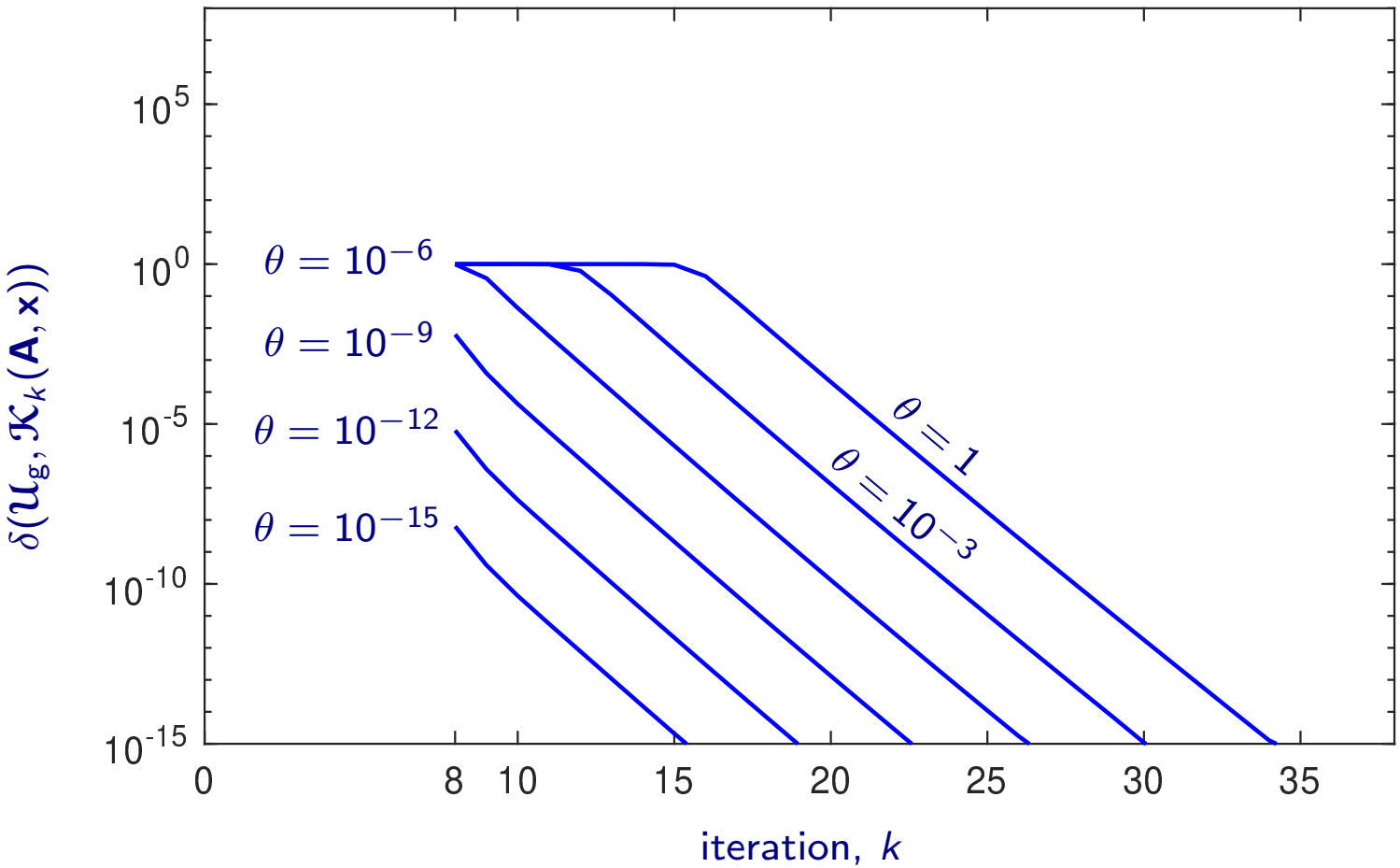
$$C_1 := \max_{\psi \in P_{m-1}} \frac{\|\psi(\mathbf{A})\mathbf{P}_b \mathbf{x}\|}{\|\psi(\mathbf{A})\mathbf{P}_g \mathbf{x}\|} = \max_{\mathbf{v} \in \mathcal{K}_m(\mathbf{A}, \mathbf{x})} \frac{\|\mathbf{P}_b \mathbf{v}\|}{\|\mathbf{P}_g \mathbf{v}\|},$$

the ratio of the bad to the good component in the *worst approximation to* $\mathcal{U}_g$ *from the m-dimensional Krylov space.*

effect of starting vector on convergence

$\mathbf{A}$ = symmetric matrix ($n = 128$) with
8 desired eigenvalues in $[1, 2]$;
120 undesired eigenvalues in $[-1, 0]$.

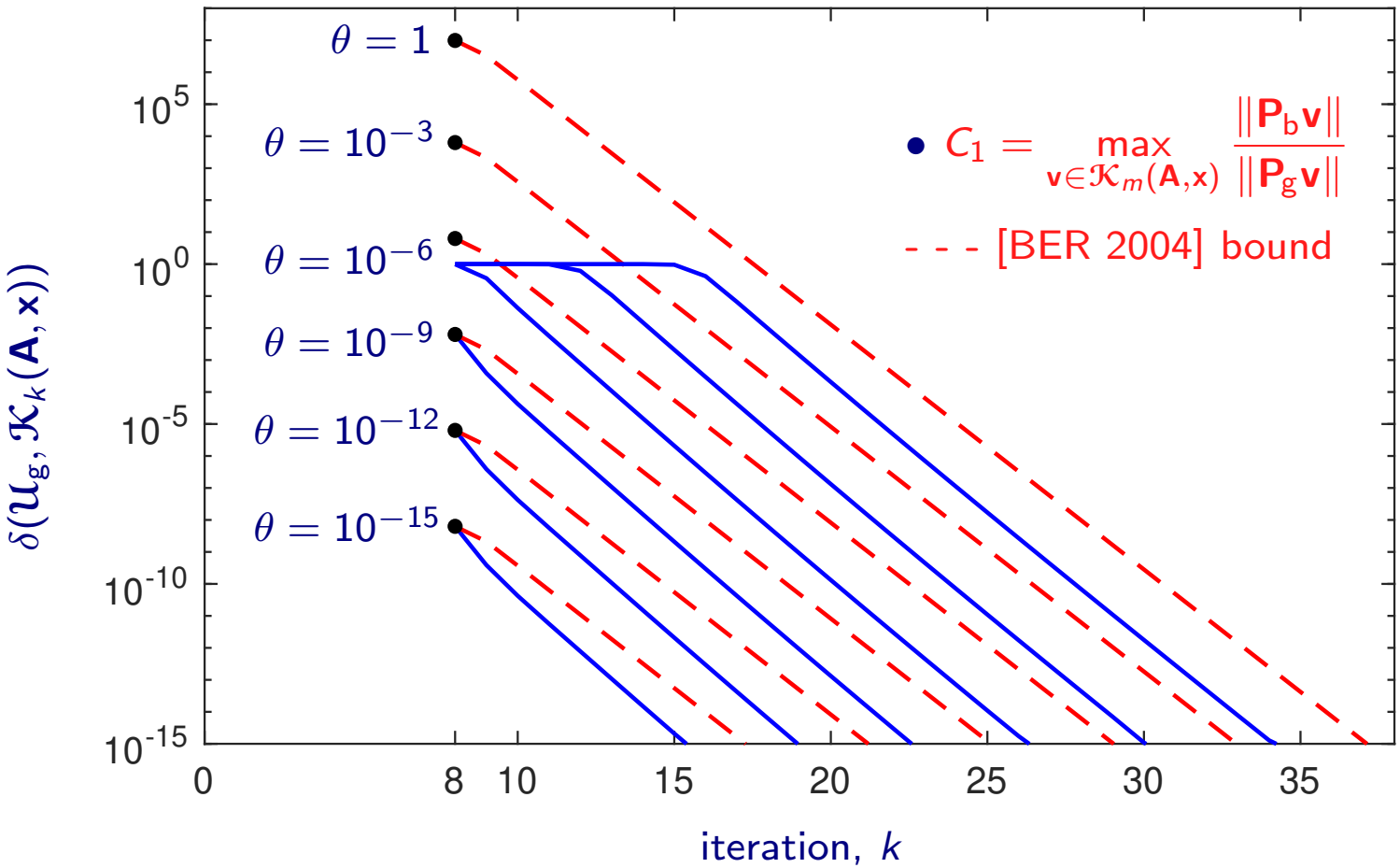
$\theta = \angle(\mathbf{x}, \mathcal{U}_g)$, the angle between $\mathbf{x}$ and its best approximation in $\mathcal{U}_g$.



effect of starting vector on convergence

$\mathbf{A}$ = symmetric matrix ($n = 128$) with
8 desired eigenvalues in $[1, 2]$;
120 undesired eigenvalues in $[-1, 0]$.

$\theta = \angle(\mathbf{x}, \mathcal{U}_g)$, the angle between $\mathbf{x}$ and its best approximation in $\mathcal{U}_g$.



How do the

eigenvalues of A

affect convergence ?

asymptotic convergence rate determined by eigenvalues

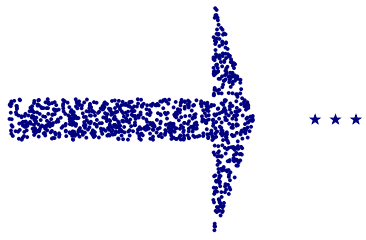
$$\min_{\phi \in P_{k-2m}} \max_{z \in \Omega_b} |1 - \alpha(z)\phi(z)|$$

- P_{k-2m} = set of polynomials of degree $k - 2m$ or less.
- $\Omega_b \subset \mathbb{C}$ contains the undesired eigenvalues.
- $\alpha(z) = (z - \lambda_1) \cdots (z - \lambda_m)$.

The polynomial approximation problem gives convergence like $C\,\gamma^k$ for some constant C and rate γ.

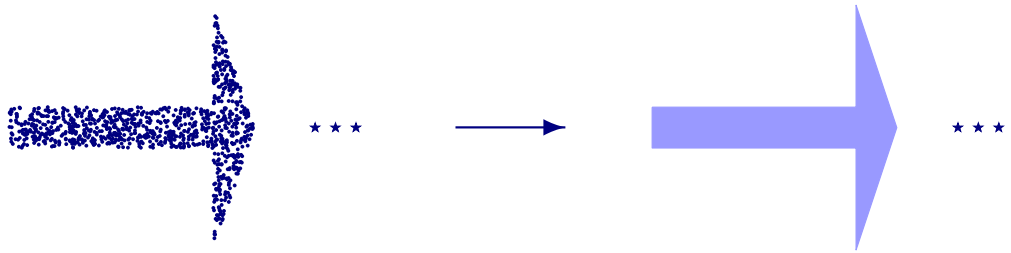
- When $\mathbf{A} = \mathbf{A}^*$, $\Omega_b = [\lambda_{m+1}, \lambda_n]$, and use Chebyshev polynomials to compute the convergence rate γ.
- When Ω_b is a simply connected open subset of $\mathbb{C}$, use conformal mapping to approach the approximation problem.

potential theoretic determination of the convergence rate



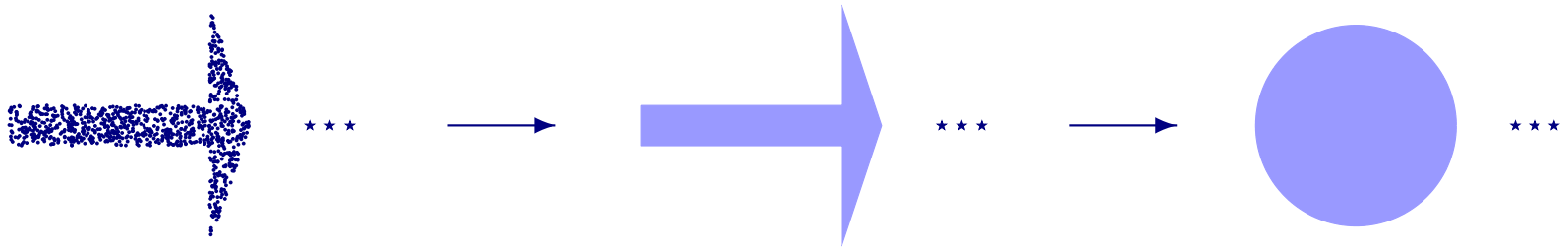
Step 1: *Begin by identifying undesired* $(\cdot)$ *and desired* $(\star)$ *eigenvalues.*

potential theoretic determination of the convergence rate



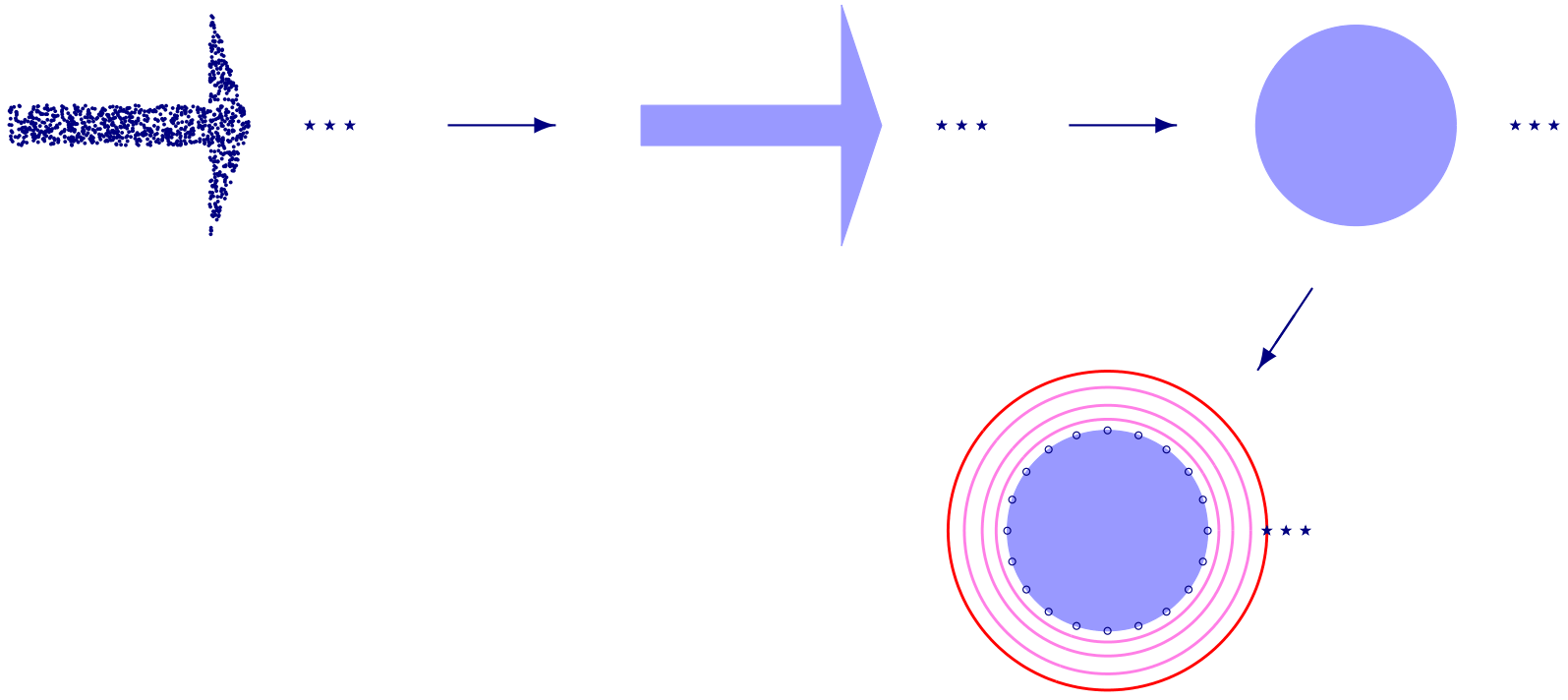
Step 2: *Bound bad eigenvalues with* Ω_{b}.

potential theoretic determination of the convergence rate



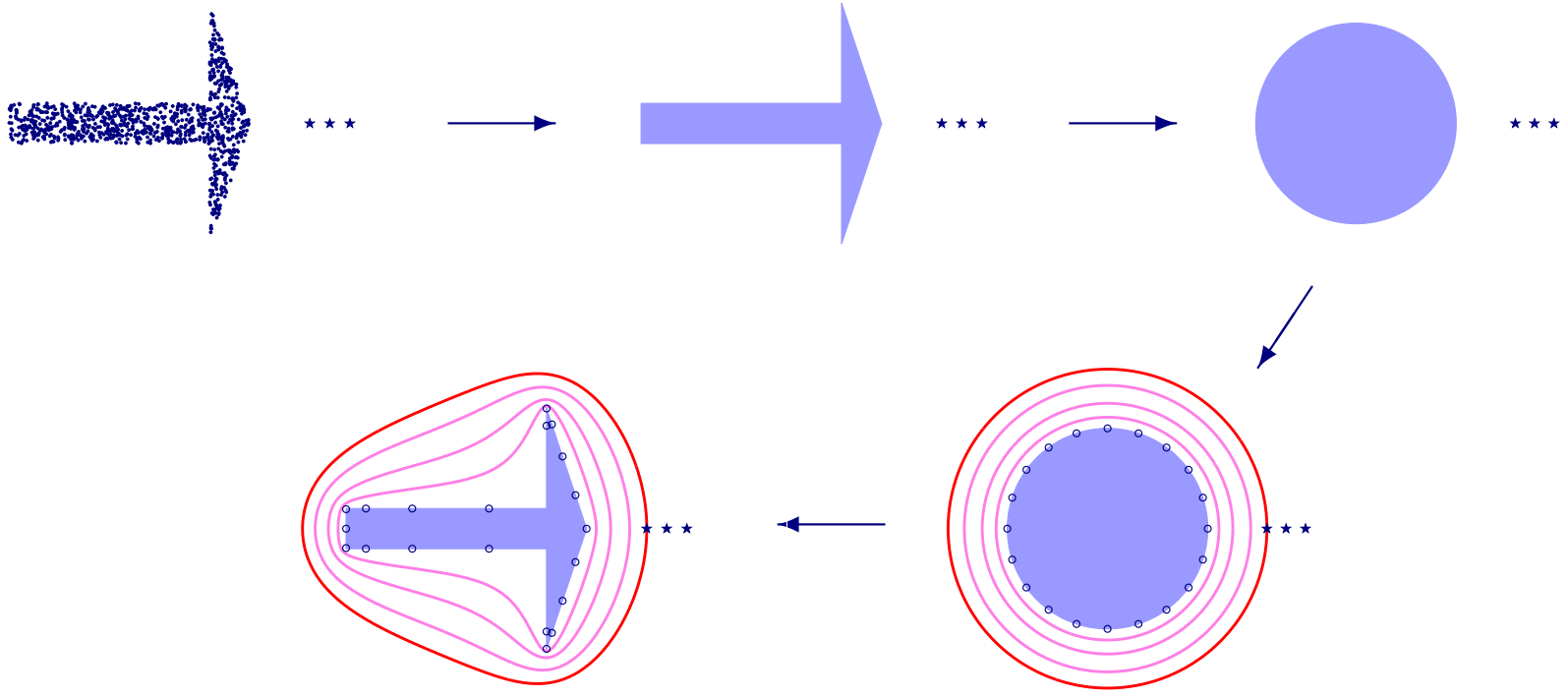
Step 3: *Conformally map* $\mathbb{C} \setminus \Omega_b$ *to the exterior of the unit disk.*

potential theoretic determination of the convergence rate



Step 4: *Find the lowest level curve of the Green's function intersecting a good eigenvalues.*

potential theoretic determination of the convergence rate



Step 5: *Invert map to obtain curves of constant convergence rate in the original domain.*

Black circles on final figure are Fejér points, asymptotically optimal interpolation points for ϕ.

convergence rate: and granularity of the spectrum

If convergence is very slow, perhaps you are solving the wrong problem.

convergence rate: and granularity of the spectrum

If convergence is very slow, perhaps you are solving the wrong problem.

Consider $m = 1$, where we can use the elementary bound [Saad 1980]

$$\sin \angle(\mathbf{u}_1, \mathcal{K}_k(\mathbf{A}, \mathbf{x})) \leq \frac{1}{\|\mathbf{P}_1 \mathbf{x}\|} \min_{\substack{\phi \in P_{k-1} \\ \phi(\lambda_1)=1}} \|(\mathbf{I} - \mathbf{P}_1)\psi(\mathbf{A})\|.$$

Suppose $\mathbf{A} = \mathbf{A}^*$ and we seek leftmost eigenvalue λ_1, where

$$\lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n.$$

The error bound suggests the progress made at each iteration is like

$$\gamma := \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}, \qquad \text{where } \kappa := \frac{\lambda_n - \lambda_1}{\lambda_2 - \lambda_1}.$$

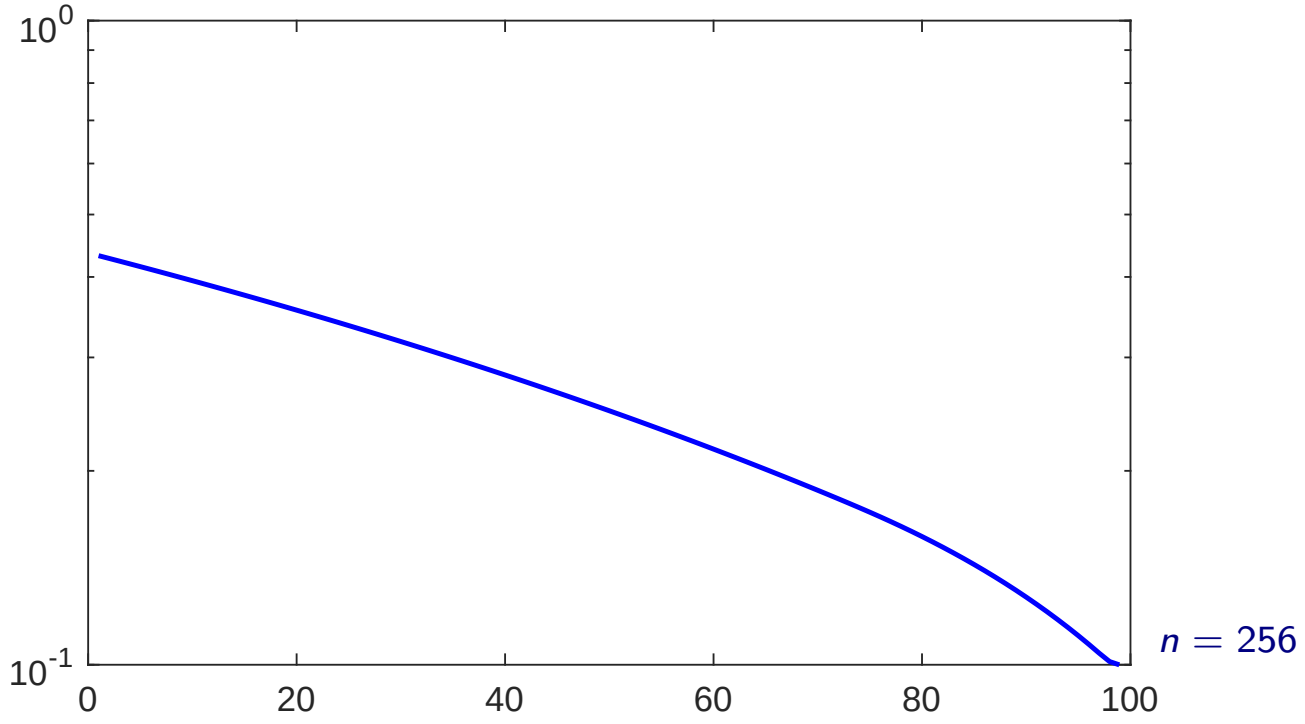
When $\mathbf{A}$ is a discretization of an *unbounded operator*, we expect $\lambda_n = \|\mathbf{A}\| \to \infty$ as $n \to \infty$. *The convergence rate goes to one as $n \to \infty$.*

Thus Krylov subspace methods often perform poorly for PDE eigenvalue problems — *unless the set-up is modified*.

convergence slows as the discretization improves

Apply the Krylov method to discretizations of the Laplacian in one dimension.

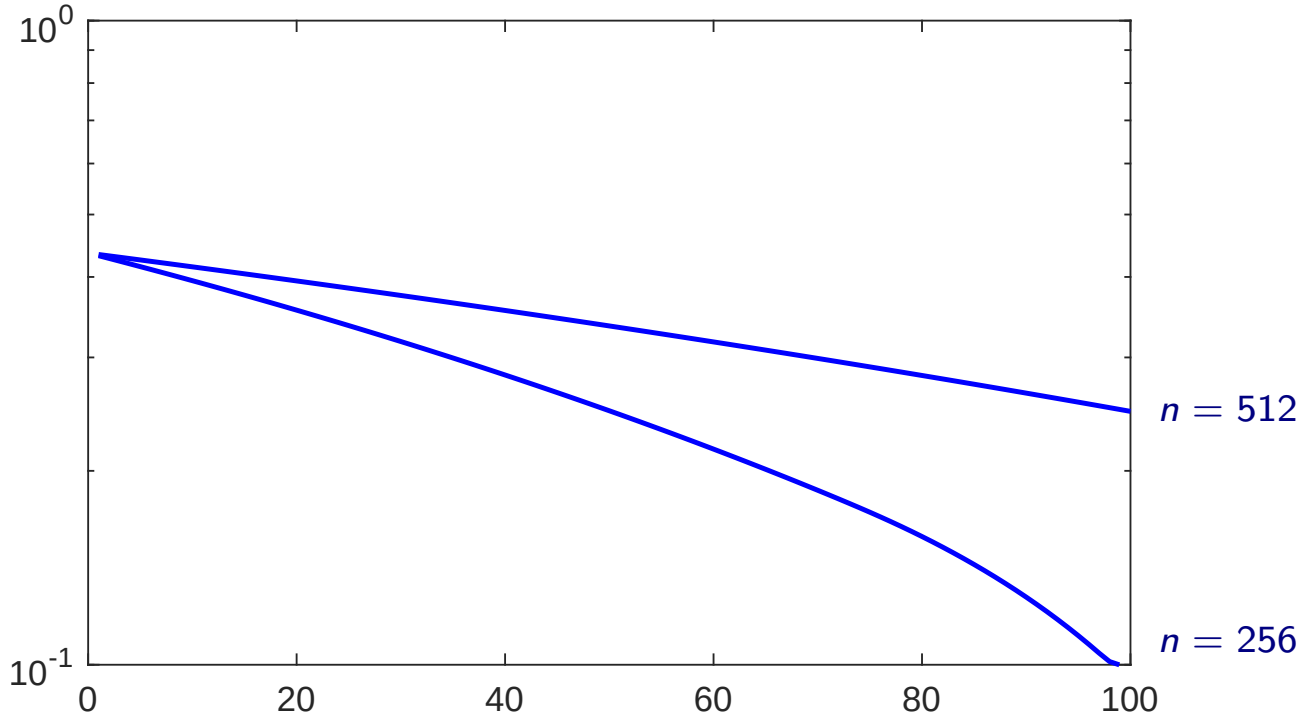
How does the convergence rate change as the discretization improves?



convergence slows as the discretization improves

Apply the Krylov method to discretizations of the Laplacian in one dimension.

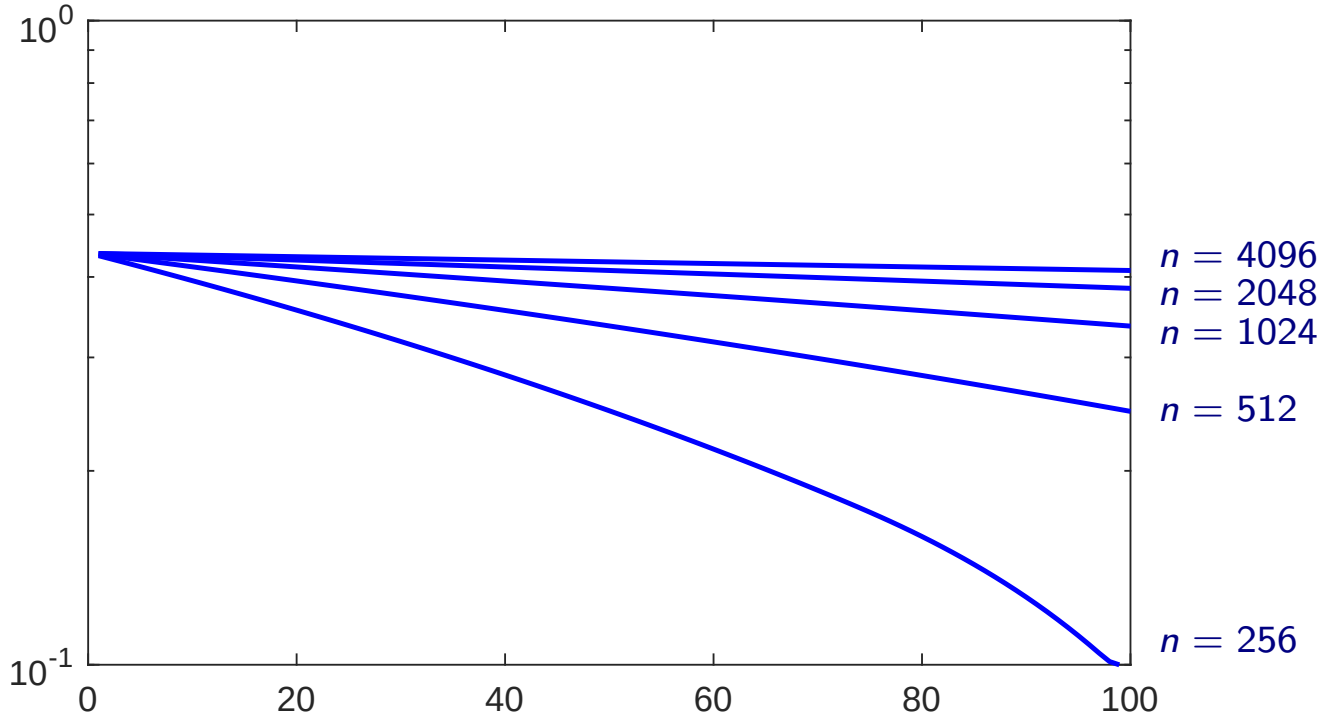
How does the convergence rate change as the discretization improves?



convergence slows as the discretization improves

Apply the Krylov method to discretizations of the Laplacian in one dimension.

How does the convergence rate change as the discretization improves?



Convergence of Krylov Subspace Projection

The problem becomes immediately apparent if we attempt to run Krylov subspace projection *on the operator itself*,

$$\mathcal{K}_k(L, f) = \text{span}\{f, Lf, \ldots, L^{k-1}f\}.$$

For $Lu = -u''$ with Dirichlet boundary conditions, $u(0) = u(1) = 1$, take some starting vector $f \in \text{Dom}(L)$, i.e.,

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

In general $Lf \notin \text{Dom(L)}$, so we cannot build the next Krylov direction $L^2 f = L(Lf)$.

The Krylov algorithm breaks down at the third step.

Convergence of Krylov Subspace Projection

The problem becomes immediately apparent if we attempt to run Krylov subspace projection *on the operator itself*,

$$\mathcal{K}_k(L, f) = \text{span}\{f, Lf, \ldots, L^{k-1}f\}.$$

For $Lu = -u''$ with Dirichlet boundary conditions, $u(0) = u(1) = 1$, take some starting vector $f \in \text{Dom}(L)$, i.e.,

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

In general $Lf \notin \text{Dom(L)}$, so we cannot build the next Krylov direction $L^2 f = L(Lf)$.

The Krylov algorithm breaks down at the third step.

The operator setting suggests that we instead apply Krylov to L^{-1}:

$$\mathcal{K}_k(L^{-1}, f) = \text{span}\{f, L^{-1}f, \ldots, L^{-(k-1)}f\}.$$

In this case, L^{-1} is a beautiful compact operator:

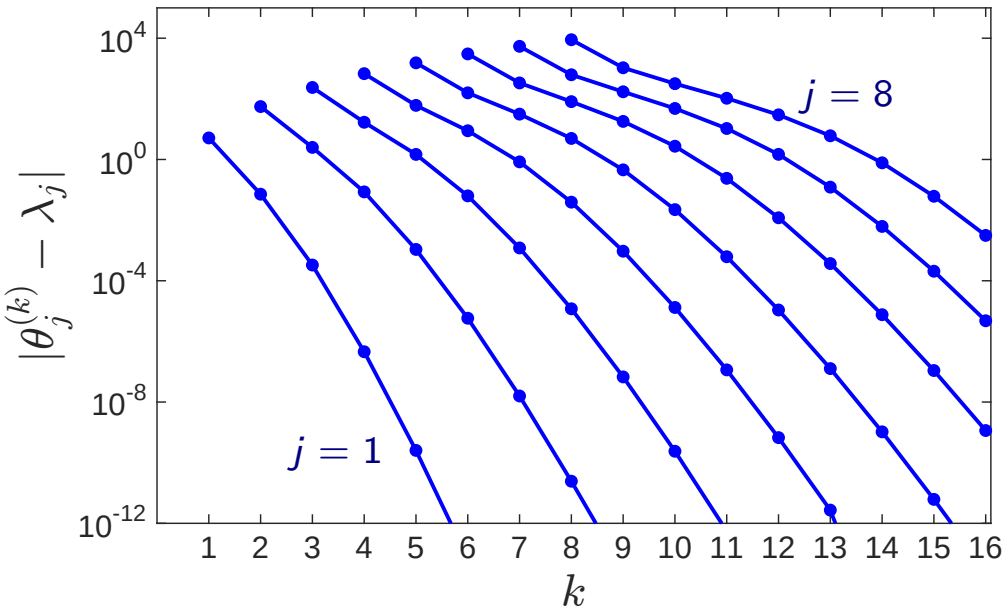
$$(L^{-1}f)(x) = \iint f + C_0 + C_1 x,$$

where we choose C_0 and C_1 so that

$$(L^{-1}f)(0) = (L^{-1}f)(1) = 0.$$

krylov projection applied to the operator

We run the Krylov method on L^{-1} exactly in Mathematica.
Denote the eigenvalue estimates at the kth iteration as $\theta_1^{(k)} \leq \theta_2^{(k)} \leq \cdots \leq \theta_k^{(k)}$.

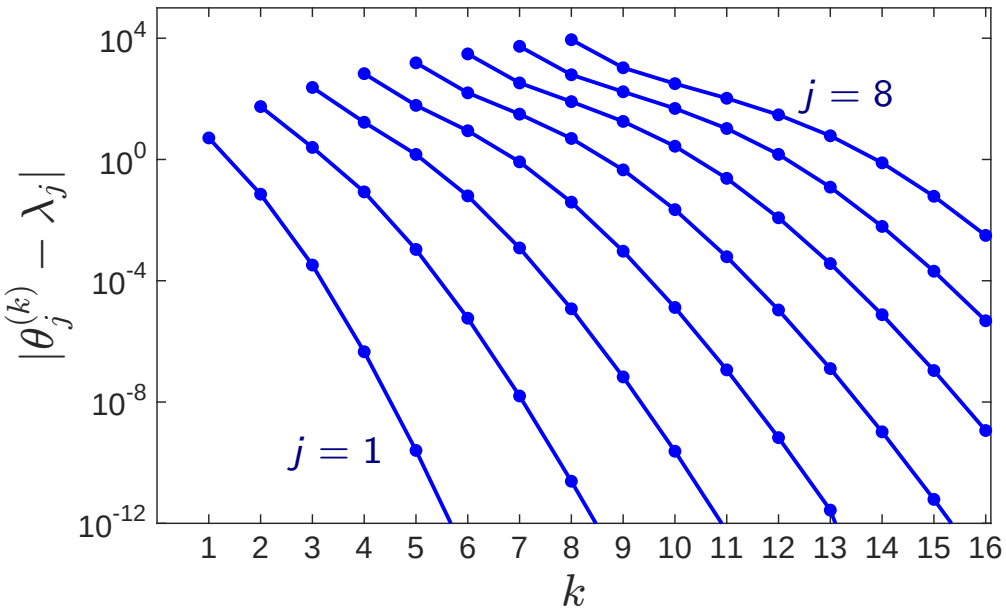


Observe *superlinear convergence as k increases.*

For CG and GMRES applied to operators, see [Winther 1980], [Nevanlinna 1993], [Moret 1997], [Olver 2009], [Kirby 2010]. For "superlinear" convergence in finite dimensional settings, see [van der Sluis, van der Vorst, 1986], [van der Vorst, Vuik, 1992], [Beattie, E., Rossi 2004], [Simoncini, Szyld 2005].

krylov projection applied to the operator

We run the Krylov method on L^{-1} exactly in Mathematica.
Denote the eigenvalue estimates at the kth iteration as $\theta_1^{(k)} \leq \theta_2^{(k)} \leq \cdots \leq \theta_k^{(k)}$.



Observe *superlinear convergence as k increases.*

For CG and GMRES applied to operators, see [Winther 1980], [Nevanlinna 1993], [Moret 1997], [Olver 2009], [Kirby 2010]. For "superlinear" convergence in finite dimensional settings, see [van der Sluis, van der Vorst, 1986], [van der Vorst, Vuik, 1992], [Beattie, E., Rossi 2004], [Simoncini, Szyld 2005].

This mode of computation is preferred for discretization matrices as well:
the *shift-invert Arnoldi method* uses $\mathcal{K}_k((\mathbf{A} - \mu\mathbf{I})^{-1}, \mathbf{x})$.

polynomial preconditioning: a cheap spectral transformation

Replace the conventional Krylov space

$$\mathcal{K}_k(\mathbf{A}, \mathbf{x}) = \text{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \mathbf{A}^2\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}$$

with the *polynomial preconditioned transformation*

$$\mathcal{K}_k(\pi(\mathbf{A}), \mathbf{x}) = \text{span}\{\mathbf{x}, \pi(\mathbf{A})\mathbf{x}, \pi(\mathbf{A})^2\mathbf{x}, \ldots, \pi(\mathbf{A})^{k-1}\mathbf{x}\}.$$

[Thornquist 2006], [E., Loe, Morgan arXiv:1806.08020]

Use the polynomial π to separate the interesting eigenvalues.

Often increases matvecs, but decreases iterations (hence orthogonalization).

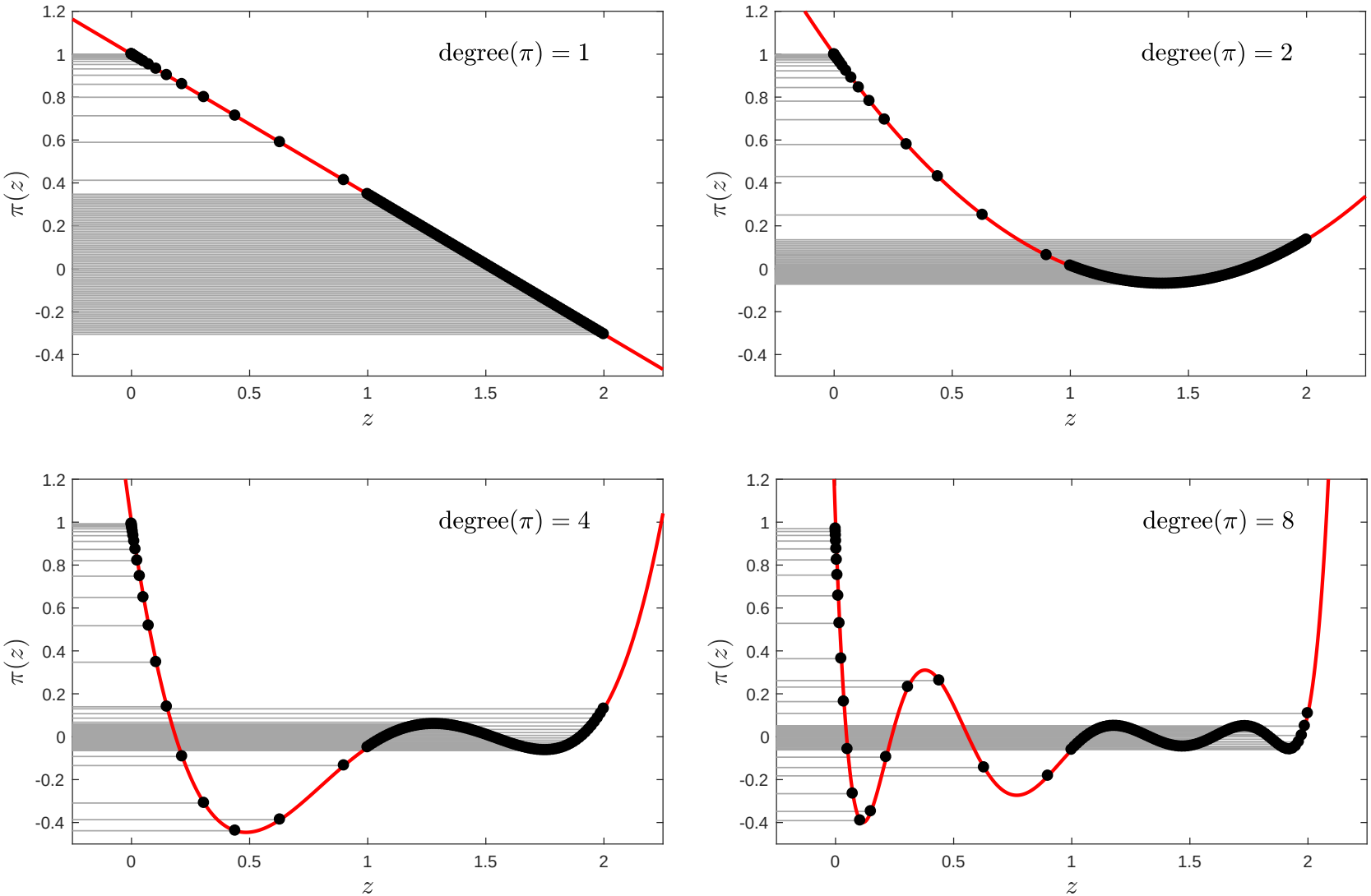
For example, with Hermitian $\mathbf{A}$, take π to be the degree-d MINRES residual polynomial [Paige & Saunders 1975], which attains

$$\min_{\substack{p \in P_d \\ p(0)=1}} \|p(\mathbf{A})\mathbf{b}\|.$$

This polynomial tends to separate smallest-magnitude eigenvalues.

polynomial preconditioning: a cheap spectral transformation

Polynomial preconditioning: Hermitian **A**, MINRES polynomial.



horizontal axis: $\sigma(\mathbf{A})$ *vertical axis* $\sigma(\pi(\mathbf{A}))$

How do the

eigenvectors of A

affect convergence ?

bounding functions of a matrix

If $\mathbf{A}$ is *normal* ($\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$), *eigenvectors are orthogonal.* For f analytic on $\sigma(\mathbf{A})$,

$$\|f(\mathbf{A})\| = \max_{\lambda \in \sigma(\mathbf{A})} |f(\lambda)|.$$

For nonnormal $\mathbf{A}$*, the situation is considerably more complicated.*

- If $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$, then

$$\|f(\mathbf{A})\| \leq \|\mathbf{U}\| \, \|\mathbf{U}^{-1}\| \max_{\lambda \in \sigma(\mathbf{A})} |f(\lambda)|.$$

- For the *numerical range* (field of values) $W(\mathbf{A}) = \{\mathbf{v}^*\mathbf{A}\mathbf{v} : \|\mathbf{v}\| = 1\}$,

$$\|f(\mathbf{A})\| \leq \left(1 + \sqrt{2}\right) \max_{z \in W(\mathbf{A})} |f(z)|.$$

- For the *ε-pseudospectrum* $\sigma_\varepsilon(\mathbf{A}) = \{z \in \sigma(\mathbf{A} + \mathbf{E}) \text{ for some } \|\mathbf{E}\| < \varepsilon\}$,

$$\|f(\mathbf{A})\| \leq \frac{L_\varepsilon}{2\pi\varepsilon} \max_{z \in \sigma_\varepsilon(\mathbf{A})} |f(z)|,$$

  where L_ε is the boundary length of $\sigma_\varepsilon(\mathbf{A})$.

constant to account for nonnormality

$C_2 = C_2(\mathbf{A}, \Omega_b)$ comes from bounding $\|f(\mathbf{A}|_{\mathcal{U}_b})\|$, $f(z) = 1 - \alpha(z)\phi(z)$.

Theorem [Beattie, E., Sorensen 2005].
Suppose $\mathcal{U}_g$ is *reachable* from the Krylov space $\mathcal{K}_k(\mathbf{A}, \mathbf{x})$.
Then for $k \geq 2m$,

$$\delta(\mathcal{U}_g, \mathcal{K}_k(\mathbf{A}, \mathbf{x})) \;\leq\; C_1\, C_2 \min_{\phi \in P_{k-2m}} \max_{z \in \Omega_b} |1 - \alpha(z)\phi(z)|.$$

- If $\Omega_b = \sigma(\mathbf{A}|_{\mathcal{U}_b})$ (no defective eigenvalues), then $C_2 = \|\mathbf{U}_b\|\,\|\mathbf{U}_b^+\|$, where the columns of $\mathbf{U}_b \in \mathbb{C}^{n\times(n-m)}$ are eigenvectors of $\mathbf{A}|_{\mathcal{U}_b}$.
- If $\Omega_b = W(\mathbf{A}|_{\mathcal{U}_b})$ then $C_2 = 1 + \sqrt{2}$.
- If $\Omega_b = \sigma_\varepsilon(\mathbf{A}|_{\mathcal{U}_b})$ then $C_2 = L_\varepsilon/(2\pi\varepsilon)$.

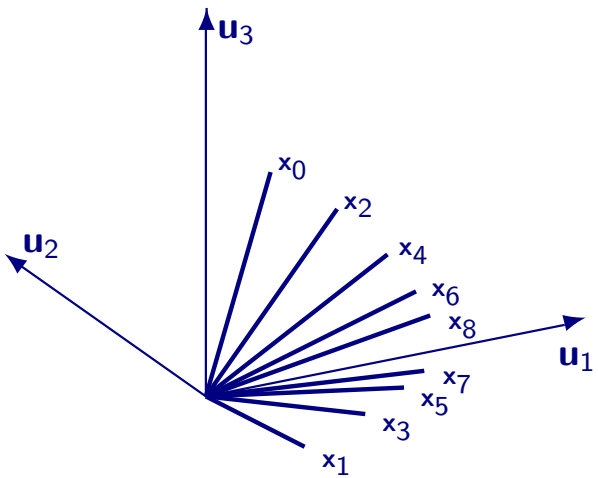
Tension: balance $C_2 \geq 1$ verses size of Ω_b.

transient behavior of the power method

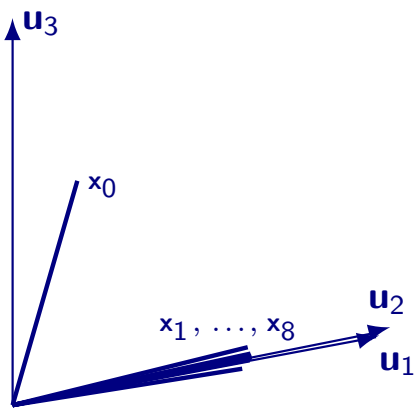
Large coefficients in the expansion of $\mathbf{x}_0$ in the eigenvector basis can lead to cancellation effects in $\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0$.

Example: here different choices of α and β affect eigenvalue conditioning,

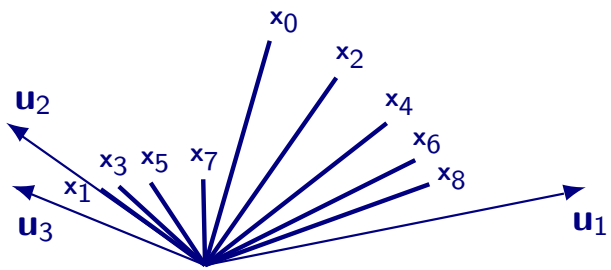
$$\mathbf{A} = \begin{bmatrix} 1 & \alpha & 0 \\ 0 & 3/4 & \beta \\ 0 & 0 & -3/4 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -4\alpha \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 8\alpha\beta/21 \\ -2\beta/3 \\ 1 \end{bmatrix}.$$



$\alpha = \beta = 0$



$\alpha = 10,\ \beta = 0$



$\alpha = 0,\ \beta = 10$

[Trefethen & E. 2005]

restarting

krylov suspace

methods

an essential tool for controlling subspace dimension

restarting krylov methods

restarted Arnoldi algorithm (`eigs`)

To compute $m < k$ eigenvalues of $\mathbf{A} \in \mathbb{C}^{n\times n}$, Arnoldi methods restrict $\mathbf{A}$ to act on the k-dimensional Krylov subspace

$$\mathrm{Ran}(\mathbf{V}) = \mathrm{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}.$$

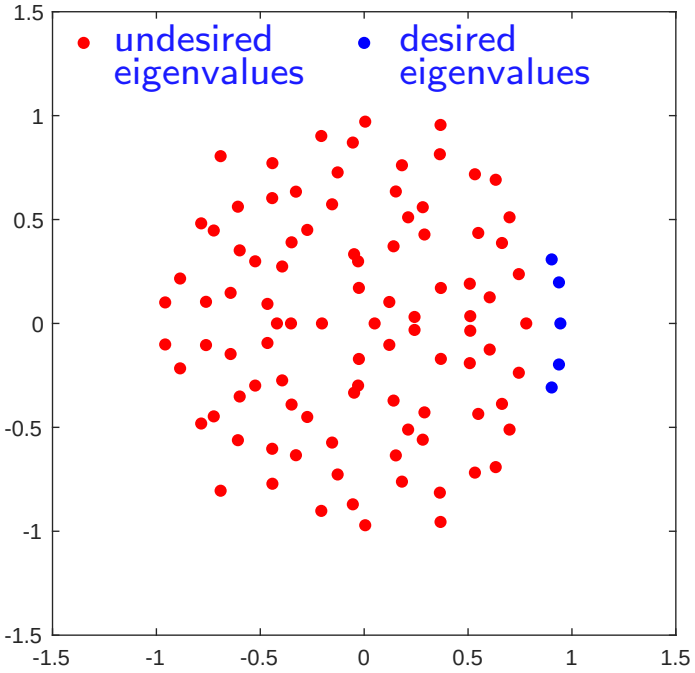
Compute eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ (*Ritz values*), and order them by relevance; e.g., if seeking the rightmost eigenvalue of $\mathbf{A}$, let

$$\mathrm{Re}\,\theta_1 \geq \mathrm{Re}\,\theta_2 \geq \cdots \geq \mathrm{Re}\,\theta_k.$$

Exact shifts [Sorensen 1992] *restart* the method, attempting to improve $\mathbf{v}$ with a *polynomial filter* having the "unwanted" Ritz values as roots:

$$\mathbf{x}_+ = (\mathbf{A} - \theta_{m+1}\mathbf{I}) \cdots (\mathbf{A} - \theta_k \mathbf{I})\mathbf{x}.$$

To understand convergence, one must understand how the Ritz values are distributed over $\sigma(\mathbf{A})$.



restarting krylov methods

restarted Arnoldi algorithm (`eigs`)

To compute $m < k$ eigenvalues of $\mathbf{A} \in \mathbb{C}^{n\times n}$, Arnoldi methods restrict $\mathbf{A}$ to act on the k-dimensional Krylov subspace

$$\mathrm{Ran}(\mathbf{V}) = \mathrm{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}.$$

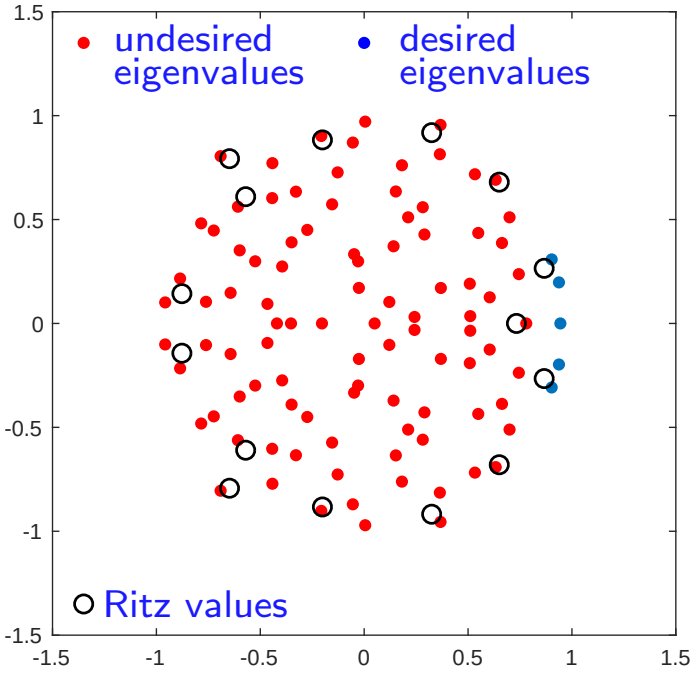
Compute eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ (*Ritz values*), and order them by relevance; e.g., if seeking the rightmost eigenvalue of $\mathbf{A}$, let

$$\mathrm{Re}\,\theta_1 \geq \mathrm{Re}\,\theta_2 \geq \cdots \geq \mathrm{Re}\,\theta_k.$$

Exact shifts [Sorensen 1992] *restart* the method, attempting to improve $\mathbf{v}$ with a *polynomial filter* having the "unwanted" Ritz values as roots:

$$\mathbf{x}_+ = (\mathbf{A} - \theta_{m+1}\mathbf{I}) \cdots (\mathbf{A} - \theta_k \mathbf{I})\mathbf{x}.$$

To understand convergence, one must understand how the Ritz values are distributed over $\sigma(\mathbf{A})$.



restarting krylov methods

restarted Arnoldi algorithm (`eigs`)

To compute $m < k$ eigenvalues of $\mathbf{A} \in \mathbb{C}^{n\times n}$, Arnoldi methods restrict $\mathbf{A}$ to act on the k-dimensional Krylov subspace

$$\mathrm{Ran}(\mathbf{V}) = \mathrm{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}.$$

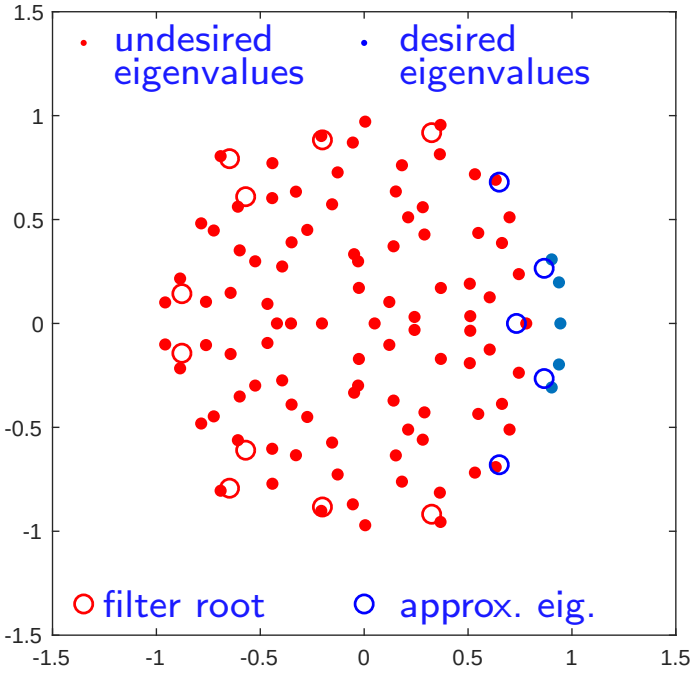
Compute eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ (*Ritz values*), and order them by relevance; e.g., if seeking the rightmost eigenvalue of $\mathbf{A}$, let

$$\mathrm{Re}\,\theta_1 \geq \mathrm{Re}\,\theta_2 \geq \cdots \geq \mathrm{Re}\,\theta_k.$$

Exact shifts [Sorensen 1992] *restart* the method, attempting to improve $\mathbf{v}$ with a *polynomial filter* having the "unwanted" Ritz values as roots:

$$\mathbf{x}_+ = (\mathbf{A} - \theta_{m+1}\mathbf{I}) \cdots (\mathbf{A} - \theta_k \mathbf{I})\mathbf{x}.$$

To understand convergence, one must understand how the Ritz values are distributed over $\sigma(\mathbf{A})$.



restarting krylov methods

restarted Arnoldi algorithm (`eigs`)

To compute $m < k$ eigenvalues of $\mathbf{A} \in \mathbb{C}^{n\times n}$, Arnoldi methods restrict $\mathbf{A}$ to act on the k-dimensional Krylov subspace

$$\mathrm{Ran}(\mathbf{V}) = \mathrm{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}.$$

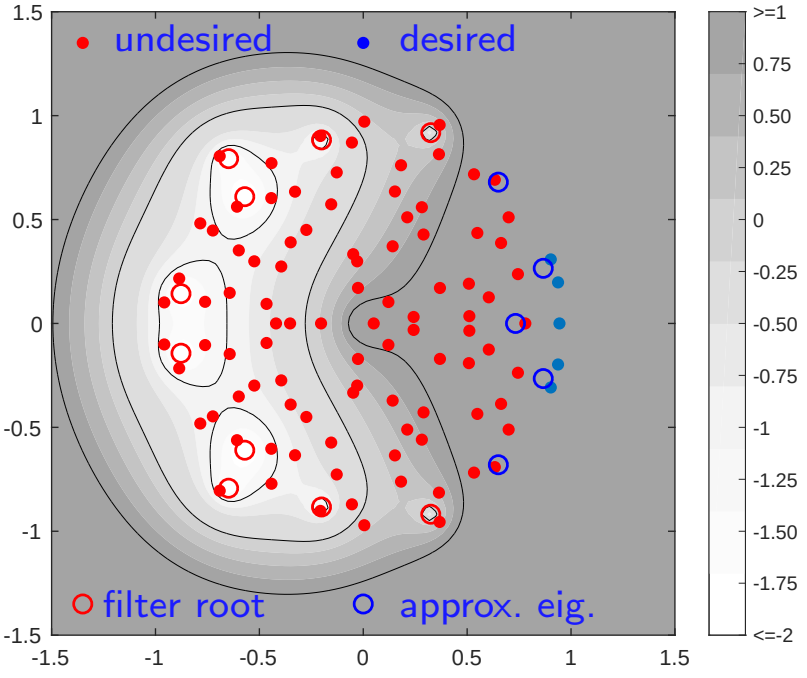
Compute eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ (*Ritz values*), and order them by relevance; e.g., if seeking the rightmost eigenvalue of $\mathbf{A}$, let

$$\mathrm{Re}\,\theta_1 \geq \mathrm{Re}\,\theta_2 \geq \cdots \geq \mathrm{Re}\,\theta_k.$$

Exact shifts [Sorensen 1992] *restart* the method, attempting to improve $\mathbf{v}$ with a *polynomial filter* having the "unwanted" Ritz values as roots:

$$\mathbf{x}_+ = (\mathbf{A} - \theta_{m+1}\mathbf{I}) \cdots (\mathbf{A} - \theta_k \mathbf{I})\mathbf{x}.$$

To understand convergence, one must understand how the Ritz values are distributed over $\sigma(\mathbf{A})$.



Shade $\sim \log_{10}$(*magnitude of filter polynomial*)

restarting krylov methods

restarted Arnoldi algorithm (`eigs`)

To compute $m < k$ eigenvalues of $\mathbf{A} \in \mathbb{C}^{n\times n}$, Arnoldi methods restrict $\mathbf{A}$ to act on the k-dimensional Krylov subspace

$$\mathrm{Ran}(\mathbf{V}) = \mathrm{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}.$$

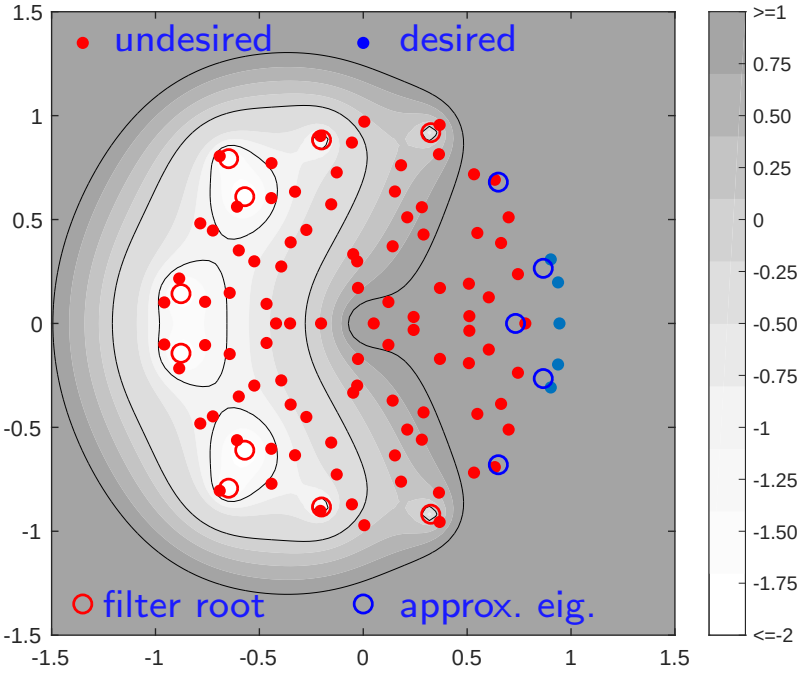
Compute eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ (*Ritz values*), and order them by relevance; e.g., if seeking the rightmost eigenvalue of $\mathbf{A}$, let

$$\mathrm{Re}\,\theta_1 \geq \mathrm{Re}\,\theta_2 \geq \cdots \geq \mathrm{Re}\,\theta_k.$$

Exact shifts [Sorensen 1992] *restart* the method, attempting to improve $\mathbf{v}$ with a *polynomial filter* having the "unwanted" Ritz values as roots:

$$\mathbf{x}_+ = (\mathbf{A} - \theta_{m+1}\mathbf{I}) \cdots (\mathbf{A} - \theta_k\mathbf{I})\mathbf{x}.$$

To understand convergence, one must understand how the Ritz values are distributed over $\sigma(\mathbf{A})$.



Shade $\sim \log_{10}$(*magnitude of filter polynomial*)

Pushing the language Haim Avron used the previous talk, a standard Krylov method (fixed k) is a sketch-and-solve method, while restarted Krylov methods sketch-to-precondition.

restarting krylov methods

restarted Arnoldi algorithm (`eigs`)

To compute $m < k$ eigenvalues of $\mathbf{A} \in \mathbb{C}^{n \times n}$, Arnoldi methods restrict $\mathbf{A}$ to act on the k-dimensional Krylov subspace

$$\mathrm{Ran}(\mathbf{V}) = \mathrm{span}\{\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{k-1}\mathbf{x}\}.$$

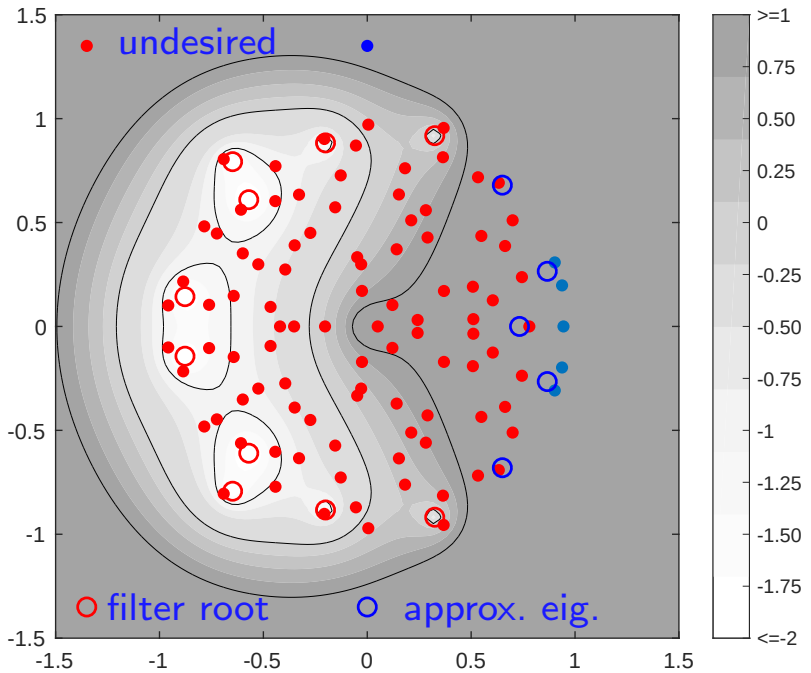
Compute eigenvalues of $\mathbf{V}^*\mathbf{A}\mathbf{V}$ (*Ritz values*), and order them by relevance; e.g., if seeking the rightmost eigenvalue of $\mathbf{A}$, let

$$\mathrm{Re}\,\theta_1 \geq \mathrm{Re}\,\theta_2 \geq \cdots \geq \mathrm{Re}\,\theta_k.$$

Exact shifts [Sorensen 1992] *restart* the method, attempting to improve $\mathbf{v}$ with a *polynomial filter* having the "unwanted" Ritz values as roots:

$$\mathbf{x}_+ = (\mathbf{A} - \theta_{m+1}\mathbf{I}) \cdots (\mathbf{A} - \theta_k \mathbf{I})\mathbf{x}.$$

To understand convergence, one must understand how the Ritz values are distributed over $\sigma(\mathbf{A})$.



Shade $\sim \log_{10}$(*magnitude of filter polynomial*)

- [Sorensen 1992] proved convergence for $\mathbf{A} = \mathbf{A}^*$.
- The process fails for some $\mathbf{A} \neq \mathbf{A}^*$ [E. 2009], [Duintjer Tebbens, Meurant 2012].
- Stringent sufficient conditions are known [Carden 2011].

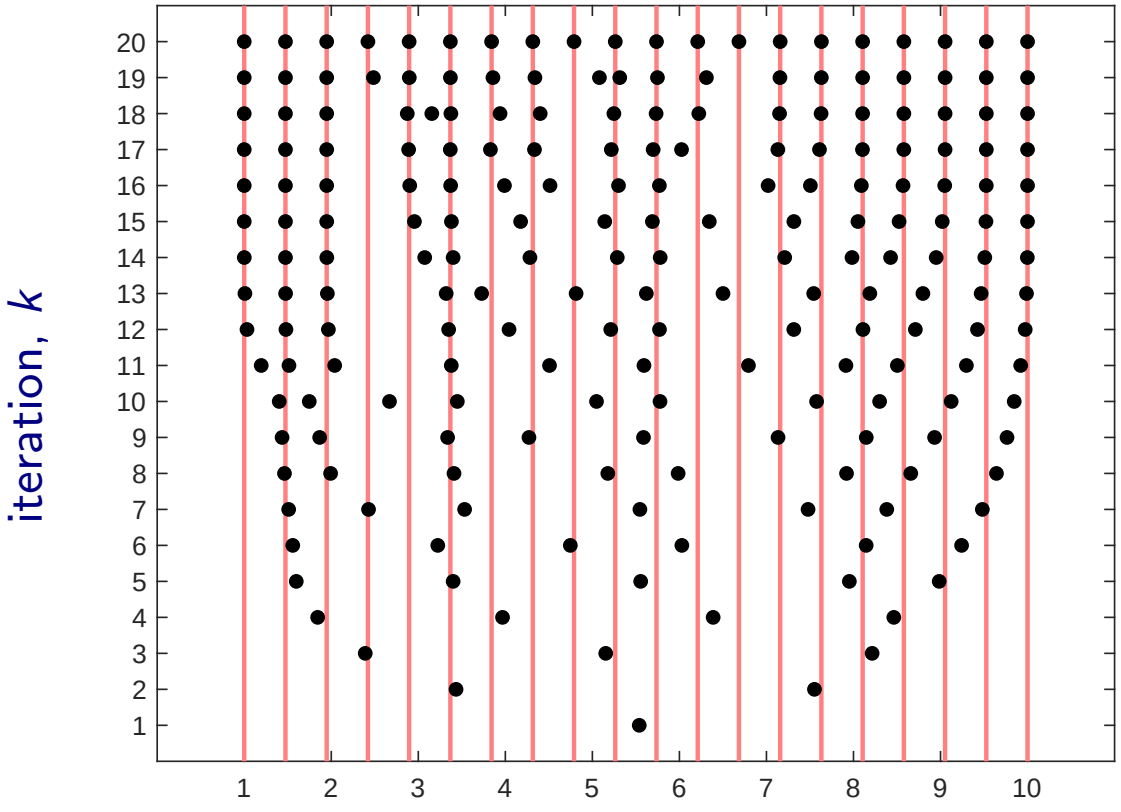
Do nonsymmetric matrices

enjoy any kind of

interlacing ?

interlacing is a key to convergence theory for $\mathbf{A} = \mathbf{A}^*$

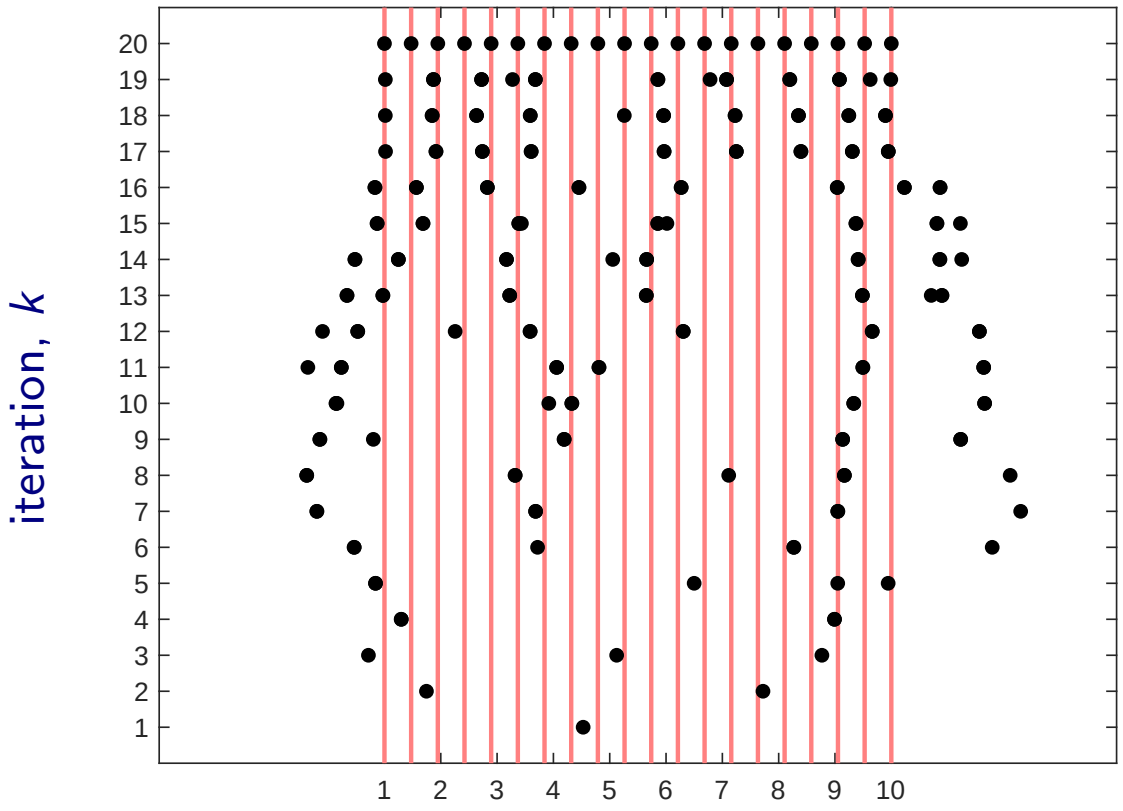
Cauchy's interlacing theorem assures us that Ritz values cannot bunch up at the ends of the spectrum.



eigenvalues (red lines) and Ritz values (black dots)

interlacing does not hold for $\mathbf{A} = \mathbf{A}^*$

The absence of interlacing for non-Hermitian problems is the major impediment to a full convergence theory – and is the mechanism that allows the method to fail (in theory).



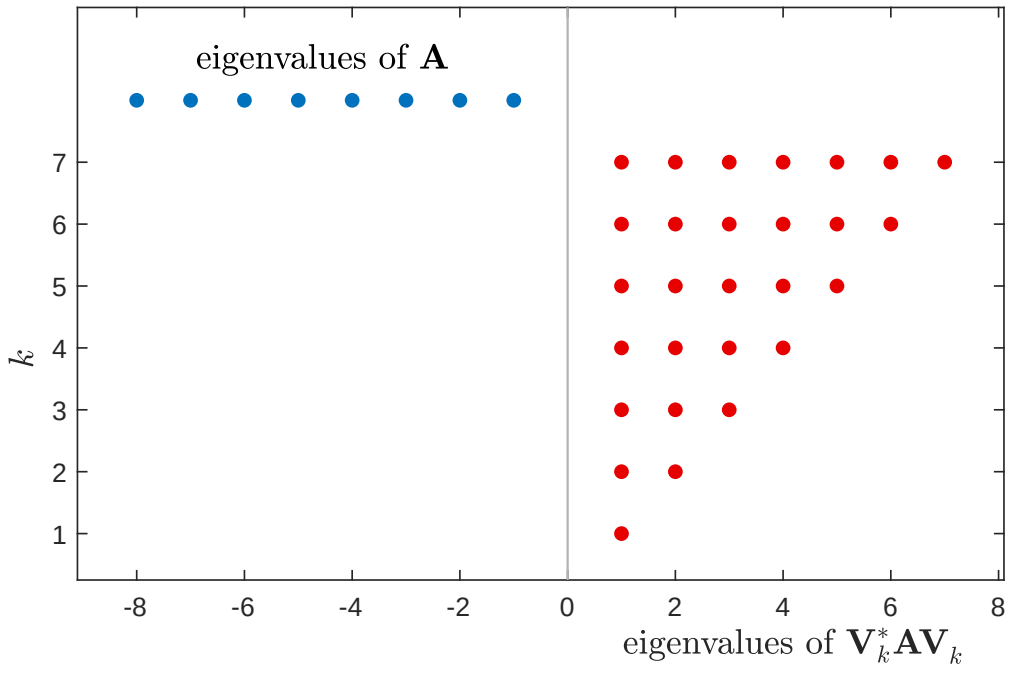
eigenvalues (red lines) and Ritz values (black dots = real parts)

a pathologically terrible example

A monster, built using the construction of [Duintjer Tebbens, Meurant 2012]:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & -362880 \\ 1 & 2 & 0 & 0 & 0 & 0 & 0 & -1451520 \\ & 1 & 3 & 0 & 0 & 0 & 0 & -1693440 \\ & & 1 & 4 & 0 & 0 & 0 & -846720 \\ & & & 1 & 5 & 0 & 0 & -211680 \\ & & & & 1 & 6 & 0 & -28224 \\ & & & & & 1 & 7 & -2016 \\ & & & & & & 1 & -64 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

arXiv:1801.00234



ritz value localization for non-hermitian matrices

Do non-Hermitian matrices obey any kind of "interlacing" theorem?

Ritz values must be contained within the *numerical range*

$$W(\mathbf{A}) = \{\mathbf{v}^* \mathbf{A}\, \mathbf{v} : \|\mathbf{v}\| = 1\},$$

a closed, convex subset of $\mathbb{C}$ that contains $\sigma(\mathbf{A})$.

ritz value localization for non-hermitian matrices

Do non-Hermitian matrices obey any kind of "interlacing" theorem?

Ritz values must be contained within the *numerical range*

$$W(\mathbf{A}) = \{\mathbf{v}^* \mathbf{A}\, \mathbf{v} : \|\mathbf{v}\| = 1\},$$

a closed, convex subset of $\mathbb{C}$ that contains $\sigma(\mathbf{A})$.

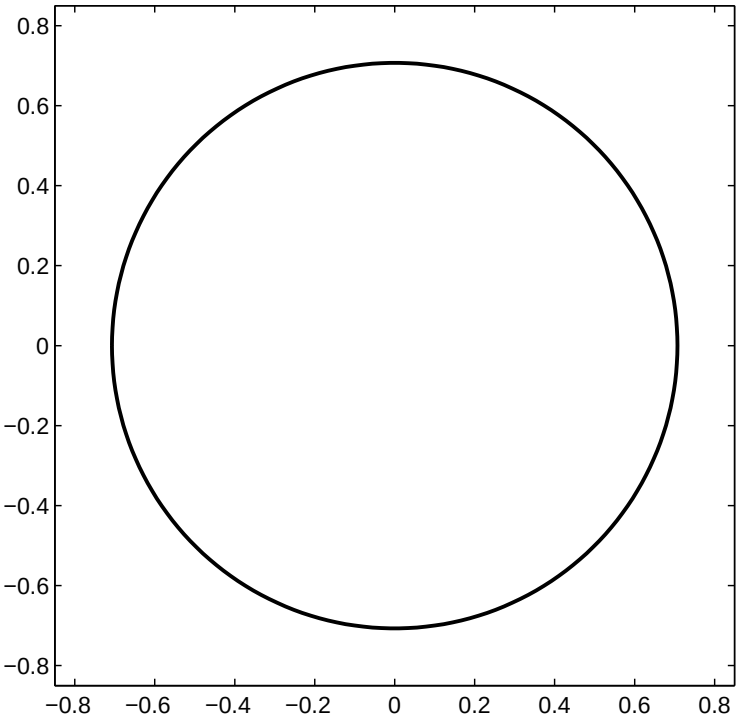
Consider an extreme example:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad W(\mathbf{A}) = \left\{ z \in \mathbb{C} : |z| \leq \frac{\sqrt{2}}{2} \right\}.$$

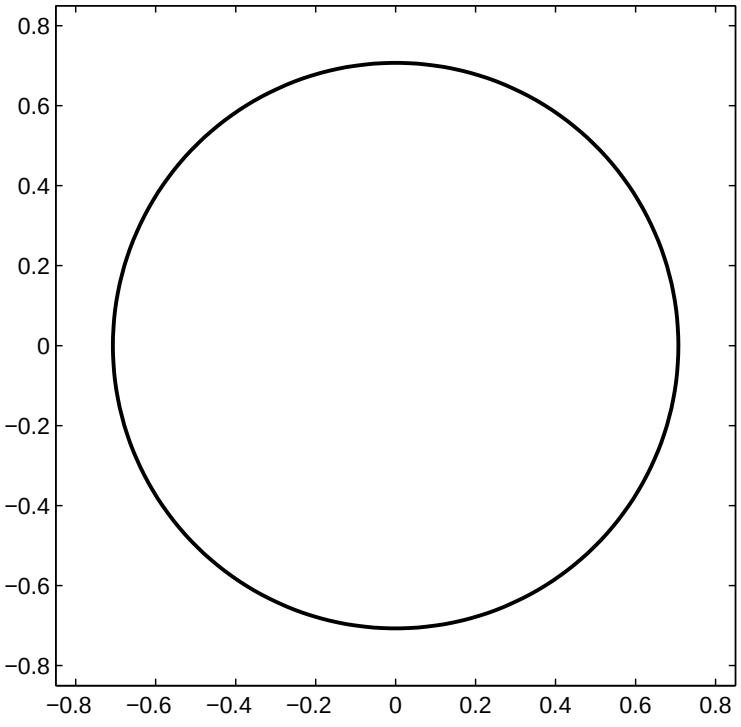
Repeat the following experiment many times:

- Generate random *two* dimensional subspaces, $\mathcal{V} = \operatorname{Ran} \mathbf{V}$, where $\mathbf{V}^*\mathbf{V} = \mathbf{I}$.
- Form $\mathbf{V}^*\mathbf{A}\mathbf{V} \in \mathbb{C}^{2\times 2}$ and compute Ritz values $\{\theta_1, \theta_2\} = \sigma(\mathbf{V}^*\mathbf{A}\mathbf{V})$.
- Identify the leftmost and rightmost Ritz values.
- Since $\sigma(\mathbf{A}) = \{0\}$, "interlacing" is meaningless here...

ritz values of a jordan block



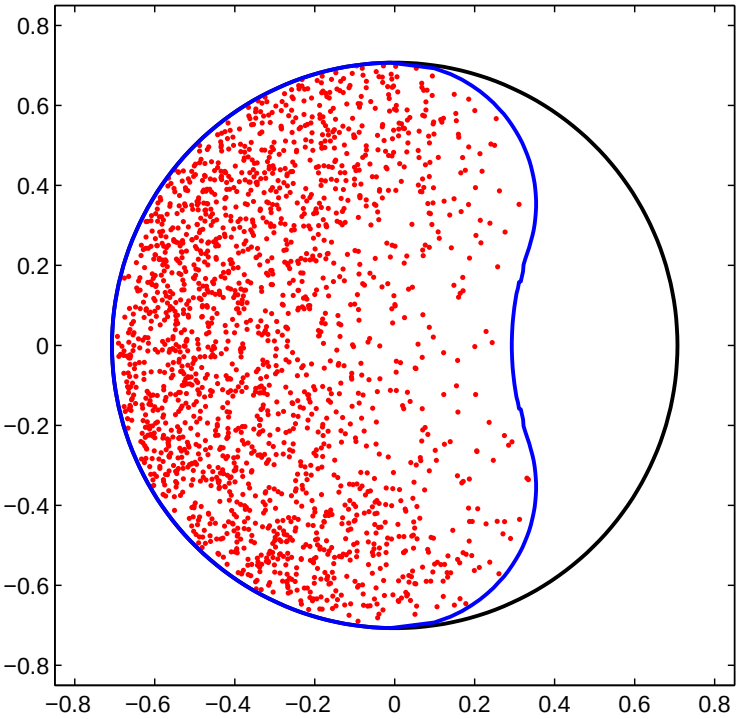
leftmost Ritz value



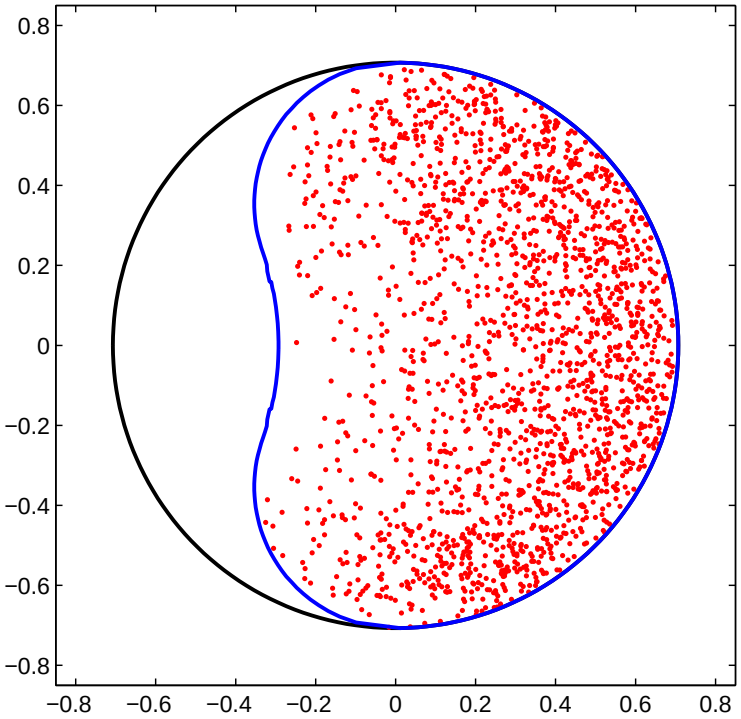
rightmost Ritz value

$$W(\mathbf{A}) = \left\{ z \in \mathbb{C} : |z| \leq \frac{\sqrt{2}}{2} \right\}$$

ritz values of a jordan block



leftmost Ritz value



rightmost Ritz value

10,000 random (complex) two dimensional subspaces

three matrices with identical $W(\mathbf{A})$

Compute $k = 4$ Ritz values for these 8×8 matrices.

$$\gamma_1 \begin{bmatrix} 0 & 1 & & & & & & \\ & 0 & & & & & & \\ & & 0 & 1 & & & & \\ & & & 0 & & & & \\ & & & & 0 & 1 & & \\ & & & & & 0 & & \\ & & & & & & 0 & 1 \\ & & & & & & & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & & & & & & \\ & 0 & 1 & & & & & \\ & & 0 & 1 & & & & \\ & & & 0 & 1 & & & \\ & & & & 0 & 1 & & \\ & & & & & 0 & 1 & \\ & & & & & & 0 & 1 \\ & & & & & & & 0 \end{bmatrix} \quad \gamma_3 \begin{bmatrix} 0 & \varrho^1 & & & & & & \\ & 0 & \varrho^2 & & & & & \\ & & 0 & \varrho^3 & & & & \\ & & & 0 & \varrho^4 & & & \\ & & & & 0 & \varrho^5 & & \\ & & & & & 0 & \varrho^6 & \\ & & & & & & 0 & \varrho^7 \\ & & & & & & & 0 \end{bmatrix}.$$

(γ_1 and γ_3 set to give same $W(\mathbf{A})$ for all examples; $\varrho = 1/8$.)

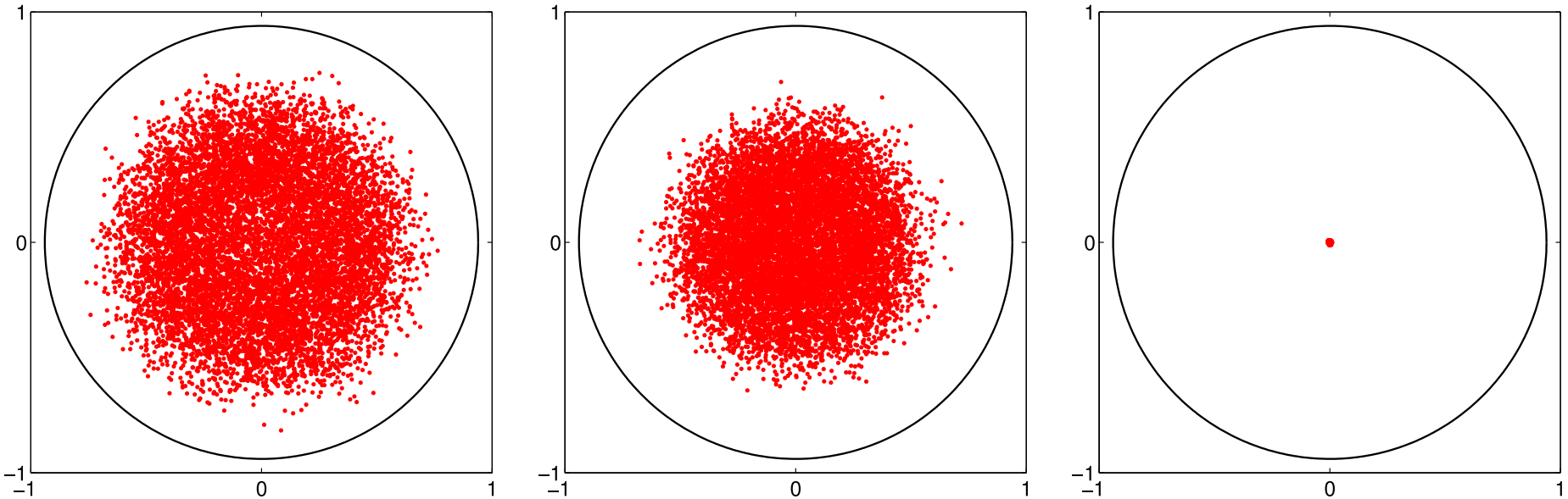
three matrices with identical $W(\mathbf{A})$

Compute $k = 4$ Ritz values for these 8×8 matrices.

$$\gamma_1 \begin{bmatrix} 0 & 1 & & & & & & \\ & 0 & & & & & & \\ & & 0 & 1 & & & & \\ & & & 0 & & & & \\ & & & & 0 & 1 & & \\ & & & & & 0 & & \\ & & & & & & 0 & 1 \\ & & & & & & & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & & & & & & \\ & 0 & 1 & & & & & \\ & & 0 & 1 & & & & \\ & & & 0 & 1 & & & \\ & & & & 0 & 1 & & \\ & & & & & 0 & 1 & \\ & & & & & & 0 & 1 \\ & & & & & & & 0 \end{bmatrix}$$

$$\gamma_3 \begin{bmatrix} 0 & \varrho^1 & & & & & & \\ & 0 & \varrho^2 & & & & & \\ & & 0 & \varrho^3 & & & & \\ & & & 0 & \varrho^4 & & & \\ & & & & 0 & \varrho^5 & & \\ & & & & & 0 & \varrho^6 & \\ & & & & & & 0 & \varrho^7 \\ & & & & & & & 0 \end{bmatrix}.$$

(γ_1 and γ_3 set to give same $W(\mathbf{A})$ for all examples; $\varrho = 1/8$.)



Smallest magnitude of $k = 4$ Ritz values, 10,000 random complex subspaces.

ritz value localization, sorted by real part

Using Schur's eigenvalue majorization theorem for Hermitian matrices, we can establish an interlacing-type result.

Theorem (Carden & E. 2012)

Let $\theta_1, \ldots, \theta_k$ denote the Ritz values of $\mathbf{A} \in \mathbb{C}^{n \times n}$ drawn from a $k < n$ dimensional subspace, labeled by decreasing real part: $\operatorname{Re}\theta_1 \geq \cdots \geq \operatorname{Re}\theta_k$. Then for $j = 1, \ldots, k$,

$$\frac{\mu_{n-k+j} + \cdots + \mu_n}{k - j + 1} \leq \operatorname{Re}\theta_j \leq \frac{\mu_1 + \cdots + \mu_j}{j},$$

where $\mu_1 \geq \cdots \geq \mu_n$ are the eigenvalues of $\frac{1}{2}(\mathbf{A} + \mathbf{A}^*)$.

ritz value localization, sorted by real part

Using Schur's eigenvalue majorization theorem for Hermitian matrices, we can establish an interlacing-type result.

Theorem (Carden & E. 2012)

Let $\theta_1, \ldots, \theta_k$ denote the Ritz values of $\mathbf{A} \in \mathbb{C}^{n\times n}$ drawn from a $k < n$ dimensional subspace, labeled by decreasing real part: $\operatorname{Re}\theta_1 \geq \cdots \geq \operatorname{Re}\theta_k$. Then for $j = 1, \ldots, k$,

$$\frac{\mu_{n-k+j} + \cdots + \mu_n}{k-j+1} \leq \operatorname{Re}\theta_j \leq \frac{\mu_1 + \cdots + \mu_j}{j},$$

where $\mu_1 \geq \cdots \geq \mu_n$ are the eigenvalues of $\frac{1}{2}(\mathbf{A} + \mathbf{A}^*)$.

- The fact that $\theta_j \in W(\mathbf{A})$ gives the well-known bound

  $$\mu_1 \leq \operatorname{Re}\theta_j \leq \mu_n, \qquad j = 1, \ldots, k.$$

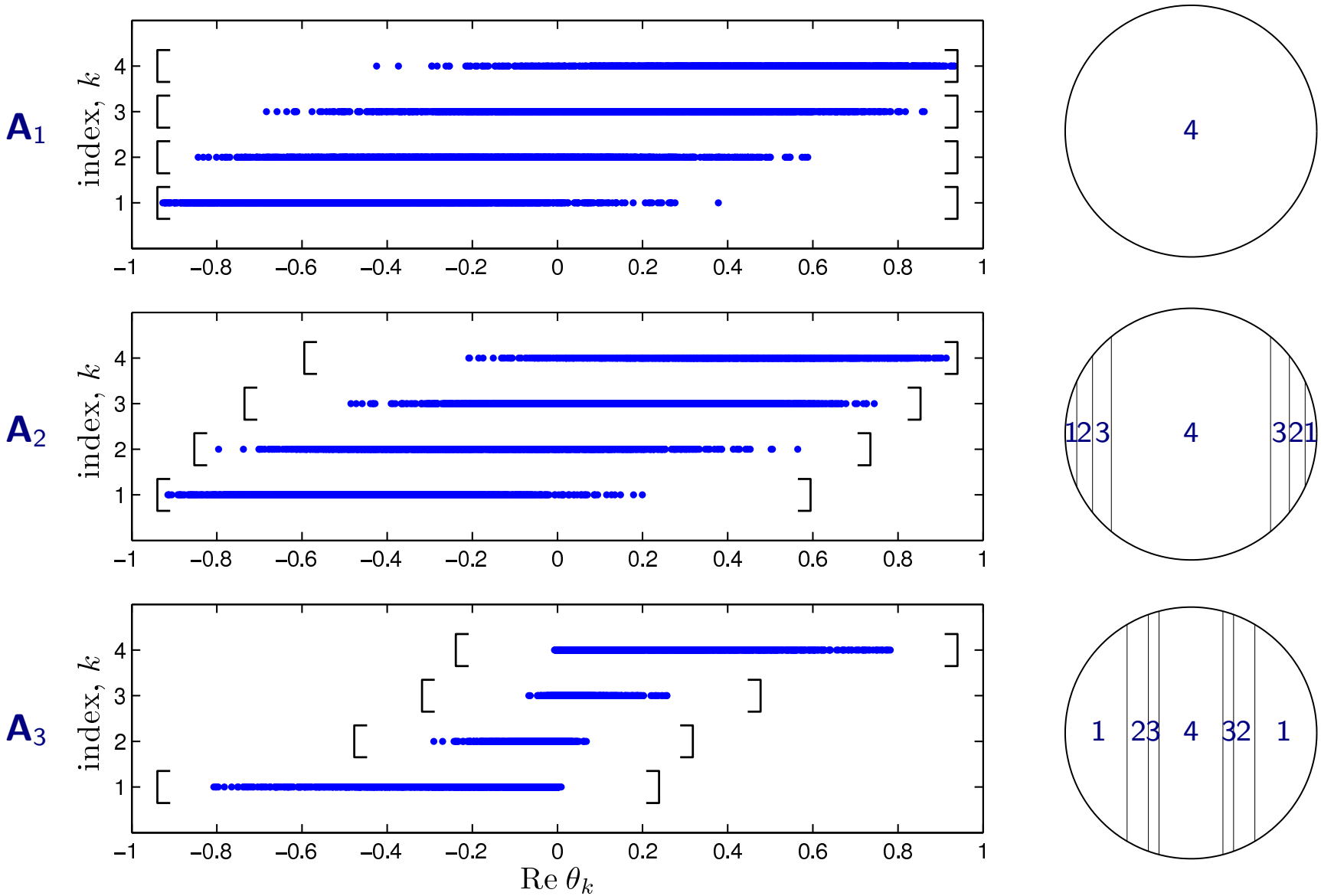
  The theorem provides sharper bounds for interior Ritz values.

- The interior eigenvalues of $\frac{1}{2}(\mathbf{A} + \mathbf{A}^*)$ give additional insight; cf. eigenvalue inclusion regions of [Psarrakos & Tsatsomeros, 2012].

- Theorem applies to any subspace Range($\mathbf{V}$): Krylov, block Krylov, etc.

three matrices with identical $W(\mathbf{A})$

Three matrices with the same $W(\mathbf{A})$, different interior structure; 2000 trials.
For $k = 4$, numbers on right indicate max Ritz values in each region.



ritz value localization, sorted by magnitude

The log-majorization of products of eigenvalues by products of singular values [Marshall, Olkin, Arnold 2011] leads to a limit on Ritz value magnitudes.

Theorem (Carden & E., 2012)

Let $\theta_1, \ldots, \theta_k$ denote the Ritz values of $\mathbf{A} \in \mathbb{C}^{n \times n}$ drawn from a $k < n$ dimensional subspace, labeled by decreasing magnitude: $|\theta_1| \geq \cdots \geq |\theta_k|$. Then for $j = 1, \ldots, k$,

$$|\theta_j| \leq \left(s_1 \cdots s_j\right)^{1/j},$$

where $s_1 \geq \cdots \geq s_n$ are the singular values of $\mathbf{A}$.

ritz value localization, sorted by magnitude

The log-majorization of products of eigenvalues by products of singular values [Marshall, Olkin, Arnold 2011] leads to a limit on Ritz value magnitudes.

Theorem (Carden & E., 2012)

Let $\theta_1, \ldots, \theta_k$ denote the Ritz values of $\mathbf{A} \in \mathbb{C}^{n \times n}$ drawn from a $k < n$ dimensional subspace, labeled by decreasing magnitude: $|\theta_1| \geq \cdots \geq |\theta_k|$. Then for $j = 1, \ldots, k$,

$$|\theta_j| \leq (s_1 \cdots s_j)^{1/j},$$

where $s_1 \geq \cdots \geq s_n$ are the singular values of $\mathbf{A}$.

Related results:

Zvonimir Bujanovic [2011] studies Ritz values of normal matrices *from Krylov subspaces* in his Ph.D. thesis (Zagreb).

Jakeniah Christiansen [2012] studies real Ritz values for $n = 3$ (SIURO).

some closing thoughts

Krylov methods can further develop as a prominent tool for RandNLA.

- Polynomials are better than powers!
  Krylov methods have great advantages over power/subspace iteration.
- Block methods hold promise but additional subtleties.
  Large subspaces can be built rapidly; must maintain linear independence.
- Restarting is crucial in engineering computations, but analysis is tricky.
  Restarting controls the subspace dimension, refines the starting vector.
- Spectral transformations (shift-invert) can vastly accelerate convergence.
  You are not entirely constrained by the eigenvalue distribution of $\mathbf{A}$.
- Non-Hermitian problems are solved everyday.
  The theory is incomplete and monsters are easy to construct,
  but the Krylov method (as implemented in `eigs`/ARPACK) works well.
- Can Random Matrix Theory shed light on Ritz value locations?
  What is the probability that $\mathbf{A}$ is stable if $\mathbf{V}^*\mathbf{A}\mathbf{V}$ is stable?