Pseudospectra and Nonnormal Dynamical Systems

Mark Embree and Russell Carden
Computational and Applied Mathematics
Rice University
Houston, Texas

ELGERSBURG

MARCH 2012

Overview of the Course

These lectures describe modern tools
for the spectral analysis of dynamical systems.
We shall cover a mix of theory, computation, and applications.

Lecture 1: Introduction to Nonnormality and Pseudospectra

Lecture 2: Functions of Matrices

Lecture 3: Toeplitz Matrices and Model Reduction

Lecture 4: Model Reduction, Numerical Algorithms, Differential Operators

Lecture 5: Discretization, Extensions, Applications

Outline for Today

Lecture 4: Differential Opeartors and Behavior of Numerical Algorithms

- Nonnormality and Lyapunov Equations
- Moment Matching Model Reduction
- Nonnormality in Iterative Linear Algebra
- Pseudospectra of Operators on Infinite Dimensional Spaces
- Pseudospectra of Differential Operators

4(a) Nonnormality and Lyapunov Equations

Lyapunov Equations

Balanced Truncation required the controlability and observability gramians

$$\mathbf{P} := \int_0^\infty \mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\,\mathrm{d}t, \qquad \mathbf{Q} := \int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}\,\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t$$

Hermitian positive definite, for a controllable and observable stable system.

Similarly, the total energy of the state in $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is given by

$$\int_0^\infty \|\mathbf{x}(t)\|^2\,\mathrm{d}t = \int_0^\infty \mathbf{x}_0^*\mathrm{e}^{t\mathbf{A}^*}\mathrm{e}^{t\mathbf{A}}\mathbf{x}_0\,\mathrm{d}t = \mathbf{x}_0^*\Big(\int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t\Big)\mathbf{x}_0 = \mathbf{x}_0^*\mathbf{E}\mathbf{x}_0.$$

Lyapunov Equations

Balanced Truncation required the controlability and observability gramians

$$\mathbf{P} := \int_0^\infty \mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\,\mathrm{d}t, \qquad \mathbf{Q} := \int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}\,\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t$$

Hermitian positive definite, for a controllable and observable stable system.

Similarly, the total energy of the state in $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is given by

$$\int_0^\infty \|\mathbf{x}(t)\|^2\,\mathrm{d}t = \int_0^\infty \mathbf{x}_0^*\mathrm{e}^{t\mathbf{A}^*}\mathrm{e}^{t\mathbf{A}}\mathbf{x}_0\,\mathrm{d}t = \mathbf{x}_0^*\Big(\int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t\Big)\mathbf{x}_0 = \mathbf{x}_0^*\mathbf{E}\mathbf{x}_0.$$

These integrals can be found as the solution of a *Lyapunov equation*.

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^* \quad = \quad \int_0^\infty \Big(\mathbf{A}\mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*} + \mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\mathbf{A}^*\Big)\,\mathrm{d}t$$

Lyapunov Equations

Balanced Truncation required the controlability and observability gramians

$$\mathbf{P} := \int_0^\infty \mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\,\mathrm{d}t, \qquad \mathbf{Q} := \int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}\,\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t$$

Hermitian positive definite, for a controllable and observable stable system.

Similarly, the total energy of the state in $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is given by

$$\int_0^\infty \|\mathbf{x}(t)\|^2\,\mathrm{d}t = \int_0^\infty \mathbf{x}_0^*\mathrm{e}^{t\mathbf{A}^*}\mathrm{e}^{t\mathbf{A}}\mathbf{x}_0\,\mathrm{d}t = \mathbf{x}_0^*\Big(\int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t\Big)\mathbf{x}_0 = \mathbf{x}_0^*\mathbf{E}\mathbf{x}_0.$$

These integrals can be found as the solution of a *Lyapunov equation*.

$$\begin{aligned}\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^* &= \int_0^\infty \Big(\mathbf{A}\mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*} + \mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\mathbf{A}^*\Big)\,\mathrm{d}t \\ &= \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\Big)\,\mathrm{d}t\end{aligned}$$

Lyapunov Equations

Balanced Truncation required the controlability and observability gramians

$$\mathbf{P} := \int_0^\infty \mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\,\mathrm{d}t, \qquad \mathbf{Q} := \int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}\,\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t$$

Hermitian positive definite, for a controllable and observable stable system.

Similarly, the total energy of the state in $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is given by

$$\int_0^\infty \|\mathbf{x}(t)\|^2\,\mathrm{d}t = \int_0^\infty \mathbf{x}_0^*\mathrm{e}^{t\mathbf{A}^*}\mathrm{e}^{t\mathbf{A}}\mathbf{x}_0\,\mathrm{d}t = \mathbf{x}_0^*\Big(\int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t\Big)\mathbf{x}_0 = \mathbf{x}_0^*\mathbf{E}\mathbf{x}_0.$$

These integrals can be found as the solution of a *Lyapunov equation*.

$$\begin{aligned}
\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^* &= \int_0^\infty \Big(\mathbf{A}\mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*} + \mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\mathbf{A}^*\Big)\,\mathrm{d}t \\
&= \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\Big)\,\mathrm{d}t \\
&= \Big[\mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\Big]_{t=0}^{\infty} = -\mathbf{b}\mathbf{b}^*.
\end{aligned}$$

Hence, $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{E}$ satisfy:

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^* = -\mathbf{b}\mathbf{b}^*, \qquad \mathbf{A}^*\mathbf{Q} + \mathbf{Q}\mathbf{A} = -\mathbf{c}^*\mathbf{c}, \qquad \mathbf{A}^*\mathbf{E} + \mathbf{E}\mathbf{A} = -\mathbf{I}.$$

Lyapunov Equations: Properties of the Solution

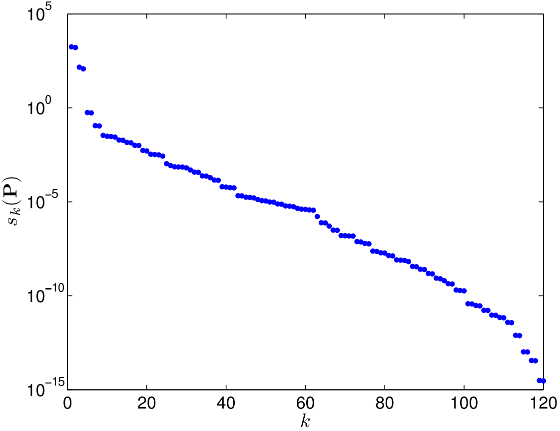
$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^* = -\mathbf{b}\mathbf{b}^* \qquad \mathbf{A}^*\mathbf{Q} + \mathbf{Q}\mathbf{A} = -\mathbf{c}^*\mathbf{c} \qquad \mathbf{A}^*\mathbf{E} + \mathbf{E}\mathbf{A} = -\mathbf{I}$$

- Under natural assumptions, **P**, **Q**, and **E** are Hermitian, positive definite.
- In general, **P**, **Q**, and **E** are *dense* $n \times n$ matrices.
- Key obstacle: We seek **P** and **Q** to perform *model reduction*, which suggests that the state space dimension n is large (e.g., 10^5 or larger). It could be impossible to store these solutions on modern computers.

Lyapunov Equations: Properties of the Solution

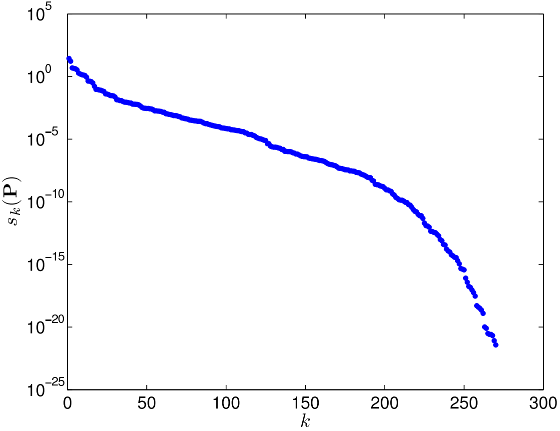
$$\mathbf{AP} + \mathbf{PA}^* = -\mathbf{bb}^* \qquad \mathbf{A}^*\mathbf{Q} + \mathbf{QA} = -\mathbf{c}^*\mathbf{c} \qquad \mathbf{A}^*\mathbf{E} + \mathbf{EA} = -\mathbf{I}$$

- Under natural assumptions, **P**, **Q**, and **E** are Hermitian, positive definite.
- In general, **P**, **Q**, and **E** are *dense* $n \times n$ matrices.
- Key obstacle: We seek **P** and **Q** to perform *model reduction*, which suggests that the state space dimension n is large (e.g., 10^5 or larger). It could be impossible to store these solutions on modern computers.
- Penzl [2000] observed, however, that when the right hand side is low rank, the singular values of the solution often *decay* rapidly. See also [Gudmundsson and Laub 1994].
- Such solutions can be approximated to high accuracy by *low-rank matrices*. Thus, the approximate solutions can be stored very efficiently.



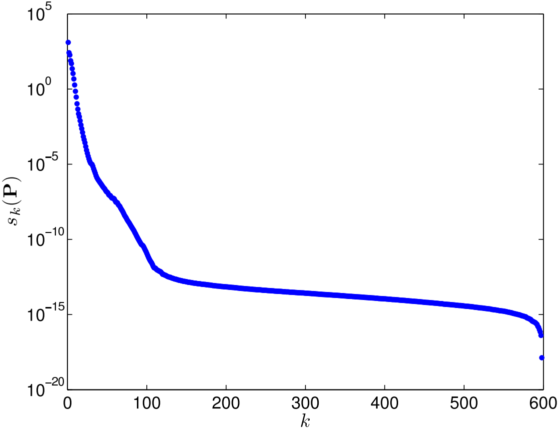
"CD Player" from SLICOT

Lyapunov Equations: Properties of the Solution



"International Space Station" from SLICOT

Lyapunov Equations: Properties of the Solution



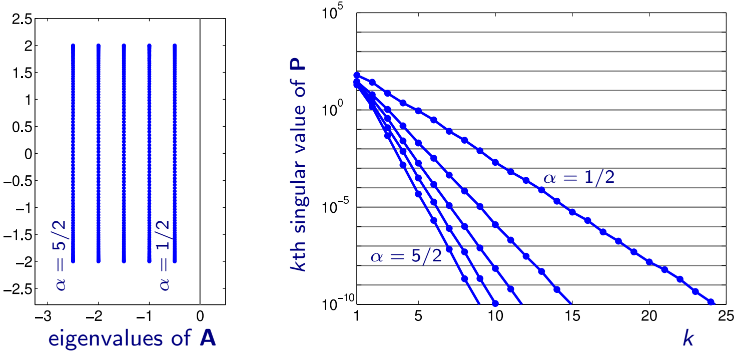
"Eady" storm track model from SLICOT

Decay of Singular Values of Lyapunov Solutions

The eigenvalues of **A** affect the rate of decay, as seen in this example.

Normal matrices: $\mathbf{A} = \text{tridiag}(-1, \alpha, 1)$ with spectrum

$$\sigma(\mathbf{A}) \subseteq \{\alpha + iy : y \in [-2, 2]\}.$$

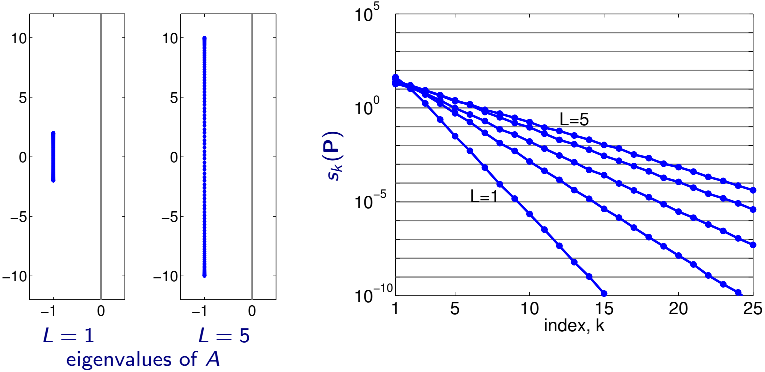


Convergence slows as $\alpha \uparrow 0$, i.e., eigenvalues move toward imaginary axis.

Decay of Singular Values of Lyapunov Solutions

Normal example: $A = \text{tridiag}(-L, -1, L)$ with spectrum

$$\sigma(A) \subseteq \{-1 + \mathrm{i}y : y \in [-2L, 2L]\}.$$



Convergence slows as L increases, though the spectral abscissa is unchanged.

Decay of Singular Values of Lyapunov Solutions

Let s_k denote the kth singular value of $\mathbf{P}$, $s_k \geq s_{k+1}$.

- For Hermitian $\mathbf{A}$, bounds on s_k/s_1 have been derived by Penzl [2000], Sabino [2006] (see Ellner & Wachspress [1991]).

- For diagonalizable $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, Antoulas, Sorensen, and Zhou [2002] prove

$$s_{k+1} \leq \kappa(\mathbf{V})^2 \delta_{k+1} \|\mathbf{b}\|^2 (N-k)^2, \qquad \delta_k = \frac{-1}{2\operatorname{Re}\lambda_k} \prod_{j=1}^{k-1} \frac{|\lambda_k - \lambda_j|^2}{|\overline{\lambda}_k + \lambda_j|^2}.$$

  The $\kappa(\mathbf{V})^2$ term imposes a significant penalty for nonnormality; cf. [Truhar, Tomljanović, Li 2009].

  We seek a bound that gives a more flexible approach, by enlarging the set over which we study rational functions like δ_k.

Decay of Singular Values of Lyapunov Solutions

The ADI iteration [Wachspress 1988; Penzl 2000] constructs a rank-k approximation $\mathbf{P}_k$ to $\mathbf{P}$.

- Pick *shifts* $\mu_1, \mu_2, \ldots, .$
- Set

$$\mathbf{A}_{\mu_j} = (\overline{\mu_j} - \mathbf{A})(\mu_j + \mathbf{A})^{-1}, \qquad \mathbf{b}_{\mu_j} = \sqrt{-2\mathrm{Re}\,\mu_j}(\mu_j + \mathbf{A})^{-1}\mathbf{b}.$$

- Set $\mathbf{P}_0 := \mathbf{0}$.
- For $k = 1, 2, \ldots, p$

$$\mathbf{P}_k := \mathbf{A}_{\mu_k}\mathbf{P}_{k-1}\mathbf{A}_{\mu_k}^* + \mathbf{b}_{\mu_k}\mathbf{b}_{\mu_k}^*.$$

- Then the solution $\mathbf{P}$ satisfies

$$\mathbf{P} - \mathbf{P}_k \;=\; \phi(\mathbf{A})\mathbf{P}\,\phi(\mathbf{A})^*,$$

  where $\mathrm{rank}(\mathbf{P}_k) \le k$ and

$$\phi(z) \;=\; \prod_{j=1}^{k} \frac{\overline{\mu}_j - z}{\mu_j + z}.$$

Decay of Singular Values of Lyapunov Solutions

- For comparison, the bound of Antoulas, Sorensen, Zhou [2002]:

$$s_{k+1} \le \kappa(\mathbf{V})^2 \delta_{k+1} \|\mathbf{b}\|^2 (N-k)^2, \qquad \delta_k = \frac{-1}{2\,\mathrm{Re}\,\lambda_k} \prod_{j=1}^{k-1} \frac{|\lambda_k - \lambda_j|^2}{|\overline{\lambda}_k + \lambda_j|^2}.$$

- The ADI iteration gives $\mathbf{P} - \mathbf{P}_k \;=\; \phi(\mathbf{A})\mathbf{P}\,\phi(\mathbf{A})^*$, where $\mathrm{rank}(\mathbf{P}_k) \le k$ and

$$\phi(z) \;=\; \prod_{j=1}^{k} \frac{\overline{\mu}_j - z}{\mu_j + z}.$$

- Hence $s_{k+1} \le \|\phi(\mathbf{A})\|^2 \|\mathbf{P}\| = \|\phi(\mathbf{A})\|^2 s_1$.

Theorem (Pseudospectral Bound on Singular Value Decay)

Using the fact that $\mathrm{rank}(\mathbf{P}_k) \le k$*,*

$$\frac{s_{k+1}}{s_1} \;\le\; \min_{\mu_1,\ldots,\mu_k} \|\phi(\mathbf{A})\|_2^2 \;\le\; \left(\frac{L_\varepsilon}{2\pi\varepsilon}\right)^2 \min_{\mu_1,\ldots,\mu_k} \; \sup_{z \in \sigma_\varepsilon(\mathbf{A})} \prod_{j=1}^{k} \frac{|\overline{\mu}_j - z|^2}{|\mu_j + z|^2},$$

where L_ε *is the boundary length of a contour enclosing* $\sigma_\varepsilon(\mathbf{A})$

[Beattie, E., Sabino; cf. Beckermann (2004)].

Singular Values of Lyapunov Solutions

Theorem (Pseudospectral Bound on Singular Value Decay)

Using the fact that $\mathrm{rank}(\mathbf{P}_k) \leq k$,

$$\frac{s_{k+1}}{s_1} \;\leq\; \min_{\mu_1,\ldots,\mu_k} \|\phi(\mathbf{A})\|_2^2 \;\leq\; \left(\frac{L_\varepsilon}{2\pi\varepsilon}\right)^2 \min_{\mu_1,\ldots,\mu_k} \sup_{z\in\sigma_\varepsilon(\mathbf{A})} \prod_{j=1}^{k} \frac{|\overline{\mu}_j - z|^2}{|\mu_j + z|^2},$$

where L_ε *is the boundary length of a contour enclosing* $\sigma_\varepsilon(\mathbf{A})$

- How should one choose the shifts, $\mu_1, \mu_2, \ldots$?

Singular Values of Lyapunov Solutions

Theorem (Pseudospectral Bound on Singular Value Decay)

Using the fact that $\operatorname{rank}(\mathbf{P}_k) \leq k$,

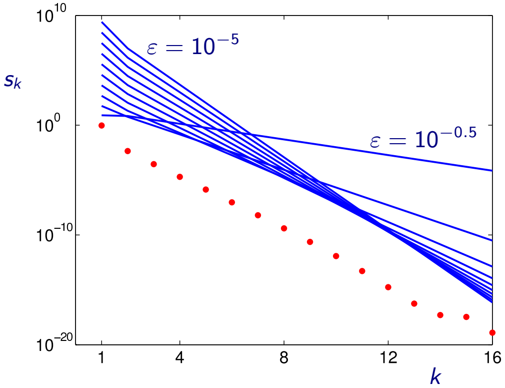
$$\frac{s_{k+1}}{s_1} \leq \min_{\mu_1,\ldots,\mu_k} \|\phi(\mathbf{A})\|_2^2 \leq \left(\frac{L_\varepsilon}{2\pi\varepsilon}\right)^2 \min_{\mu_1,\ldots,\mu_k} \sup_{z\in\sigma_\varepsilon(\mathbf{A})} \prod_{j=1}^{k} \frac{|\overline{\mu}_j - z|^2}{|\mu_j + z|^2},$$

where L_ε *is the boundary length of a contour enclosing* $\sigma_\varepsilon(\mathbf{A})$

- How should one choose the shifts, $\mu_1, \mu_2, \ldots$?
- One can pick the shifts to minimize the upper bound, i.e., be asymptotically optimal rational interpolation points for $\sigma_\varepsilon(\mathbf{A})$ for some $\varepsilon > 0$.
- Other heuristics are also effective. For example, Sabino [2006] shows that MATLAB's `fminsearch` can give excellent results for a few shifts.

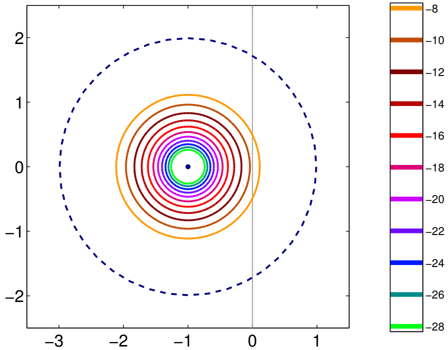
Examples of the Decay Bound Based on Pseudospectra

$$\mathbf{A} = \begin{bmatrix} -1 & 1/2 & & & \\ & -1 & \ddots & & \\ & & \ddots & 1/2 & \\ & & & -1 & 1/2 \\ & & & & -1 \end{bmatrix} \in \mathbf{C}^{16\times 16}$$



Examples of the Decay Bound Based on Pseudospectra

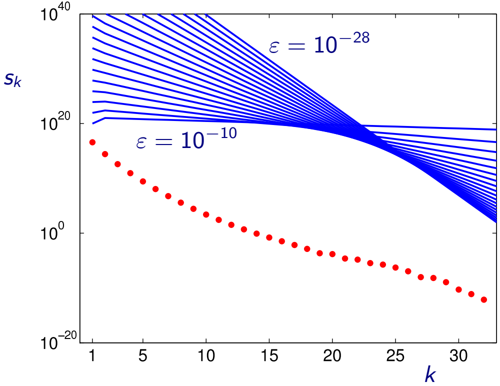
$$\mathbf{A} = \begin{bmatrix} -1 & 2 & & & \\ & -1 & \ddots & & \\ & & \ddots & 2 & \\ & & & -1 & 2 \\ & & & & -1 \end{bmatrix} \in \mathbf{C}^{32\times 32}$$



Spectrum, pseudospectra, and numerical range (boundary = dashed line)

Examples of the Decay Bound Based on Pseudospectra

$$\mathbf{A} = \begin{bmatrix} -1 & 2 & & & \\ & -1 & \ddots & & \\ & & \ddots & 2 & \\ & & & -1 & 2 \\ & & & & -1 \end{bmatrix} \in \mathbb{C}^{32\times 32}$$



Any Decay Possible for Any Spectrum

Theorem (Penzl, 2000)

Let $\mathbf{A}$ *be stable and* $\mathbf{b}$ *some vector such that* $(\mathbf{A}, \mathbf{b})$ *is controllable. Given any Hermitian positive definite* $\mathbf{P}$*, there exists some invertible matrix* $\mathbf{S}$ *such that*

$$(\mathbf{S}\mathbf{A}\mathbf{S}^{-1})\mathbf{P} + \mathbf{P}(\mathbf{S}\mathbf{A}\mathbf{S}^{-1}) = -(\mathbf{S}\mathbf{b})(\mathbf{S}\mathbf{b})^*.$$

Any singular value decay is possible for a matrix with any eigenvalues.

Proof. The proof is a construction.

- Solve $\mathbf{A}\mathbf{Y} + \mathbf{Y}\mathbf{A}^* = -\mathbf{b}\mathbf{b}^*$ for $\mathbf{Y}$.
  ($\mathbf{Y}$ is Hermitian positive definite, since $(\mathbf{A}, \mathbf{b})$ controllable.)
- Set $\mathbf{S} := \mathbf{P}^{1/2}\mathbf{Y}^{-1/2}$.
- Notice that $\mathbf{S}\mathbf{Y}\mathbf{S}^* = \mathbf{P}^{1/2}\mathbf{Y}^{-1/2}\mathbf{Y}\mathbf{Y}^{-1/2}\mathbf{P}^{1/2} = \mathbf{P}$.
- Define $\widehat{\mathbf{A}} := \mathbf{S}\mathbf{A}\mathbf{S}^{-1}$, $\widehat{\mathbf{b}} := \mathbf{S}\mathbf{b}$.

Now it is easy to verify that $\mathbf{P}$ solves the desired Lyapunov equation:

$$\begin{aligned}\widehat{\mathbf{A}}\mathbf{P} + \mathbf{P}\widehat{\mathbf{A}}^* &= (\mathbf{S}\mathbf{A}\mathbf{S}^{-1})(\mathbf{S}\mathbf{Y}\mathbf{S}^*) + (\mathbf{S}\mathbf{Y}\mathbf{S}^*)(\mathbf{S}^{-*}\mathbf{A}\mathbf{S}^*) \\ &= \mathbf{S}(\mathbf{A}\mathbf{Y} + \mathbf{Y}\mathbf{A})\mathbf{S}^* = -(\mathbf{S}\mathbf{b})(\mathbf{S}\mathbf{b})^* = -\widehat{\mathbf{b}}\widehat{\mathbf{b}}^*. \qquad \square\end{aligned}$$

A Nonnormal Anomaly

The pseudospectral bound and the bound of Antoulas, Sorensen, and Zhou both predict that *the decay rate slows* as nonnormality increases.

However, for the solutions to Lyapunov equations this intuition can be wrong [Sabino 2006].

A Nonnormal Anomaly

The pseudospectral bound and the bound of Antoulas, Sorensen, and Zhou both predict that *the decay rate slows* as nonnormality increases.

However, for the solutions to Lyapunov equations this intuition can be wrong [Sabino 2006].

Consider

$$\mathbf{A} = \begin{bmatrix} -1 & \alpha \\ 0 & -1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} t \\ 1 \end{bmatrix}, \quad t \in \mathbf{R}.$$

As α grows, **A**'s departure from normality grows.
All bounds suggest that the 'decay' rate should worsen as α increases.

A Nonnormal Anomaly

The pseudospectral bound and the bound of Antoulas, Sorensen, and Zhou both predict that *the decay rate slows* as nonnormality increases.

However, for the solutions to Lyapunov equations this intuition can be wrong [Sabino 2006].

Consider

$$\mathbf{A} = \begin{bmatrix} -1 & \alpha \\ 0 & -1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} t \\ 1 \end{bmatrix}, \quad t \in \mathbf{R}.$$

As α grows, **A**'s departure from normality grows.
All bounds suggest that the 'decay' rate should worsen as α increases.

The Lyapunov equation $\mathbf{AP} + \mathbf{PA}^* = -\mathbf{bb}^*$ has the solution

$$\mathbf{P} = \frac{1}{4} \begin{bmatrix} 2t^2 + 2\alpha t + \alpha^2 & \alpha + 2t \\ \alpha + 2t & 2 \end{bmatrix}.$$

For each α, we wish to pick t to maximize the 'decay', i.e., s_2/s_1.

A Nonnormal Anomaly

The pseudospectral bound and the bound of Antoulas, Sorensen, and Zhou both predict that *the decay rate slows* as nonnormality increases.

However, for the solutions to Lyapunov equations this intuition can be wrong [Sabino 2006].

Consider

$$\mathbf{A} = \begin{bmatrix} -1 & \alpha \\ 0 & -1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} t \\ 1 \end{bmatrix}, \quad t \in \mathbf{R}.$$

As α grows, **A**'s departure from normality grows.
All bounds suggest that the 'decay' rate should worsen as α increases.

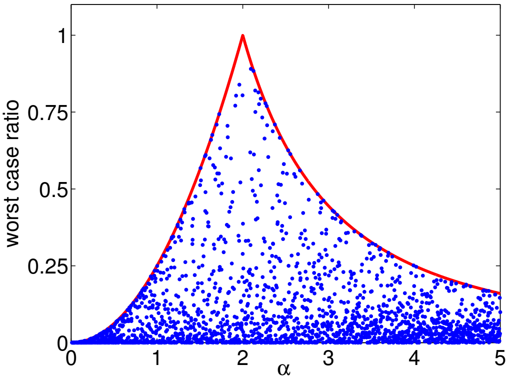
The Lyapunov equation $\mathbf{AP} + \mathbf{PA}^* = -\mathbf{bb}^*$ has the solution

$$\mathbf{P} = \frac{1}{4} \begin{bmatrix} 2t^2 + 2\alpha t + \alpha^2 & \alpha + 2t \\ \alpha + 2t & 2 \end{bmatrix}.$$

For each α, we wish to pick t to maximize the 'decay', i.e., s_2/s_1.
This is accomplished for $t = -\alpha/2$, giving

$$\frac{s_2}{s_1} = \begin{cases} \alpha^2/4, & 0 < \alpha \le 2; \\ 4/\alpha^2, & 2 \le \alpha. \end{cases}$$

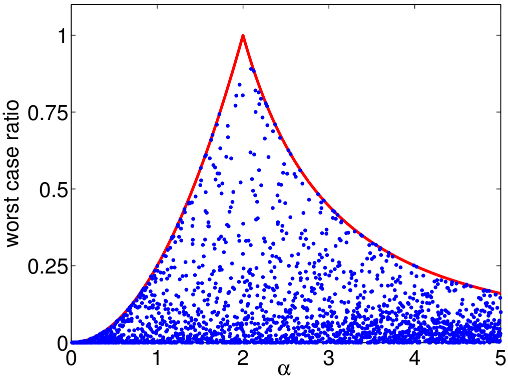
Jordan block: 2×2 case, numerical illustration



Red line: $\dfrac{s_2}{s_1} = \begin{cases} \alpha^2/4, & 0 < \alpha \leq 2; \\ 4/\alpha^2, & 2 \leq \alpha. \end{cases}$

Blue dots: s_2/s_1 for random $\mathbf{b}$ and 2000 α values.

Jordan block: 2×2 case, numerical illustration



Red line: $\dfrac{s_2}{s_1} = \begin{cases} \alpha^2/4, & 0 < \alpha \le 2; \\ 4/\alpha^2, & 2 \le \alpha. \end{cases}$

Blue dots: s_2/s_1 for random $\mathbf{b}$ and 2000 α values.

The effect of nonnormality on Lyapunov solutions
remains only partially understood.

4(d) Moment Matching Model Reduction

Krylov methods for moment matching model reduction

We now turn to a model reduction approach where nonnormality plays a crucial role. Once again, begin with the SISO system

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t)\\ y(t) &= \mathbf{c}\mathbf{x}(t) + du(t),\end{aligned}$$

with $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\mathbf{b}, \mathbf{c}^{\mathsf{T}} \in \mathbb{C}^n$ and initial condition $\mathbf{x}(0) = \mathbf{x}_0$.

Compress the state space to dimension $k \ll n$ via a projection method:

$$\begin{aligned}\dot{\widehat{\mathbf{x}}}(t) &= \widehat{\mathbf{A}}\widehat{\mathbf{x}}(t) + \widehat{\mathbf{b}}u(t)\\ \widehat{y}(t) &= \widehat{\mathbf{c}}\widehat{\mathbf{x}}(t) + du(t),\end{aligned}$$

where

$$\widehat{\mathbf{A}} = \mathbf{W}^*\mathbf{A}\mathbf{V} \in \mathbb{C}^{k\times k}, \qquad \widehat{\mathbf{b}} = \mathbf{W}^*\mathbf{b} \in \mathbb{C}^{k\times 1}, \qquad \widehat{\mathbf{c}} = \mathbf{c}\mathbf{V} \in \mathbb{C}^{1\times k}$$

for some $\mathbf{V}, \mathbf{W} \in \mathbb{C}^{n\times k}$ with $\mathbf{W}^*\mathbf{V} = \mathbf{I}$.

The matrices $\mathbf{V}$ and $\mathbf{W}$ are constructed by a Krylov subspace method.

Krylov methods for moment matching model reduction

$$\widehat{\mathbf{A}} = \mathbf{W}^*\mathbf{A}\mathbf{V} \in \mathbb{C}^{k\times k}, \qquad \widehat{\mathbf{b}} = \mathbf{W}^*\mathbf{b} \in \mathbb{C}^{k\times 1}, \qquad \widehat{\mathbf{c}} = \mathbf{c}\mathbf{V} \in \mathbb{C}^{1\times k}$$

Arnoldi Reduction

If $\mathbf{V} = \mathbf{W}$ and the columns of $\mathbf{V}$ span the kth Krylov subspace,

$$\mathrm{Ran}(\mathbf{V}) = \mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \mathrm{span}\{\mathbf{b}, \mathbf{A}\mathbf{b}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\},$$

then the reduced model matches k moments of the system:

$$\widehat{\mathbf{c}}\widehat{\mathbf{A}}^j\widehat{\mathbf{b}} = \mathbf{c}\mathbf{A}^j\mathbf{b}, \qquad j = 0, \ldots, k-1.$$

Bi-Lanczos Reduction

If the columns and $\mathbf{V}$ and $\mathbf{W}$ span the Krylov subspaces

$$\begin{aligned}
\mathrm{Ran}(\mathbf{V}) &= \mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \mathrm{span}\{\mathbf{b}, \mathbf{A}\mathbf{b}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\} \\
\mathrm{Ran}(\mathbf{W}) &= \mathcal{K}_k(\mathbf{A}^*, \mathbf{c}^*) = \mathrm{span}\{\mathbf{c}^*, \mathbf{A}^*\mathbf{c}^*, \ldots, (\mathbf{A}^*)^{k-1}\mathbf{c}^*\},
\end{aligned}$$

then the reduced model matches $2k$ moments of the system:

$$\widehat{\mathbf{c}}\widehat{\mathbf{A}}^j\widehat{\mathbf{b}} = \mathbf{c}\mathbf{A}^j\mathbf{b}, \qquad j = 0, \ldots, 2k-1.$$

Stability of Reduced Order Models

Question

Does the reduced model inherit properties of the original system?

Properties include stability, passivity, second-order structure, etc.

In this lecture we are concerned with *stability* – and, more generally, the behavior of eigenvalues of the reduced matrix $\widehat{\mathbf{A}}$.

Stability of Reduced Order Models

Question

Does the reduced model inherit properties of the original system?

Properties include stability, passivity, second-order structure, etc.

In this lecture we are concerned with *stability* – and, more generally, the behavior of eigenvalues of the reduced matrix $\widehat{\mathbf{A}}$.

Observation

For Arnoldi reduction, the eigenvalues of $\widehat{\mathbf{A}}$ *are contained in the numerical range*

$$W(\mathbf{A}) = \{\mathbf{x}^* \mathbf{A} \mathbf{x} : \|\mathbf{x}\| = 1\}.$$

Proof: If $\widehat{\mathbf{A}}\mathbf{z} = \theta \mathbf{z}$ for $\|\mathbf{z}\| = 1$, then $\theta = \mathbf{z}^* \mathbf{V}^* \mathbf{A} \mathbf{V}^* \mathbf{z} = (\mathbf{V}\mathbf{z})^* \mathbf{A} (\mathbf{V}\mathbf{z})$, where $\|\mathbf{V}\mathbf{z}\| = \|\mathbf{z}\| = 1$ $\square$

Stability of Reduced Order Models

Question

Does the reduced model inherit properties of the original system?

Properties include stability, passivity, second-order structure, etc.

In this lecture we are concerned with *stability* – and, more generally, the behavior of eigenvalues of the reduced matrix $\widehat{\mathbf{A}}$.

Observation

For Arnoldi reduction, the eigenvalues of $\widehat{\mathbf{A}}$ *are contained in the numerical range*

$$W(\mathbf{A}) = \{\mathbf{x}^*\mathbf{A}\mathbf{x} : \|\mathbf{x}\| = 1\}.$$

Proof: If $\widehat{\mathbf{A}}\mathbf{z} = \theta\mathbf{z}$ for $\|\mathbf{z}\| = 1$, then $\theta = \mathbf{z}^*\mathbf{V}^*\mathbf{A}\mathbf{V}^*\mathbf{z} = (\mathbf{V}\mathbf{z})^*\mathbf{A}(\mathbf{V}\mathbf{z})$, where $\|\mathbf{V}\mathbf{z}\| = \|\mathbf{z}\| = 1$ $\square$

No such bound for Bi-Lanczos: the decomposition may not even exist.
For the famous 'CD Player' model, $\mathbf{cb} = \mathbf{0}$: method breaks down at first step.

Stability and Bi-Lanczos

For Bi-Lanczos, the stability question is rather more subtle.

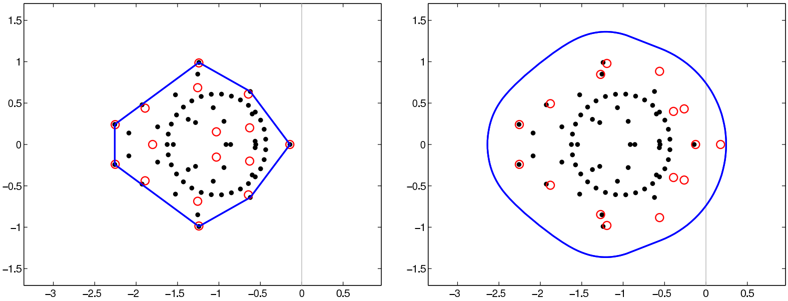
For example, if $\mathbf{b}$ and $\mathbf{c}$ are nearly orthogonal,
the (1,1) entry in $\widehat{\mathbf{A}} = \mathbf{W}^*\mathbf{AV}$ will generally be very large.

Given a fixed $\mathbf{b}$, one has much freedom to rig poor results via $\mathbf{c}$:

> [A]ny three-term recurrence (run for no more than $(n+2)/2$ steps, where n is the size of the matrix) *is* the two-sided Lanczos algorithm for some left starting vector. [Greenbaum, 1998]

Example: Arnoldi Reduction for Normal versus Nonnormal

The eigenvalues of $\widehat{\mathbf{A}}$ are known as *Ritz values*.



normal matrix　　　nonnormal matrix

Eigenvalues (●), Ritz values (○), and numerical range for isospectral matrices.

A Remedy for Unstable Arnoldi Models?

One can counteract instability by *restarting* the Arnoldi algorithm to shift out unstable eigenvalues [Grimme, Sorensen, Van Dooren, 1994]; cf. [Jaimoukha and Kasenally, 1997].

- $\widehat{\mathbf{A}} := \mathbf{V}^*\mathbf{A}\mathbf{V}$ has eigenvalues $\theta_1, \ldots, \theta_k$ (Ritz values for $\mathbf{A}$)
- Suppose $\theta_1, \ldots, \theta_p$ are in the right half plane.
- Replace the starting vector $\mathbf{b}$ by the filtered vector

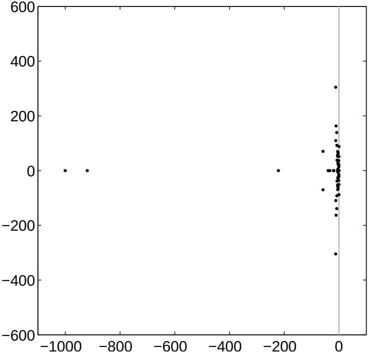
  $$\mathbf{b}_+ = \psi(\mathbf{A})\mathbf{b}, \qquad \psi(z) = \prod_{j=1}^{p}(z - \theta_j),$$

  where the *filter polynomial* ψ "discourages" future Ritz values near the *shifts* $\theta_1, \ldots, \theta_p$.
- Build new matrices $\mathbf{V}$, $\widehat{\mathbf{A}}$ with starting vector $\mathbf{b}_+$ (implicit restart).
- Now *modified* moments, $\mathbf{b}^*\psi(\mathbf{A})^*\mathbf{A}^j\psi(\mathbf{A})\mathbf{b}$, will be matched.

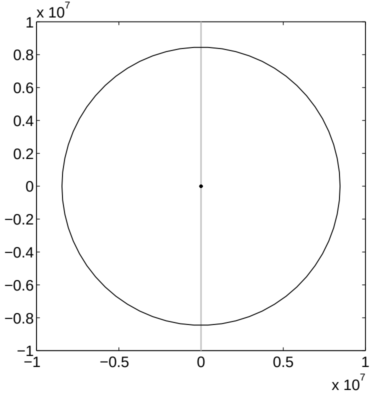
Repeat this process until $\widehat{\mathbf{A}}$ has no unstable modes.

Matching the Moments of a Nonnormal Matrix

Model of flutter in a Boeing 767 from SLICOT ($n = 55$),
stabilized by Burke, Lewis, Overton [2003].



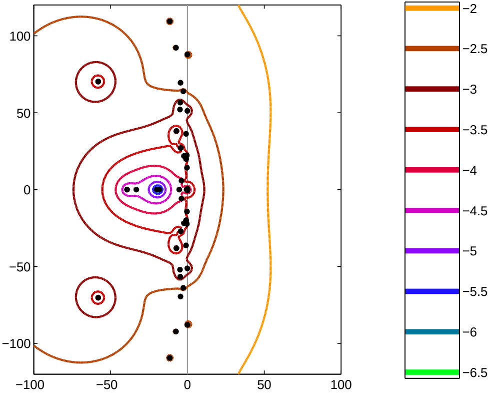
$\sigma(\mathbf{A})$



$W(\mathbf{A})$

Matching the Moments of a Nonnormal Matrix

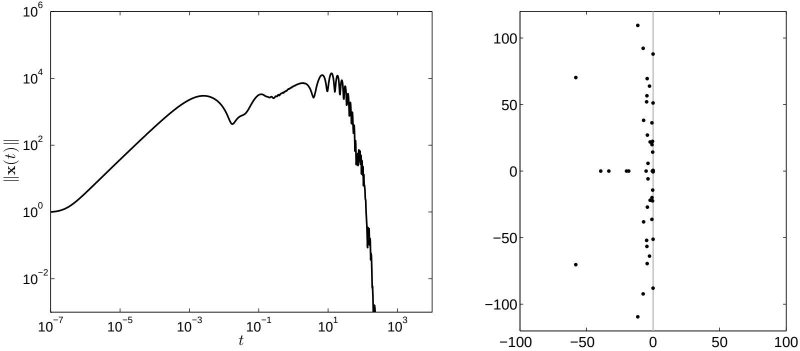
Model of flutter in a Boeing 767 from SLICOT ($n = 55$), stabilized by Burke, Lewis, Overton [2003].



$$\sigma_\varepsilon(\mathbf{A}) = \{z \in \mathbf{C} : \|(z - \mathbf{A})^{-1}\| > 1/\varepsilon\}$$

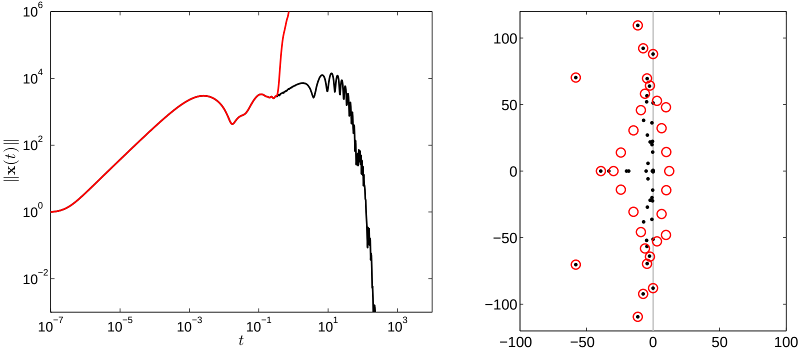
see [Trefethen & E. 2005]

Reduction via Restarted Arnoldi



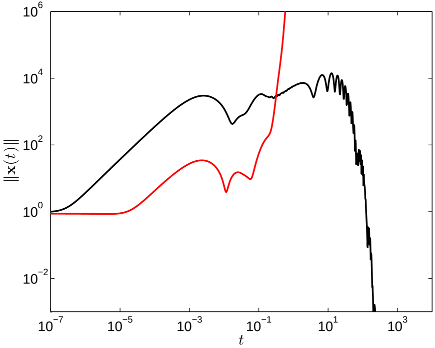
Transient behavior and spectrum of the original system

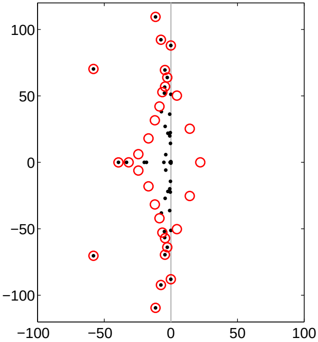
Reduction via Restarted Arnoldi



Transient behavior and spectrum for reduced system, $m = 40$

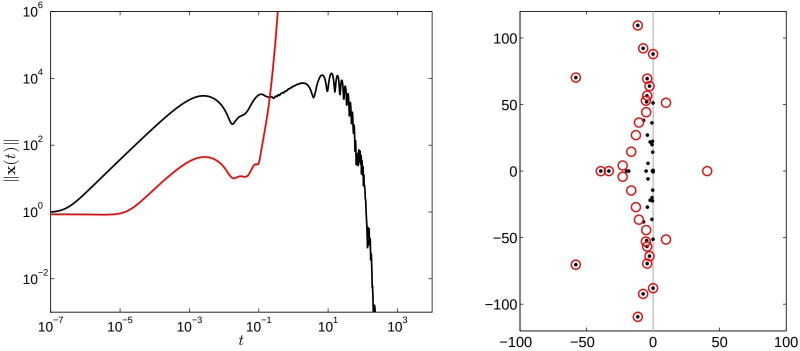
Reduction via Restarted Arnoldi





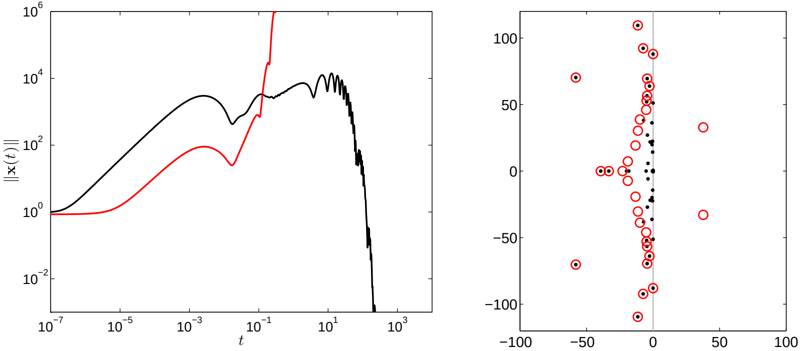
Transient behavior and spectrum for reduced system, $m = 40$,
after one implicit restart

Reduction via Restarted Arnoldi



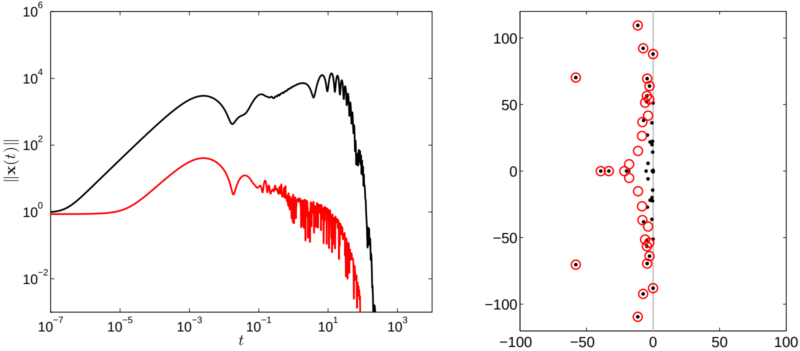
Transient behavior and spectrum for reduced system, $m = 40$,
after two implicit restarts

Reduction via Restarted Arnoldi



Transient behavior and spectrum for reduced system, $m = 40$, after three implicit restarts

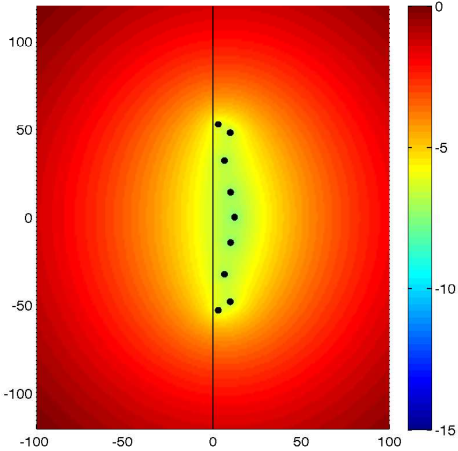
Reduction via Restarted Arnoldi



Transient behavior and spectrum for reduced system, $m = 40$,
after four implicit restarts

Reduced system is stable, but underestimates
transient growth by several orders of magnitude.

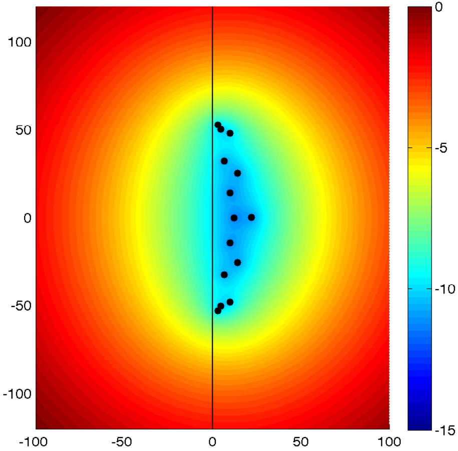
Aggregate Filter Polynomials



Color indicates relative size of $\log_{10} |\psi(z)|$.

first restart

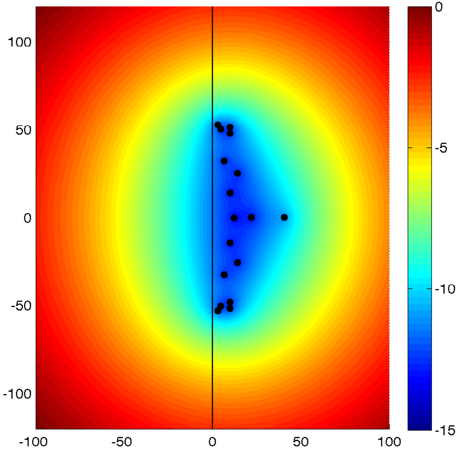
Aggregate Filter Polynomials



Color indicates relative size of $\log_{10} |\psi(z)|$.

second restart

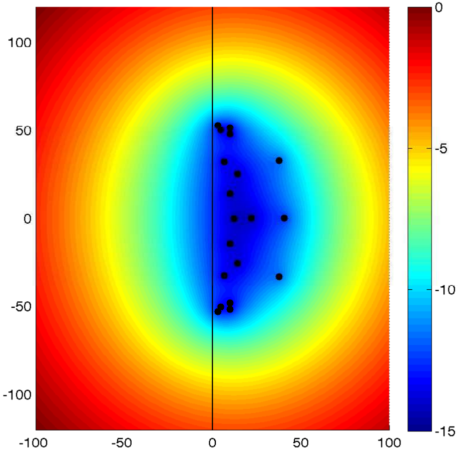
Aggregate Filter Polynomials



Color indicates relative size of $\log_{10} |\psi(z)|$.

third restart

Aggregate Filter Polynomials



Color indicates relative size of $\log_{10}|\psi(z)|$.

fourth restart

Nonlinear Heat Equation: Revisited

Linear models often arise as linearizations of nonlinear equations.

Consider the nonlinear heat equation on $x \in [-1, 1]$ with $u(-1, t) = u(1, t) = 0$

$$u_t(x,t) = \nu u_{xx}(x,t) + \sqrt{\nu} u_x(x,t) + \tfrac{1}{8} u(x,t) + u(x,t)^p$$

with $\nu > 0$ and $p > 1$ [Demanet, Holmer, Zworski].

The linearization L, an advection–diffusion operator,

$$Lu = \nu u_{xx} + \sqrt{\nu} u_x + \tfrac{1}{8} u$$

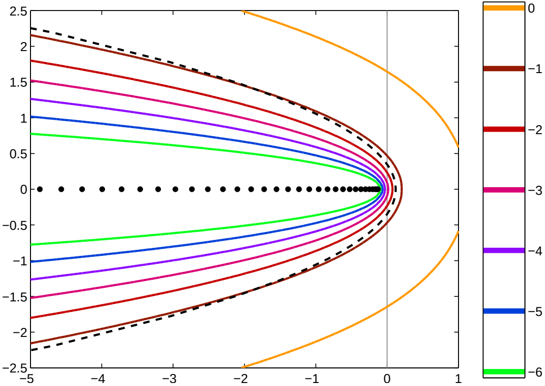
has eigenvalues and eigenfunctions

$$\lambda_n = -\frac{1}{8} - \frac{n^2\pi^2\nu}{4}, \qquad u_n(x) = e^{-x/(2\sqrt{\nu})}\sin(n\pi x/2);$$

see, e.g., [Reddy & Trefethen 1994].

The linearized operator is stable for all $\nu > 0$, but has interesting transients . . . .

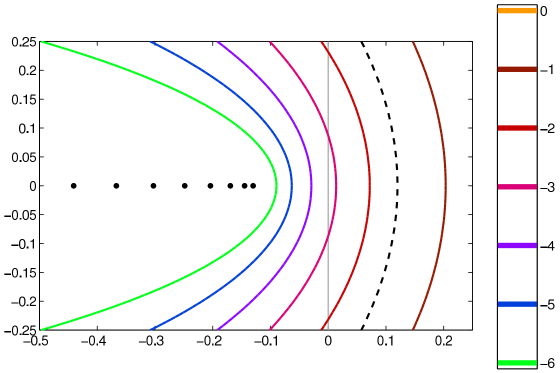
Nonnormality in the Linearization



Spectrum, pseudospectra, and numerical range (L^2 norm, $\nu = 0.02$)

Transient growth can feed the nonlinearity
cf. [Trefethen, Trefethen, Reddy, Driscoll 1993], ...

Nonnormality in the Linearization

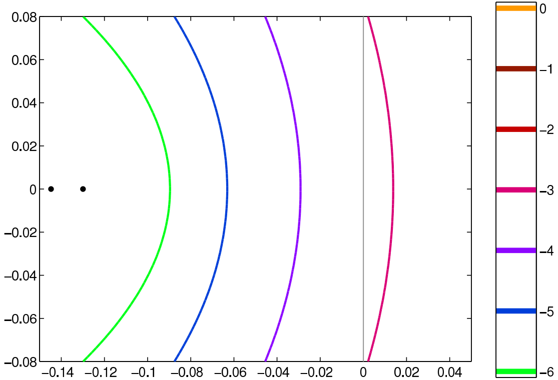


Spectrum, pseudospectra, and numerical range (L^2 norm, $\nu = 0.02$)

Transient growth can feed the nonlinearity

cf. [Trefethen, Trefethen, Reddy, Driscoll 1993], ...

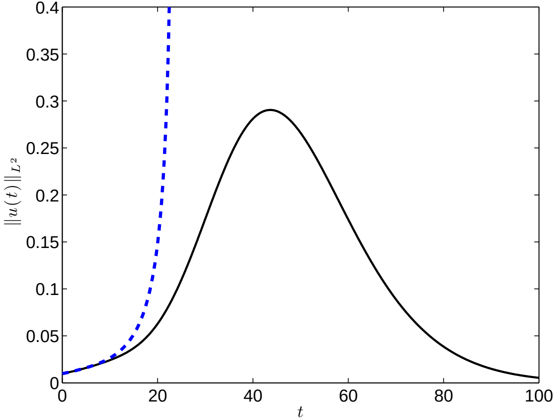
Nonnormality in the Linearization



Spectrum, pseudospectra, and numerical range (L^2 norm, $\nu = 0.02$)

Transient growth can feed the nonlinearity
cf. [Trefethen, Trefethen, Reddy, Driscoll 1993], ...

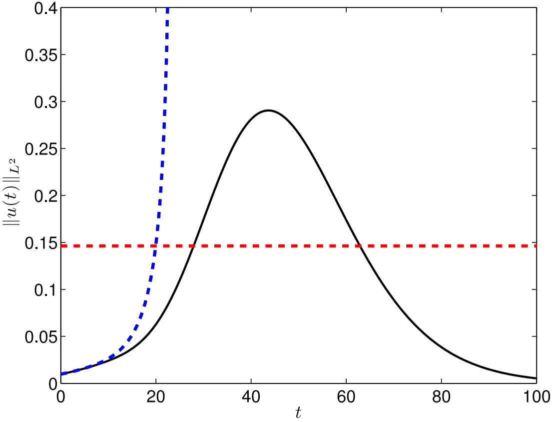
Transient Behavior



Linearized system (black) and nonlinear system (dashed blue)

Nonnormal growth feeds the nonlinear instability.

Transient Behavior

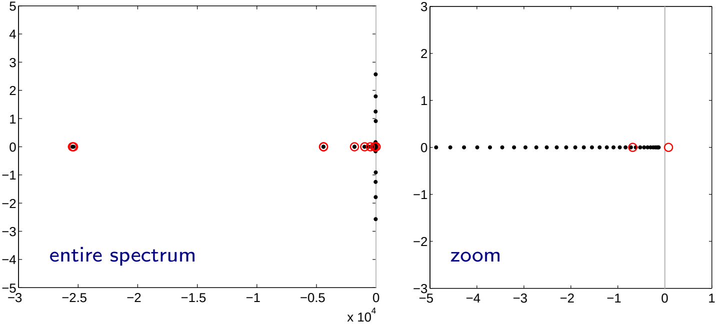


Linearized system (black) and nonlinear system (dashed blue)

Nonnormal growth feeds the nonlinear instability.

Red line: estimated “lower bound”: $z\|(z-\mathbf{A})^{-1}\|\|u_0(t)\|_{L^2}$ for $z = 0.02$.

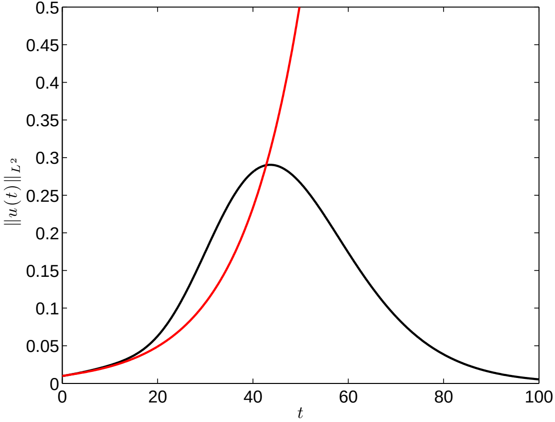
Transient Behavior: Reduction of the Linearized Model



Spectral discretization, $n = 128$ (black) and Arnoldi reduction, $m = 10$ (red).

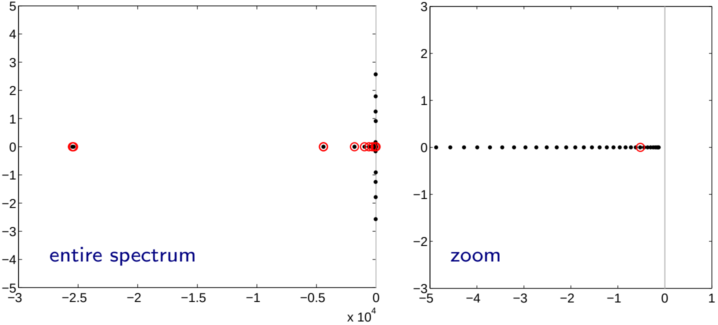
[Many Ritz values capture *spurious* eigenvalues of the discretization of the left.]

Transient Behavior: Reduction of the Linearized Model



Spectral discretization, $n = 128$ (black) and Arnoldi reduction, $m = 10$ (red).

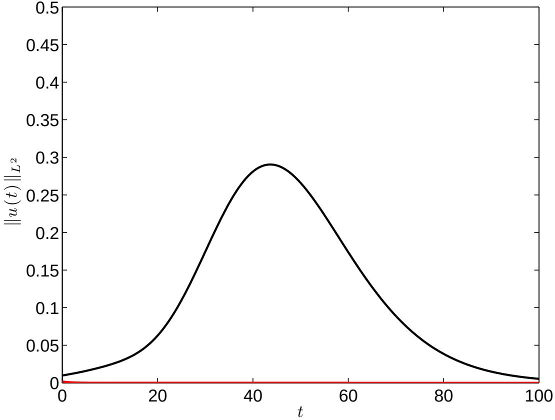
Transient Behavior: Reduction of the Linearized Model



Spectral discretization, $n = 128$ (black) and Arnoldi reduction, $m = 10$ (red) after a restart to remove the spurious eigenvalue.

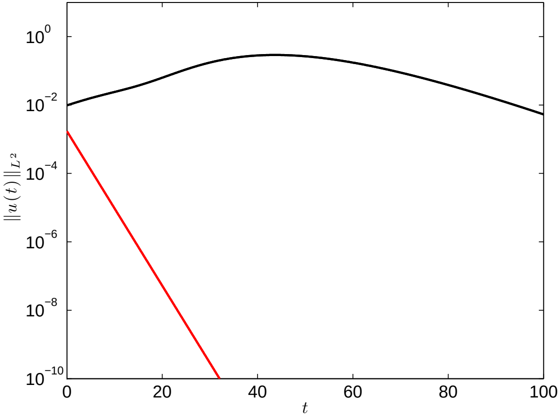
[This effectively pushes Ritz values to the left.]

Transient Behavior: Reduction of the Linearized Model



Spectral discretization, $n = 128$ (black) and Arnoldi reduction, $m = 10$ (red) after one restart to remove the spurious eigenvalue.

Transient Behavior: Reduction of the Linearized Model



Spectral discretization, $n = 128$ (black) and Arnoldi reduction, $m = 10$ (red) after one restart to remove the spurious eigenvalue.

4(c) Nonnormality in Iterative Linear Algebra

We wish to describe the role of nonnormality in the iterative solution of differential equations.

The GMRES algorithm produces optimal iterates $\mathbf{x}_k$ whose residuals $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$ satisfy

$$\|\mathbf{r}_k\| \le \min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \|p(\mathbf{A})\mathbf{b}\|,$$

where $\mathcal{P}_k$ denotes the set of polynomials of degree k or less.

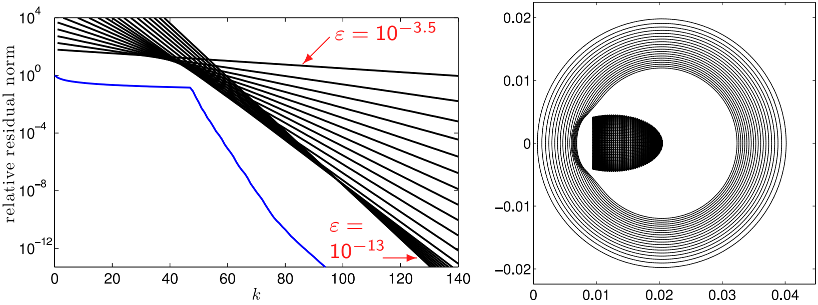
Applying the Cauchy integral theorem bound,

$$\frac{\|\mathbf{r}_k\|}{\|\mathbf{b}\|} \le \frac{L_\varepsilon}{2\pi\varepsilon} \min_{\substack{p \in \mathcal{P}_k \\ p(0)=1}} \sup_{z \in \sigma_\varepsilon(\mathbf{A})} |p(z)|.$$

Different ε values give the best bounds at different stages of convergence.

Illustrations and applications: [Trefethen 1990; E. 2000; Sifuentes, E., Morgan 2011].

GMRES Convergence: Example



Pseudospectral bound for convection–diffusion problem, $n = 2304$.

$$\varepsilon = 10^{-3.5}, 10^{-3}, 10^{-3.5}, \ldots, 10^{-13}$$

Bound $\sigma_\varepsilon(\mathbf{A})$ with a circle centered at c, use $p_k(z) = (1 - z/c)^k$.

Perturbed GMRES and Approximate Preconditioning

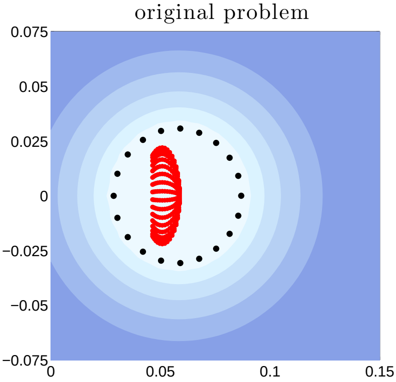
- We shall show how psedospectra can be used to bound GMRES convergence when the coefficient matrix $\mathbf{A}$ is slightly perturbed. [Sifuentes, E., Morgan, 2011]
- The primary application result of this work is *approximate preconditioning*, where we aim to solve $\mathbf{Ax} = \mathbf{b}$ by the modified system

  $$\mathbf{PAx} = \mathbf{Pb}$$

  for some invertible $\mathbf{P}$.
- Often one knows an "ideal" preconditioner that is expensive to compute, but for which $\sigma(\mathbf{PA})$ is easy to analyze. In practice, $\mathbf{P}$ is only applied approximately. How accurate must it be to provide similar convergence to the exact preconditioner?

An illustrative example

2d Advection–Diffusion discretized with SUPG
[Fischer, Ramage, Silvester, Wathen 1999]
[Ernst 2000], [Liesen, Strakoš 2005]

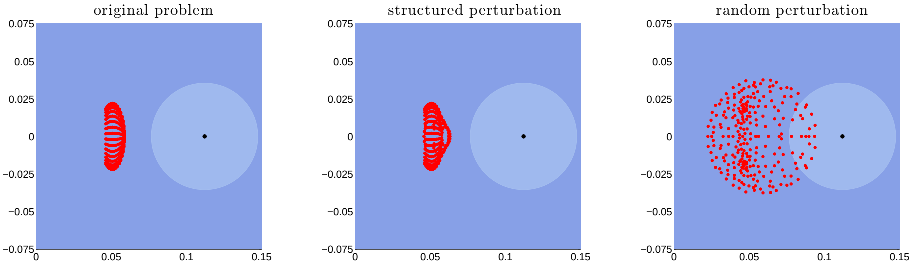


Matrix is of order $n = 16^2$.
$\mathbf{A} + \mathbf{E}_1$: slightly change wind direction
$\mathbf{A} + \mathbf{E}_2$: random perturbation
$\|\mathbf{E}_1\| = \|\mathbf{E}_2\| \approx 5.86 \times 10^{-4}$

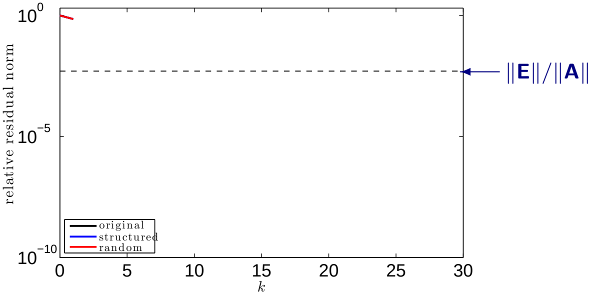
- eigenvalues of coefficient matrix
- harmonic Ritz values (roots of GMRES residual polynomial)

  blue color denotes $\log_{10} |p_k(z)|$ in the complex plane (white $\Longrightarrow$ smaller)

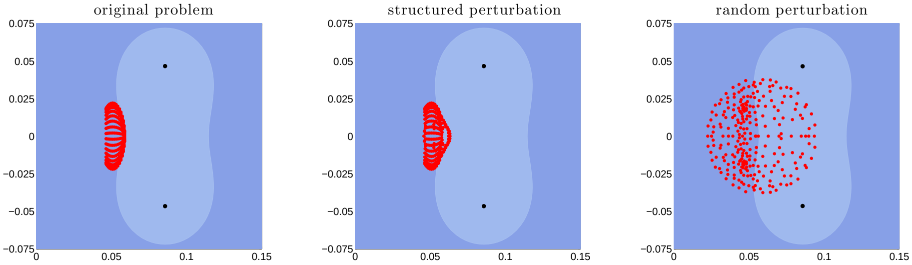
An illustrative example



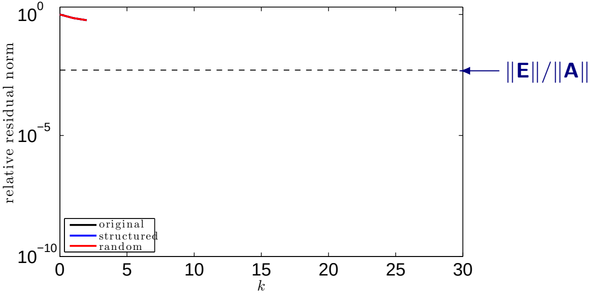
Iteration 1



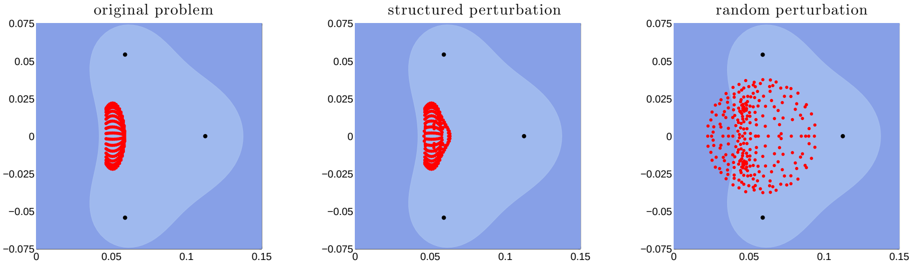
An illustrative example



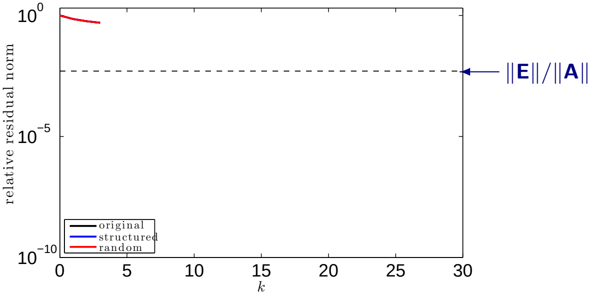
Iteration 2



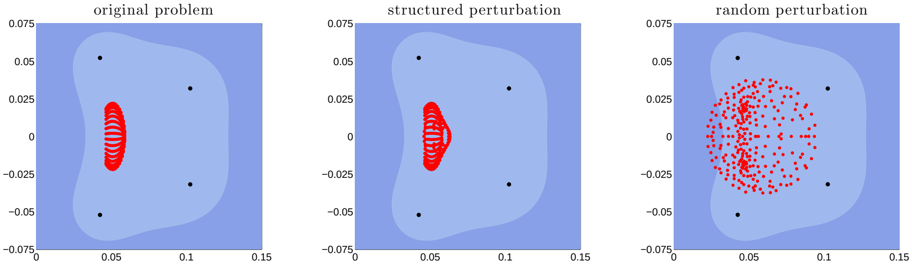
An illustrative example



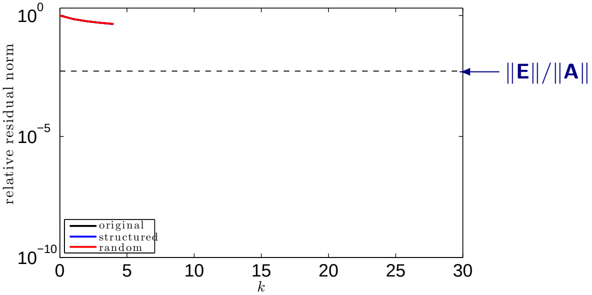
Iteration 3



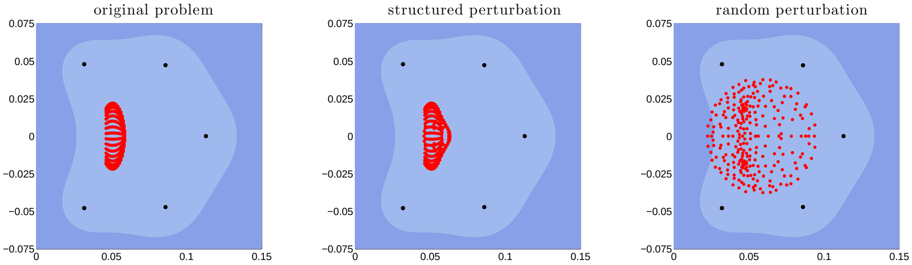
An illustrative example



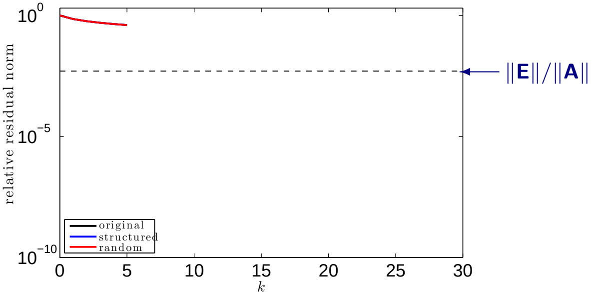
Iteration 4



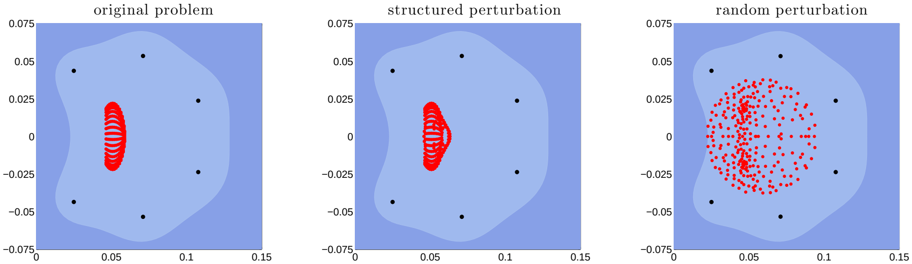
An illustrative example



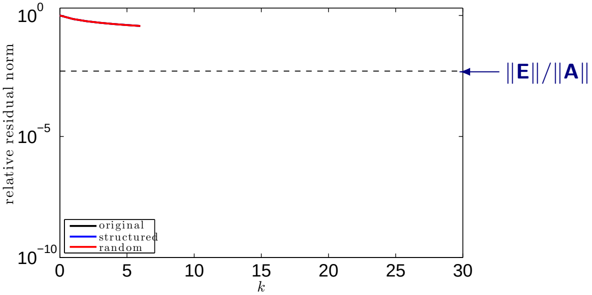
Iteration 5



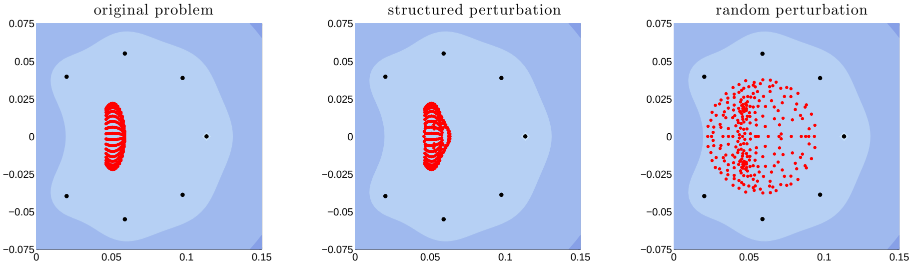
An illustrative example



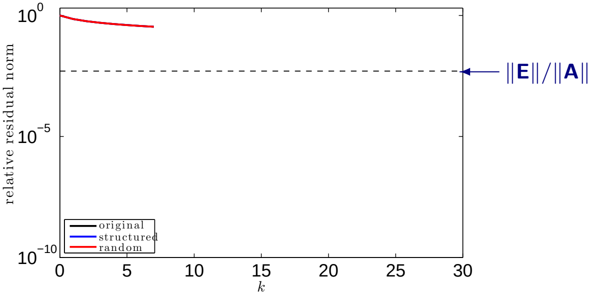
Iteration 6



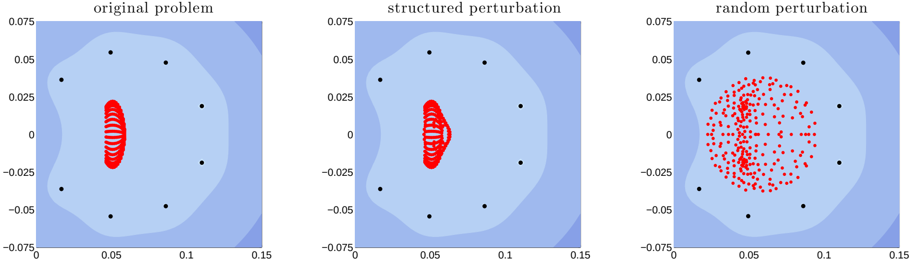
An illustrative example



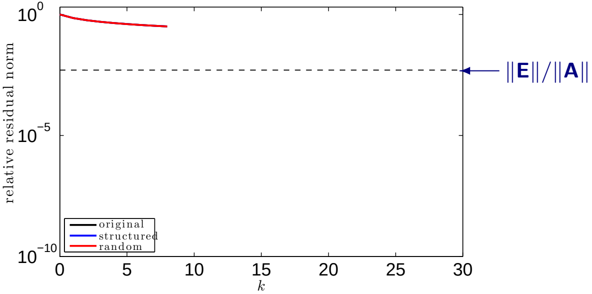
Iteration 7



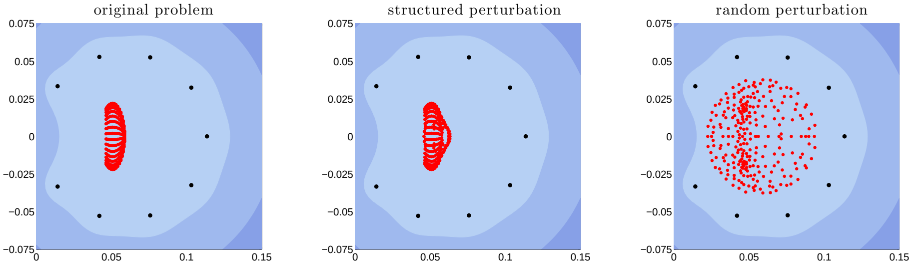
An illustrative example



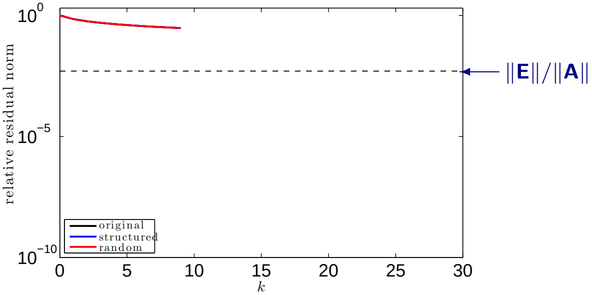
Iteration 8



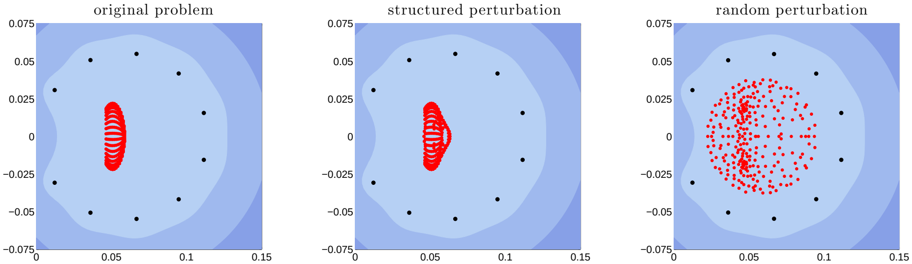
An illustrative example



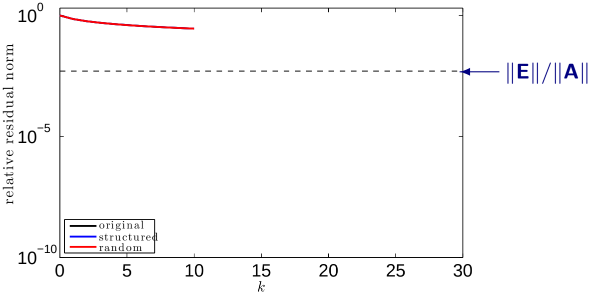
Iteration 9



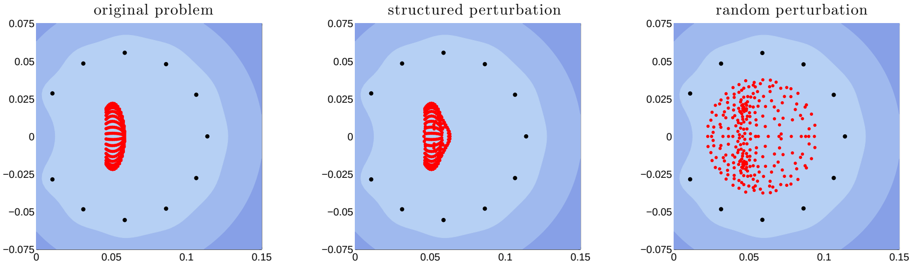
An illustrative example



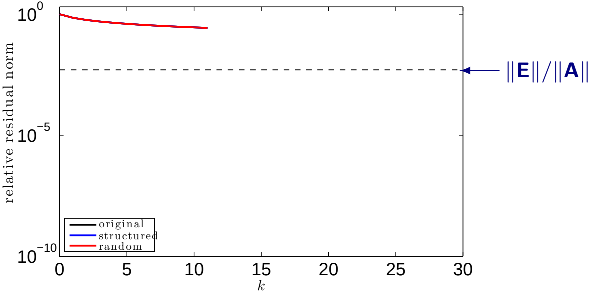
Iteration 10



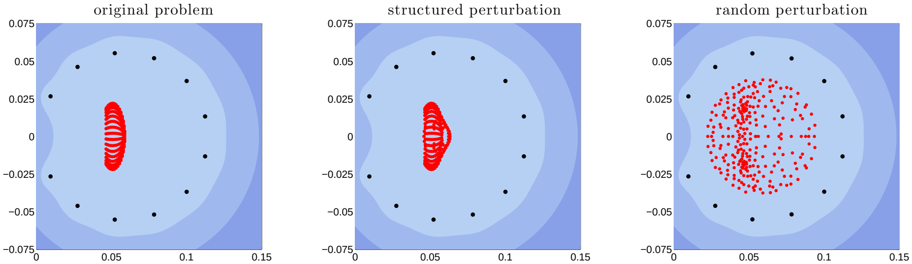
An illustrative example



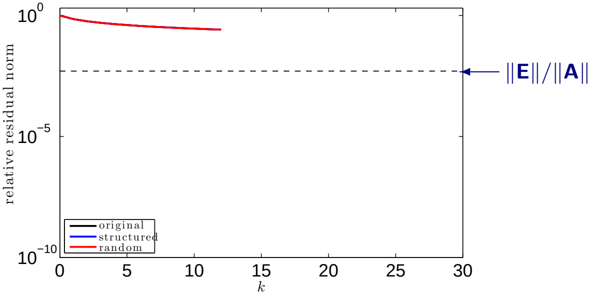
Iteration 11



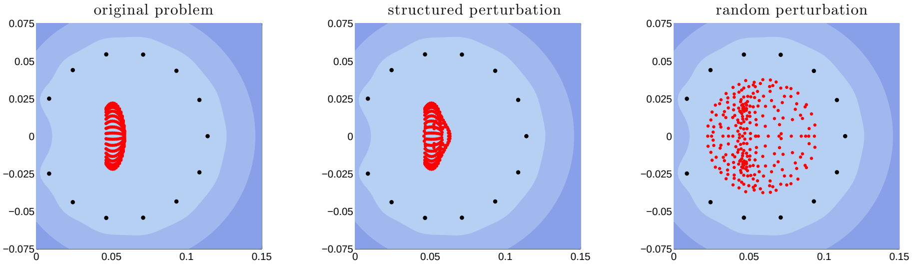
An illustrative example



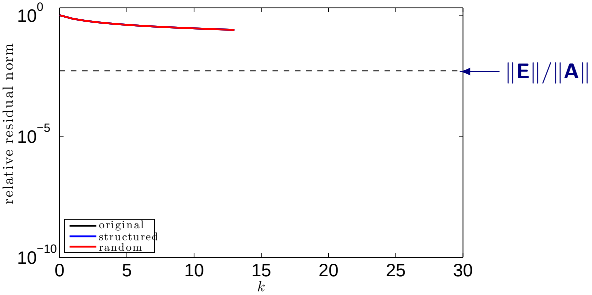
Iteration 12



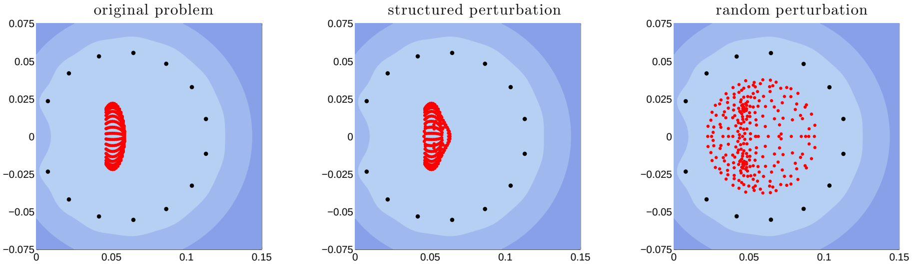
An illustrative example



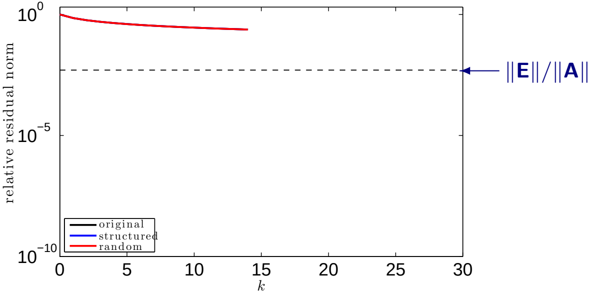
Iteration 13



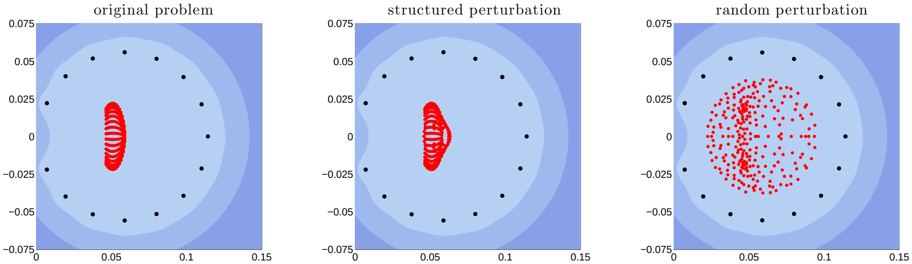
An illustrative example



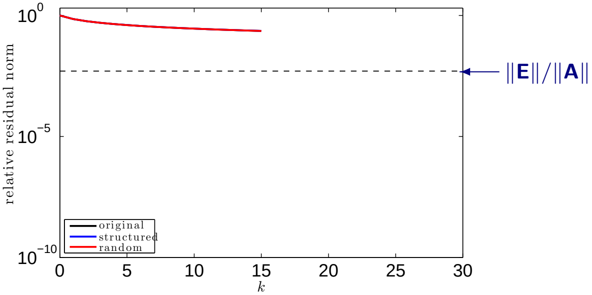
Iteration 14



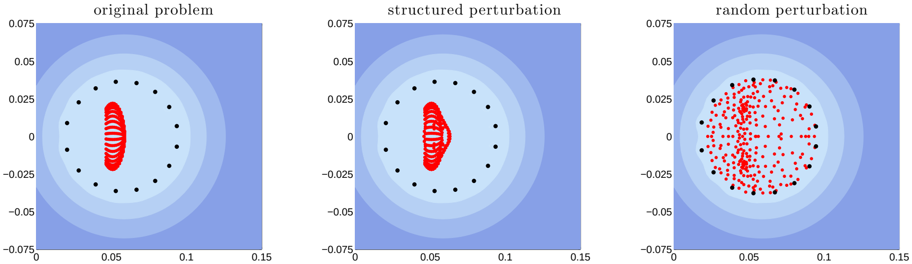
An illustrative example



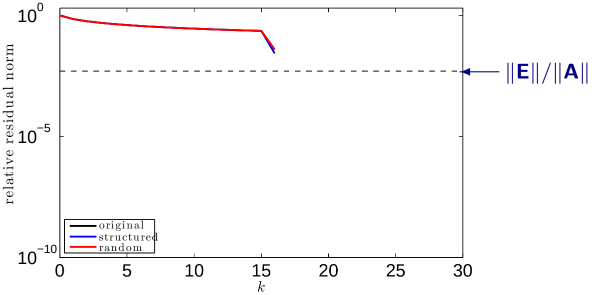
Iteration 15



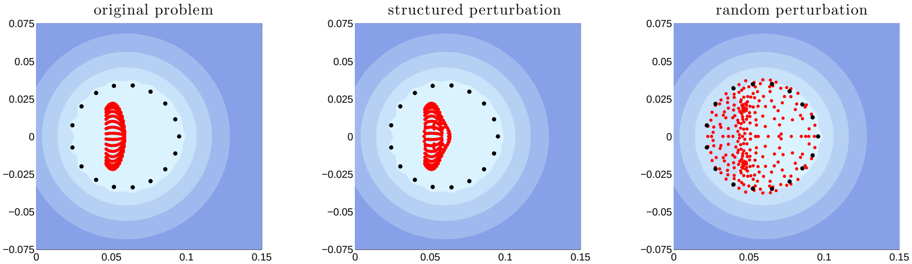
An illustrative example



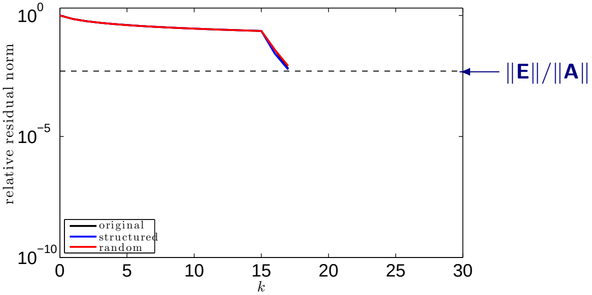
Iteration 16



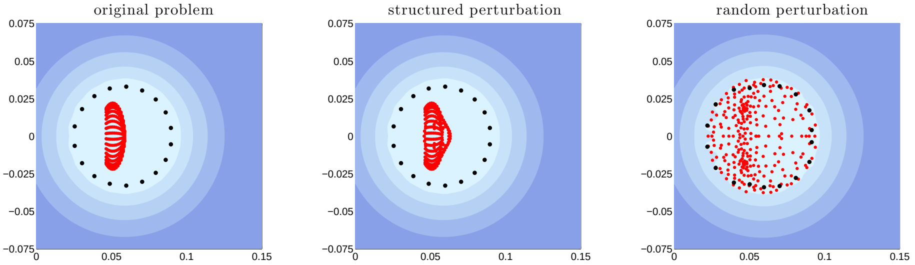
An illustrative example



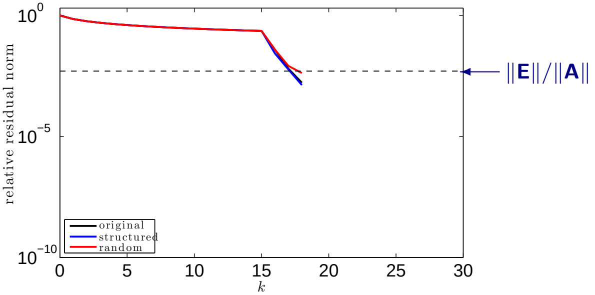
Iteration 17



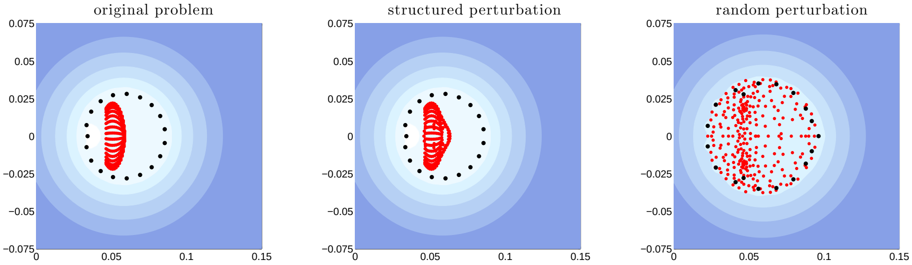
An illustrative example



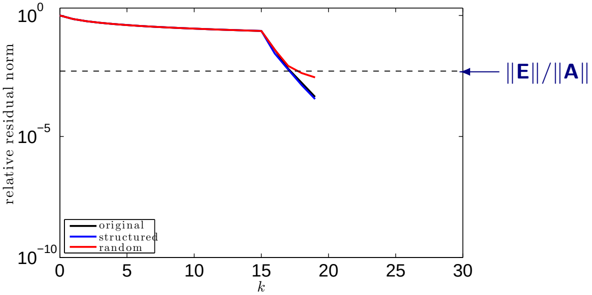
Iteration 18



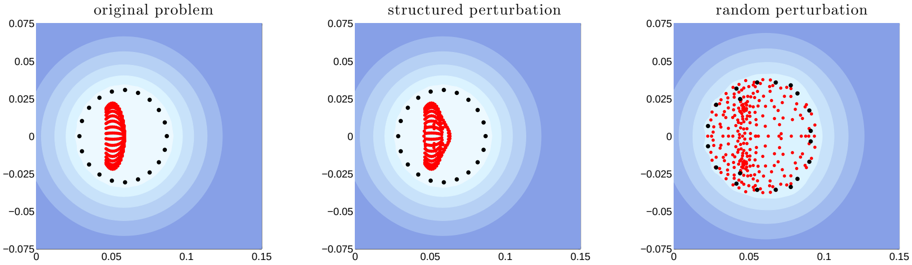
An illustrative example



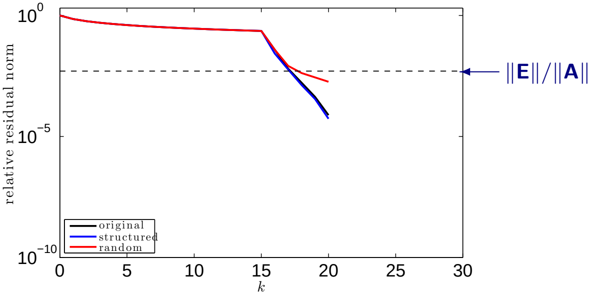
Iteration 19



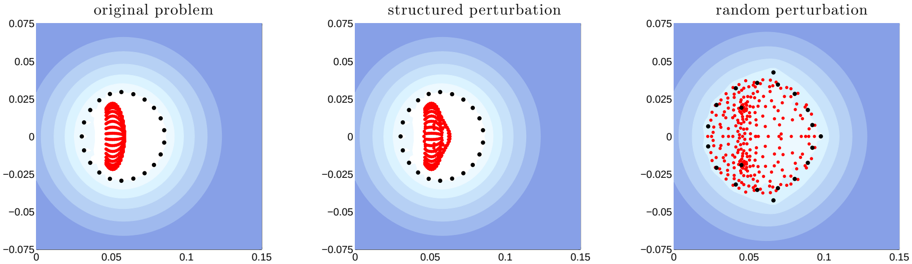
An illustrative example



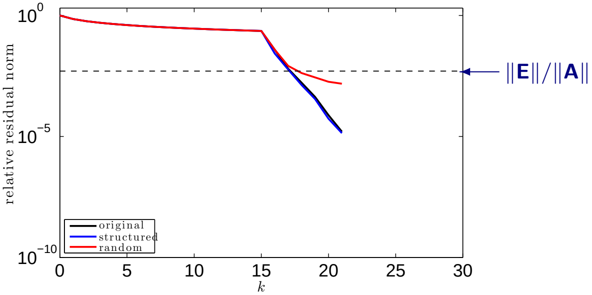
Iteration 20



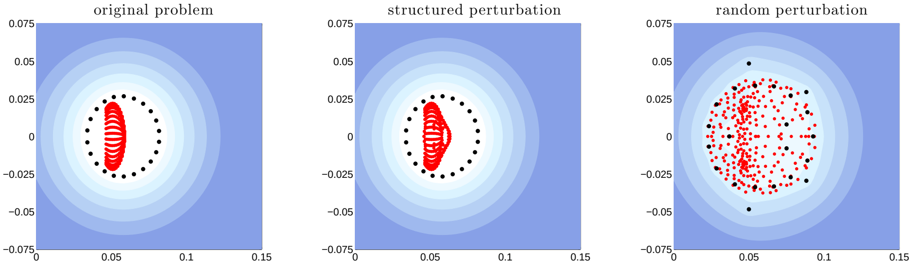
An illustrative example



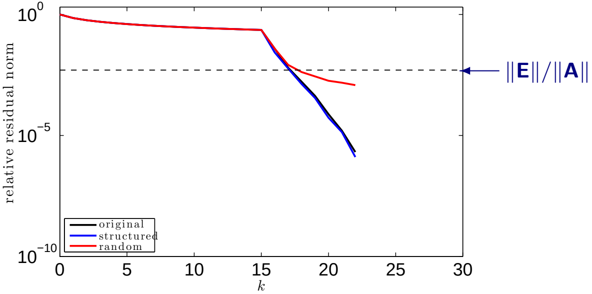
Iteration 21



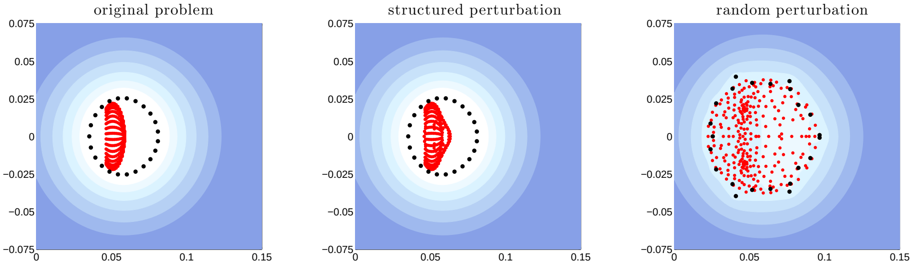
An illustrative example



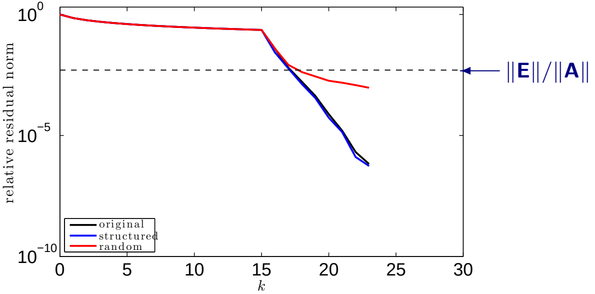
Iteration 22



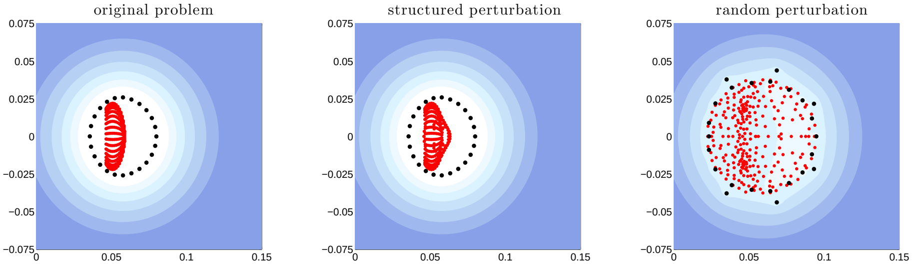
An illustrative example



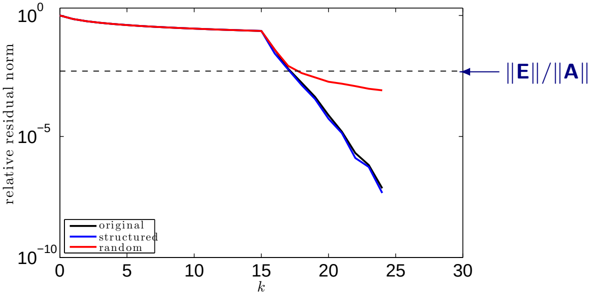
Iteration 23



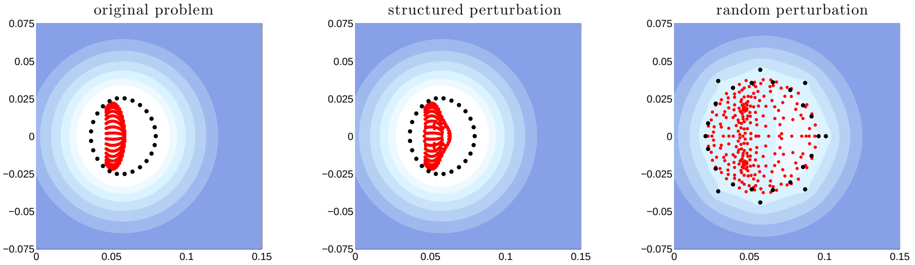
An illustrative example



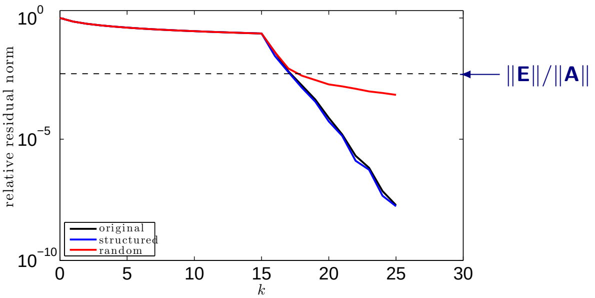
Iteration 24



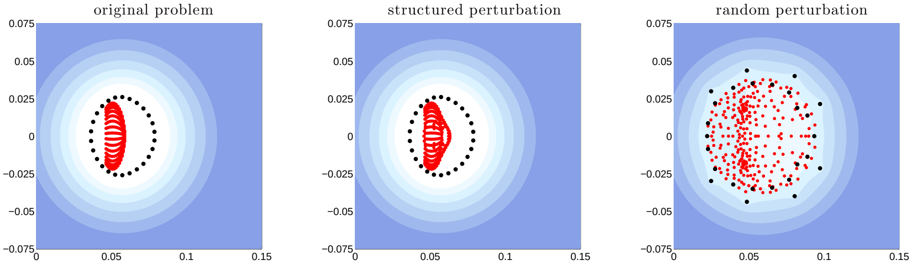
An illustrative example



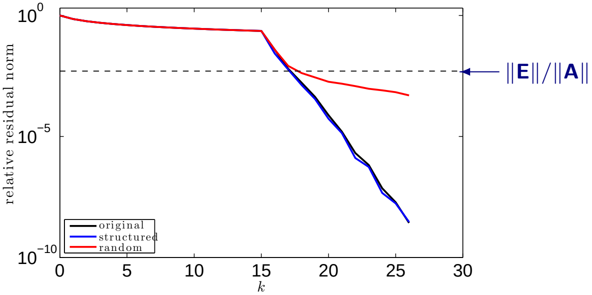
Iteration 25



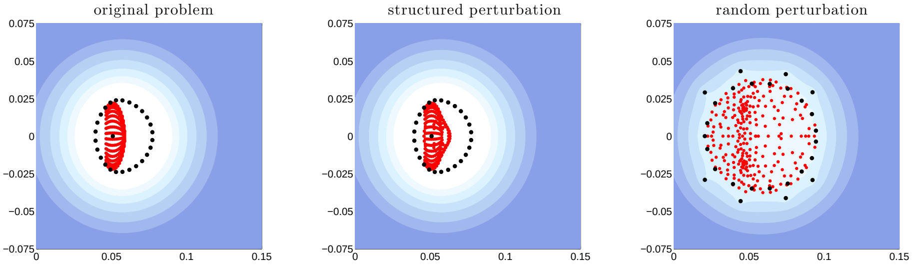
An illustrative example



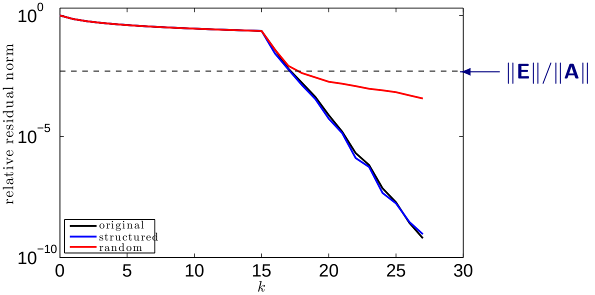
Iteration 26



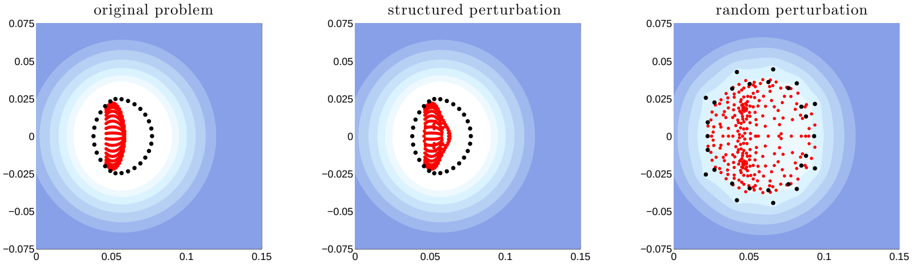
An illustrative example



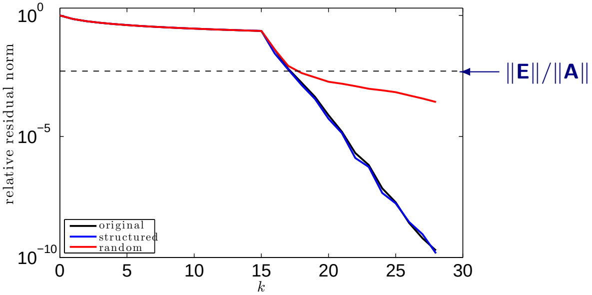
Iteration 27



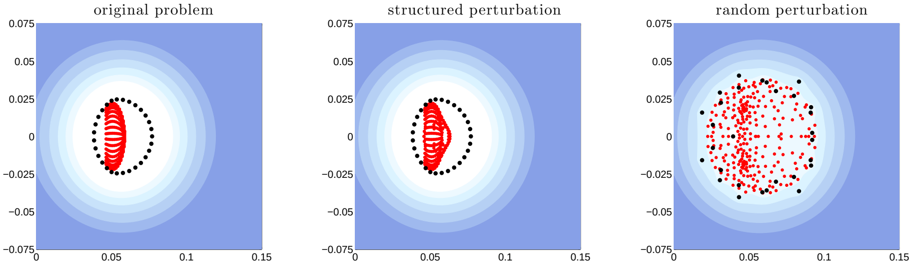
An illustrative example



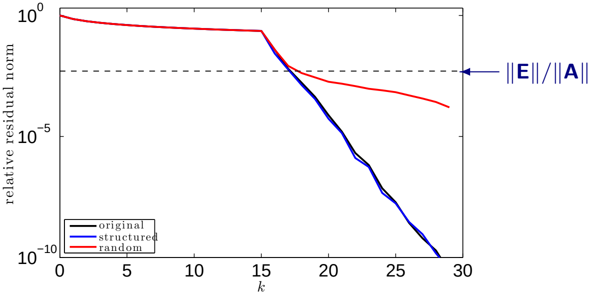
Iteration 28



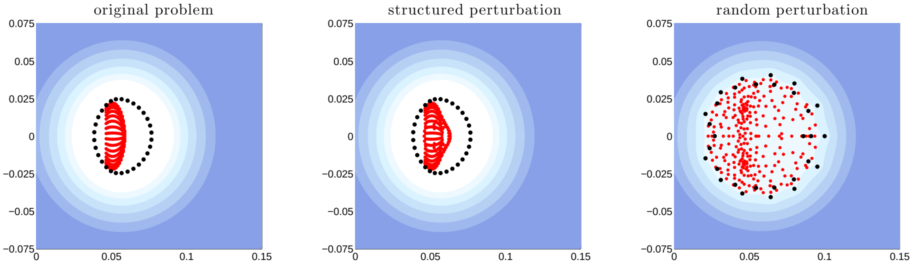
An illustrative example



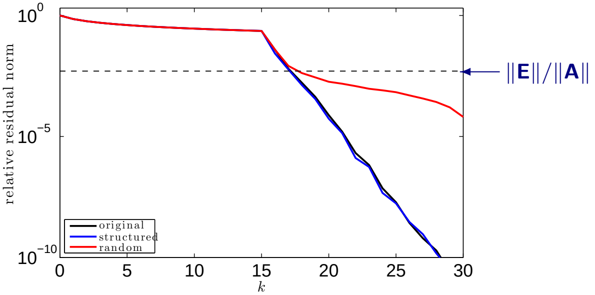
Iteration 29



An illustrative example



Iteration 30



Analysis via Cauchy integrals

Define

$$\begin{aligned} \textbf{original problem:} \quad & \mathbf{r}_k = p_k(\mathbf{A})\mathbf{r}_0 \\ \textbf{perturbed problem:} \quad & \boldsymbol{\rho}_k = \phi_k(\mathbf{A}+\mathbf{E})\mathbf{r}_0 \end{aligned}$$

Note that

$$\begin{aligned} \|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\| &= \|\phi_k(\mathbf{A}+\mathbf{E})\mathbf{r}_0\| - \|p_k(\mathbf{A})\mathbf{r}_0\| \\ &\le \|p_k(\mathbf{A}+\mathbf{E})\mathbf{r}_0\| - \|p_k(\mathbf{A})\mathbf{r}_0\| \\ &\le \big\|\big(p_k(\mathbf{A}+\mathbf{E}) - p_k(\mathbf{A})\big)\mathbf{r}_0\big\|. \end{aligned}$$

Analysis via Cauchy integrals

Define

$$\textbf{original problem:} \quad \mathbf{r}_k = p_k(\mathbf{A})\mathbf{r}_0$$
$$\textbf{perturbed problem:} \quad \boldsymbol{\rho}_k = \phi_k(\mathbf{A}+\mathbf{E})\mathbf{r}_0$$

Note that

$$\begin{aligned}\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\| &= \|\phi_k(\mathbf{A}+\mathbf{E})\mathbf{r}_0\| - \|p_k(\mathbf{A})\mathbf{r}_0\| \\ &\le \|p_k(\mathbf{A}+\mathbf{E})\mathbf{r}_0\| - \|p_k(\mathbf{A})\mathbf{r}_0\| \\ &\le \big\|\big(p_k(\mathbf{A}+\mathbf{E}) - p_k(\mathbf{A})\big)\mathbf{r}_0\big\|.\end{aligned}$$

For *appropriate* Γ, we write

$$\begin{aligned}p_k(\mathbf{A}) &= \frac{1}{2\pi i}\int_\Gamma p_k(z)(z-\mathbf{A})^{-1}\,\mathrm{d}z \\ p_k(\mathbf{A}+\mathbf{E}) &= \frac{1}{2\pi i}\int_\Gamma p_k(z)(z-\mathbf{A}-\mathbf{E})^{-1}\,\mathrm{d}z.\end{aligned}$$

Sharper analysis via Cauchy integrals

$$p_k(\mathbf{A}) = \frac{1}{2\pi i}\int_\Gamma p_k(z)(z-\mathbf{A})^{-1}\,\mathrm{d}z$$
$$p_k(\mathbf{A}+\mathbf{E}) = \frac{1}{2\pi i}\int_\Gamma p_k(z)(z-\mathbf{A}-\mathbf{E})^{-1}\,\mathrm{d}z.$$

Key fact: the resolvent is robust to perturbations if Γ is chosen properly:

$$\begin{aligned}(z-\mathbf{A}-\mathbf{E})^{-1} &= \big((z-\mathbf{A})(\mathbf{I}-(z-\mathbf{A})^{-1}\mathbf{E})\big)^{-1}\\ &= (\mathbf{I}-(z-\mathbf{A})^{-1}\mathbf{E})^{-1}(z-\mathbf{A})^{-1}.\end{aligned}$$

Sharper analysis via Cauchy integrals

$$\begin{aligned} p_k(\mathbf{A}) &= \frac{1}{2\pi i}\int_\Gamma p_k(z)(z-\mathbf{A})^{-1}\,\mathrm{d}z \\ p_k(\mathbf{A}+\mathbf{E}) &= \frac{1}{2\pi i}\int_\Gamma p_k(z)(z-\mathbf{A}-\mathbf{E})^{-1}\,\mathrm{d}z. \end{aligned}$$

Key fact: the resolvent is robust to perturbations if Γ is chosen properly:

$$\begin{aligned} (z-\mathbf{A}-\mathbf{E})^{-1} &= \Big((z-\mathbf{A})(\mathbf{I}-(z-\mathbf{A})^{-1}\mathbf{E})\Big)^{-1} \\ &= (\mathbf{I}-(z-\mathbf{A})^{-1}\mathbf{E})^{-1}(z-\mathbf{A})^{-1}. \end{aligned}$$

Our bound thus becomes:

$$\frac{\|\boldsymbol{\rho}_k\|-\|\mathbf{r}_k\|}{\|\mathbf{r}_0\|} \le \frac{L_\Gamma}{2\pi}\Big(\max_{z\in\Gamma}|p_k(z)|\Big)\Big(\max_{z\in\Gamma}\|(\mathbf{I}-(z-\mathbf{A})^{-1}\mathbf{E})^{-1}\|\Big)\Big(\max_{z\in\Gamma}\|(z-\mathbf{A})^{-1}\|\Big).$$

Bounding the perturbed resolvent

Two options $\|(\mathbf{I} - (z-\mathbf{A})^{-1}\mathbf{E})^{-1}\|$:

- if $\mathrm{rank}(\mathbf{E}) \ll n$, use Sherman-Morrison;
- if $\|\mathbf{E}\| < 1/\|(z-\mathbf{A})^{-1}\|$ for $z \in \Gamma$, then use a Neumann series:

$$\begin{aligned} \|(\mathbf{I} - (z-\mathbf{A})^{-1}\mathbf{E})^{-1}\| &= \Big\| \sum_{k=1}^{\infty} (z-\mathbf{A})^{-1}\mathbf{E} \,\Big\| \\ &= \frac{\|\mathbf{E}\| \, \|(z-\mathbf{A})^{-1}\|}{1 - \|\mathbf{E}\| \, \|(z-\mathbf{A})^{-1}\|}. \end{aligned}$$

Thus, our bound takes the form

$$\frac{\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\|}{\|\mathbf{r}_0\|} \le \frac{L_\Gamma}{2\pi} \Big(\max_{z\in\Gamma} |p_k(z)|\Big) \Big(\max_{z\in\Gamma} \frac{\|\mathbf{E}\| \, \|(z-\mathbf{A})^{-1}\|^2}{1 - \|\mathbf{E}\| \, \|(z-\mathbf{A})^{-1}\|}\Big).$$

For this approach to perturbed matrix functions:
R. F. Rinehart, "The Derivative of a Matric Function," *Proc. A.M.S.*, 1956.
E. B. Davies, "Approximate Diagonalization," *SIAM J. Matrix Anal.*, 2007.

How to select the contour Γ ?

$$\frac{\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\|}{\|\mathbf{r}_0\|} \leq \frac{L_\Gamma}{2\pi}\left(\max_{z\in\Gamma}|p_k(z)|\right)\left(\max_{z\in\Gamma}\frac{\|\mathbf{E}\|\,\|(z-\mathbf{A})^{-1}\|^2}{1-\|\mathbf{E}\|\,\|(z-\mathbf{A})^{-1}\|}\right).$$

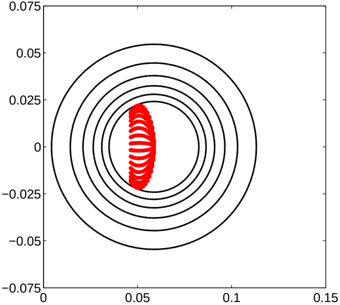
We have competing goals that affect the selection of Γ:

- Γ should be close enough to the spectrum of $\mathbf{A}$ that $|p_k(z)|$ is small;
- Γ should be far enough from the spectrum of $\mathbf{A}$ that $\|(z-\mathbf{A})^{-1}\|$ is small.

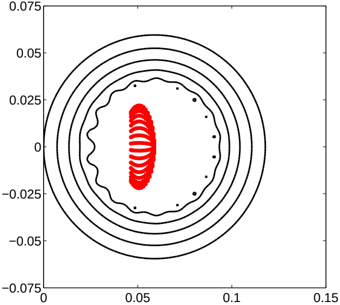
How to select the contour Γ ?

$$\frac{\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\|}{\|\mathbf{r}_0\|} \le \frac{L_\Gamma}{2\pi}\left(\max_{z\in\Gamma}|p_k(z)|\right)\left(\max_{z\in\Gamma}\frac{\|\mathbf{E}\|\,\|(z-\mathbf{A})^{-1}\|^2}{1-\|\mathbf{E}\|\,\|(z-\mathbf{A})^{-1}\|}\right).$$

Take Γ to be the boundary of the δ-pseudospectrum, $\sigma_\delta(\mathbf{A})$.



level sets of $\|(z-\mathbf{A})^{-1}\|$
$\sigma_\delta(\mathbf{A})$, $\delta = 10^{-2.5}, 10^{-3.5}, \ldots, 10^{-7.5}$



level sets of $|p_{18}(z)|$
$|p_{18}(z)| = 10^0, 10^{-1}, \ldots, 10^{-5}$

cf. [Trefethen 1990; Toh & Trefethen 1998]

Perturbation bounds for GMRES based on pseudospectra

$$\frac{\|\boldsymbol{\rho}_k\| - \|\mathbf{r}_k\|}{\|\mathbf{r}_0\|} \leq \frac{L_\Gamma}{2\pi}\left(\max_{z\in\Gamma}|p_k(z)|\right)\left(\max_{z\in\Gamma}\frac{\|\mathbf{E}\|\,\|(z-\mathbf{A})^{-1}\|^2}{1-\|\mathbf{E}\|\,\|(z-\mathbf{A})^{-1}\|}\right).$$

Select $\Gamma = \partial\sigma_\delta(\mathbf{A})$ for some $\delta > \varepsilon := \|\mathbf{E}\|$:

$$\boxed{\frac{\|\boldsymbol{\rho}_k\|}{\|\mathbf{r}_0\|} \;\leq\; \frac{\|\mathbf{r}_k\|}{\|\mathbf{r}_0\|} + \frac{L_\delta}{2\pi\delta}\left(\frac{\varepsilon}{\delta-\varepsilon}\right)\left(\max_{z\in\sigma_\delta(\mathbf{A})}|p_k(z)|\right).}$$

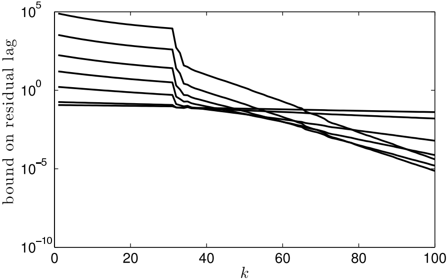
Note that $\sigma(\mathbf{A}+\mathbf{E}) \subset \sigma_\varepsilon(\mathbf{A}) \subset \sigma_\delta(\mathbf{A})$.

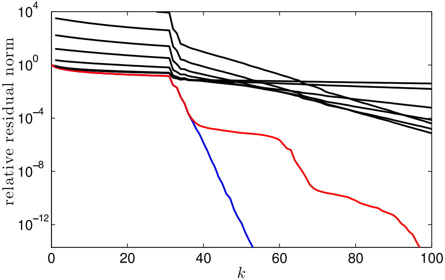
Alternatively, we can bound $\|\boldsymbol{\rho}_k\|$ above independent of the specific p_k:

$$\boxed{\frac{\|\boldsymbol{\rho}_k\|}{\|\mathbf{r}_0\|} \;\leq\; \frac{L_\delta}{2\pi\delta}\left(\frac{\delta}{\delta-\varepsilon}\right)\left(\min_{\substack{p\in\mathcal{P}_k\\ p(0)=1}}\max_{z\in\sigma_\delta(\mathbf{A})}|p(z)|\right).}$$

Example: SUPG Matrix for N = 32, $\nu = 0.01$

$$N = 32,\ \ \nu = 0.01 \quad \Longrightarrow \quad \mathbf{A} \in \mathbf{R}^{1024 \times 1024}.$$



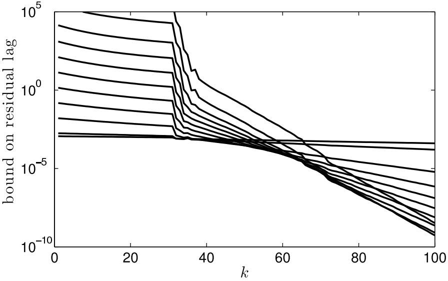


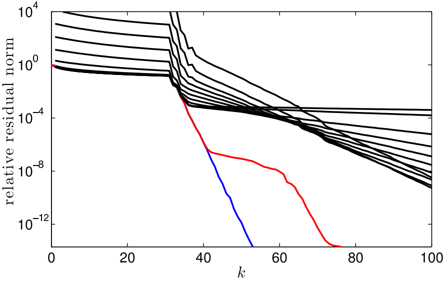
perturbation size: $\|\mathbf{E}\| = 10^{-4}\|\mathbf{A}\|$

- original GMRES residual curve for $\mathbf{A}\mathbf{x} = \mathbf{b}$
- perturbed GMRES residual curve for $(\mathbf{A} + \mathbf{E})\mathbf{x} = \mathbf{b}$
- upper bound on perturbed residual
  $(\delta = 10^{-5.2}, 10^{-5}, 10^{-4.5}, \ldots, 10^{-3}, 10^{-2.9})$

Example: SUPG Matrix for N = 32, $\nu = 0.01$

$$N = 32,\ \nu = 0.01 \quad \Longrightarrow \quad \mathbf{A} \in \mathbb{R}^{1024\times 1024}.$$



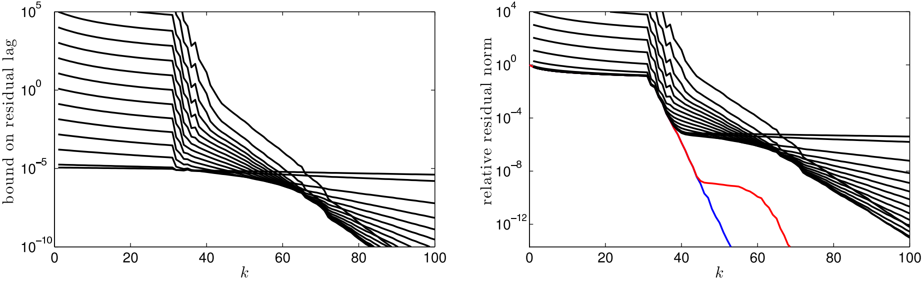


perturbation size: $\|\mathbf{E}\| = 10^{-6}\|\mathbf{A}\|$

- original GMRES residual curve for $\mathbf{A}\mathbf{x} = \mathbf{b}$
- perturbed GMRES residual curve for $(\mathbf{A}+\mathbf{E})\mathbf{x} = \mathbf{b}$
- upper bound on perturbed residual
  $(\delta = 10^{-7.2}, 10^{-7}, 10^{-6.5}, \ldots, 10^{-3}, 10^{-2.9})$

Example: SUPG Matrix for $\mathbf{N = 32}$, $\nu = 0.01$

$$N = 32,\ \nu = 0.01 \quad \Longrightarrow \quad \mathbf{A} \in \mathbf{R}^{1024 \times 1024}.$$



perturbation size: $\|\mathbf{E}\| = 10^{-8}\|\mathbf{A}\|$

- original GMRES residual curve for $\mathbf{Ax} = \mathbf{b}$
- perturbed GMRES residual curve for $(\mathbf{A} + \mathbf{E})\mathbf{x} = \mathbf{b}$
- upper bound on perturbed residual
  $(\delta = 10^{-9.2}, 10^{-9}, 10^{-8.5}, \ldots, 10^{-3}, 10^{-2.9})$

Nonnormality in Eigenvalue Computations

Here we shall only consider the simplest iterative method for computing eigenvalues: the power method.

Given a starting vector $\mathbf{x}_0$, repeatedly apply $\mathbf{A}$:

$$\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0.$$

Nonnormality in Eigenvalue Computations

Here we shall only consider the simplest iterative method for computing eigenvalues: the power method.

Given a starting vector $\mathbf{x}_0$, repeatedly apply $\mathbf{A}$:

$$\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0.$$

Assume $\mathbf{A}$ is diagonalizable,

$$\mathbf{A} = \sum_{j=1}^{n} \lambda_j \mathbf{v}_j \widehat{\mathbf{v}}_j^*$$

with eigenvalues

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Then if $\mathbf{x}_0 = \sum \gamma_j \mathbf{v}_j$, then

$$\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0 = \sum_{j=1}^{n} \lambda_j^k \mathbf{v}_j.$$

Hence if $\gamma_1 \neq 0$, as $k \to \infty$

$$\angle(\mathbf{x}_k, \mathbf{v}_1) \to 0$$

at the asymptotic rate $|\lambda_2|/|/|\lambda_1|$.

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,

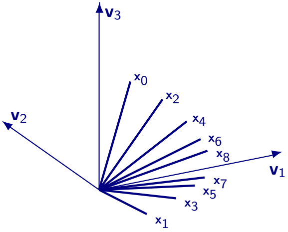
$$\mathbf{A} = \begin{bmatrix} 1 & \alpha & 0 \\ 0 & 3/4 & \beta \\ 0 & 0 & -3/4 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -4\alpha \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 8\alpha\beta/21 \\ -2\beta/3 \\ 1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

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,

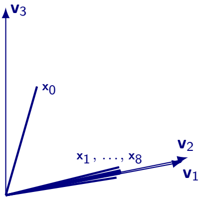
$$\mathbf{A} = \begin{bmatrix} 1 & \alpha & 0 \\ 0 & 3/4 & \beta \\ 0 & 0 & -3/4 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -4\alpha \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 8\alpha\beta/21 \\ -2\beta/3 \\ 1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



$\alpha = \beta = 0$

normal

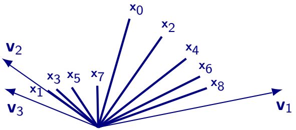
$\mathbf{v}_1 \perp \mathbf{v}_2 \perp \mathbf{v}_3$



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

nonnormal

$\mathbf{v}_1 \not\perp \mathbf{v}_2 \perp \mathbf{v}_3$



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

nonnormal

$\mathbf{v}_1 \perp \mathbf{v}_2 \not\perp \mathbf{v}_3$

Cf. [Beattie, E., Rossi 2004; Beattie, E., Sorensen 2005]

4(d) Pseudospectra of Operators

Pseudospectra with More General Norms

Recall our definition of pseudospectra of a matrix from Lecture 1.

Definition (ε-pseudospectrum)

For any $\varepsilon > 0$, the *ε-pseudospectrum of* $\mathbf{A}$, denoted $\sigma_\varepsilon(\mathbf{A})$, is the set

$$\sigma_\varepsilon(\mathbf{A}) = \{z \in \mathbf{C} : z \in \sigma(\mathbf{A} + \mathbf{E}) \text{ for some } \mathbf{E} \in \mathbf{C}^{n \times n} \text{ with } \|\mathbf{E}\| < \varepsilon\}.$$

At the time, we assumed that $\|\cdot\|$ was the vector 2-norm on $\mathbf{C}^n$ and the "operator norm" it induces,

$$\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|.$$

Now we address various generalizations:

- More general induced/operator norms on $\mathbf{C}^n$;
- Matrix norms on $\mathbf{C}^{n \times n}$ not induced by vector norms;
- Infinite dimensional Banach or Hilbert spaces, with the operator norm.

In applications, the use of the proper norm is an essential (but sometimes overlooked) aspect of pseudospectral theory.

Equivalent definitions in the induced 2-norm

Suppose for the moment that we still use the vector 2-norm, $\|\mathbf{x}\| = \sqrt{\mathbf{x}^*\mathbf{x}}$.

In our discussion (especially about computing pseudospectra) we have occasionally used the fact that

$$\|(z - \mathbf{A})^{-1}\| = \frac{1}{s_{\min}(z - \mathbf{A})}.$$

Theorem

The following four definitions of the ε-pseudospectrum are equivalent:

1. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : z \in \sigma(\mathbf{A} + \mathbf{E})$ *for some* $\mathbf{E} \in C^{n \times n}$ *with* $\|\mathbf{E}\| < \varepsilon\}$;

2. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : \|(z - \mathbf{A})^{-1}\| > 1/\varepsilon\}$;

2a. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : s_{\min}(z - \mathbf{A}) < \varepsilon\}$;

3. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : \|\mathbf{A}\mathbf{v} - z\mathbf{v}\| < \varepsilon$ *for some unit vector* $\mathbf{v} \in C^n\}$.

This equivalence (2)=(2a) is a property of the vector 2-norm that can be extended to any norm induced by an inner product, $\langle \mathbf{x}, \mathbf{y} \rangle := \overline{\mathbf{y}}^{\mathsf{T}} \mathbf{G} \mathbf{x}$ for $\mathbf{G}$ Hermitian positive definite; the adjoint becomes $\mathbf{A}^\star = \mathbf{G}^{-1} \overline{\mathbf{A}}^{\mathsf{T}} \mathbf{G}$.

Pseudospectra with Induced Norms on $\mathbf{C}^n$

Let $\|\cdot\|$ denote any vector norm on $\mathbf{C}^n$ and let

$$\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|.$$

The equivalence of the three definitions of the ε-pseudospectrum still hold, but one proof gets generalized.

Theorem

The following three definitions of the ε-pseudospectrum are equivalent:

1. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : z \in \sigma(\mathbf{A}+\mathbf{E})$ *for some* $\mathbf{E} \in C^{n\times n}$ *with* $\|\mathbf{E}\| < \varepsilon\}$;
2. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : \|(z-\mathbf{A})^{-1}\| > 1/\varepsilon\}$;
3. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : \|\mathbf{A}\mathbf{v} - z\mathbf{v}\| < \varepsilon$ *for some unit vector* $\mathbf{v} \in C^n\}$.

Pseudospectra with Induced Norms on $\mathbb{C}^n$

Let $\|\cdot\|$ denote any vector norm on $\mathbb{C}^n$ and let

$$\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|.$$

The equivalence of the three definitions of the ε-pseudospectrum still hold, but one proof gets generalized.

Theorem

The following three definitions of the ε-pseudospectrum are equivalent:

1. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : z \in \sigma(\mathbf{A}+\mathbf{E})$ *for some* $\mathbf{E} \in C^{n\times n}$ *with* $\|\mathbf{E}\| < \varepsilon\}$;
2. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : \|(z-\mathbf{A})^{-1}\| > 1/\varepsilon\}$;
3. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : \|\mathbf{A}\mathbf{v} - z\mathbf{v}\| < \varepsilon$ *for some unit vector* $\mathbf{v} \in C^n\}$.

Proof. (3) $\Longrightarrow$ (1)

Given a unit vector $\mathbf{v}$ such that $\|\mathbf{A}\mathbf{v} - z\mathbf{v}\| < \varepsilon$, define $\mathbf{r} := \mathbf{A}\mathbf{v} - z\mathbf{v}$.
Now set $\mathbf{E} := -\mathbf{r}\mathbf{w}^*$, where $\mathbf{w}$ is a vector such that $\mathbf{w}^*\mathbf{v} = 1$
(possible by the theory of dual norms, or Hahn–Banach), so that

$$(\mathbf{A}+\mathbf{E})\mathbf{v} = (\mathbf{A} - \mathbf{r}\mathbf{w}^*)\mathbf{v} = \mathbf{A}\mathbf{v} - \mathbf{r} = z\mathbf{v}.$$

Hence $z \in \sigma(\mathbf{A}+\mathbf{E})$. $\square$

Pseudospectra with Matrix Norms on $\mathbb{C}^{n\times n}$

We can define pseudospectra for a matrix norm that *is not induced by a vector norm*, such as the Frobenius norm (a.k.a. Schatten 2-norm):

$$\|\mathbf{A}\|_F = \Big(\sum_{j,k} |a_{j,k}|^2\Big)^{1/2} = \sqrt{s_1^2 + \cdots + s_n^2}.$$

Observation

If $\|\cdot\|$ is a matrix norm not induced by a vector norm, the following definitions need not be equivalent:

1. $\sigma_\varepsilon(\mathbf{A}) = \{z \in \mathbb{C} : z \in \sigma(\mathbf{A}+\mathbf{E})$ *for some* $\mathbf{E} \in \mathbb{C}^{n\times n}$ *with* $\|\mathbf{E}\| < \varepsilon\}$;
2. $\sigma_\varepsilon(\mathbf{A}) = \{z \in \mathbb{C} : \|(z-\mathbf{A})^{-1}\| > 1/\varepsilon\}$.

Pseudospectra with Matrix Norms on $\mathbb{C}^{n\times n}$

$$\|\mathbf{A}\|_F = \Big(\sum_{j,k} |a_{j,k}|^2\Big)^{1/2} = \sqrt{s_1^2 + \cdots + s_n^2}.$$

Observation

If $\|\cdot\|$ is a matrix norm not induced by a vector norm, the following definitions need not be equivalent:

1. $\sigma_\varepsilon(\mathbf{A}) = \{z \in \mathbb{C} : z \in \sigma(\mathbf{A}+\mathbf{E})$ *for some* $\mathbf{E} \in \mathbb{C}^{n\times n}$ *with* $\|\mathbf{E}\| < \varepsilon\}$;
2. $\sigma_\varepsilon(\mathbf{A}) = \{z \in \mathbb{C} : \|(z-\mathbf{A})^{-1}\| > 1/\varepsilon\}$.

Let $(s, \mathbf{u}, \mathbf{v})$ be the smallest singular value and corresponding singular vectors for $z - \mathbf{A}$, so that $(z-\mathbf{A})\mathbf{v} = s\mathbf{u}$. Define $\mathbf{E} = -s\mathbf{u}\mathbf{v}^*$, so that

$$(\mathbf{A}+\mathbf{E})\mathbf{v} = (\mathbf{A}+s\mathbf{u}\mathbf{v}^*)\mathbf{v} = \mathbf{A}\mathbf{v} + s\mathbf{u} = z\mathbf{v},$$

so $z \in \sigma(\mathbf{A}+\mathbf{E})$. Notice that $\|\mathbf{E}\|_F = \|\mathbf{E}\|_2 = s$.

Since $\|\mathbf{E}\|_F \geq \|\mathbf{E}\|_2$ for all $\mathbf{E}$, we conclude that the 2-norm and Frobenius-norm pseudospectra are identical, by definition (1).

Pseudospectra with Matrix Norms on $\mathbb{C}^{n\times n}$

$$\|\mathbf{A}\|_F = \Big(\sum_{j,k} |a_{j,k}|^2\Big)^{1/2} = \sqrt{s_1^2 + \cdots + s_n^2}.$$

Observation

If $\|\cdot\|$ is a matrix norm not induced by a vector norm, the following definitions need not be equivalent:

1. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : z \in \sigma(\mathbf{A}+\mathbf{E})$ *for some* $\mathbf{E} \in C^{n\times n}$ *with* $\|\mathbf{E}\| < \varepsilon\}$;
2. $\sigma_\varepsilon(\mathbf{A}) = \{z \in C : \|(z-\mathbf{A})^{-1}\| > 1/\varepsilon\}$.

On the other hand, if $(z-\mathbf{A})^{-1}$ has singular values $s_1, \ldots, s_n$, then

$$\sqrt{s_1^2 + \cdots + s_n^2} = \|(z-\mathbf{A})^{-1}\|_F \geq \|(z-\mathbf{A})^{-1}\|_2 = s_1.$$

Hence in general, the 2-norm ε-pseudospectrum is *strictly contained* inside the Frobenius norm ε-pseudospectrum, if defined by (2).

Using this definition (2), the Frobenius norm pseudospectra "determine matrix behavior," as described in Lecture 2.

Computation of Pseudospectra in General Norms

When describing computation of $\sigma_\varepsilon(\mathbf{A})$, our algorithms (and EigTool) used the 2-norm. Different norms require further consideration.

- Norm induced by $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{y}^*\mathbf{G}\mathbf{x}$.
  Factor $\mathbf{G} = \mathbf{R}^*\mathbf{R}$ (e.g., Cholesky decomposition, matrix square root).
  Then for any matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$,

$$\|\mathbf{A}\|^2 = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|^2}{\|\mathbf{x}\|^2} = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^*\mathbf{A}^*\mathbf{G}\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{G}\mathbf{x}}$$

$$= \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^*\mathbf{R}^*\mathbf{R}^{-*}\mathbf{A}^*\mathbf{R}^*\mathbf{R}\mathbf{A}\mathbf{R}^{-1}\mathbf{R}\mathbf{x}}{\mathbf{x}^*\mathbf{R}^*\mathbf{R}\mathbf{x}}$$

$$= \sup_{\substack{\mathbf{y}=\mathbf{R}\mathbf{x}\\ \mathbf{x}\neq\mathbf{0}}} \frac{\mathbf{y}(\mathbf{R}^{-*}\mathbf{A}^*\mathbf{R}^*)(\mathbf{R}\mathbf{A}\mathbf{R}^{-1})\mathbf{y}}{\mathbf{y}^*\mathbf{y}} = \|\mathbf{R}\mathbf{A}\mathbf{R}^{-1}\|_2^2.$$

  Hence, the $\mathbf{G}$-norm pseudospectra of $\mathbf{A}$ are the same as the 2-norm pseudospectra of $\mathbf{R}\mathbf{A}\mathbf{R}^{-1}$. This permits the use of EigTool without modification: just call `eigtool(R*A/R)`.

- To compute pseudospectra in other norms, one typically needs to explicitly compute $(z-\mathbf{A})^{-1}$ at each point on a computational grid.

Pseudospectra in Infinite Dimensional Spaces

Let X denote a Banach space.

Let $\mathbf{A}$ be a closed (possibly unbounded) linear operator, $\mathbf{A} : \mathrm{Dom}(\mathbf{A}) \to X$.

Definition (Spectrum, Resolvent Set)

The *resolvent set* of $\mathbf{A}$ is

$$\varrho(\mathbf{A}) = \{z \in \mathbf{C} : (z - \mathbf{A}) \text{ has a bounded, densely defined inverse}\}.$$

The *spectrum* is the complement of the resolvent set,

$$\sigma(\mathbf{A}) = \mathbf{C} \setminus \varrho(\mathbf{A}) = \{z \in \mathbf{C} : z \notin \varrho(\mathbf{A})\}.$$

Pseudospectra in Infinite Dimensional Spaces

Let X denote a Banach space.

Let $\mathbf{A}$ be a closed (possibly unbounded) linear operator, $\mathbf{A} : \mathrm{Dom}(\mathbf{A}) \to X$.

Definition (Spectrum, Resolvent Set)

The *resolvent set* of $\mathbf{A}$ is

$$\varrho(\mathbf{A}) = \{z \in \mathbf{C} : (z - \mathbf{A}) \text{ has a bounded, densely defined inverse}\}.$$

The *spectrum* is the complement of the resolvent set,

$$\sigma(\mathbf{A}) = \mathbf{C} \setminus \varrho(\mathbf{A}) = \{z \in \mathbf{C} : z \notin \varrho(\mathbf{A})\}.$$

- The spectrum can contain points that are not eigenvalues, i.e., for which there exist no nonzero $\mathbf{v} \in \mathrm{Dom}(\mathbf{A})$ for which $\mathbf{A}\mathbf{v} = z\mathbf{v}$.
- For unbounded operators, it is possible that $\sigma(\mathbf{A}) = \emptyset$.
- By convention, we shall write $\mathbf{A}^{-1}$ when $\mathbf{A}$ has a *bounded* inverse with $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ on X and $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ on $\mathrm{Dom}(\mathbf{A})$.
  If such an inverse does not exist, we write $\|\mathbf{A}^{-1}\| = \infty$.

Pseudospectra

Let $\mathcal{B}(X)$ denote the set of bounded operators defined on all X.

The three main equivalent definitions of $\sigma_\varepsilon(\mathbf{A})$ hold with only slight modifications:

Theorem

The following three definitions of the ε-pseudospectrum are equivalent.

1. $\sigma_\varepsilon(\mathbf{A}) = \{z \in \boldsymbol{C} : z \in \sigma(\mathbf{A} + \mathbf{E})$ *for some* $\mathbf{E} \in \mathcal{B}(X)$ *with* $\|\mathbf{E}\| < \varepsilon\}$;
2. $\sigma_\varepsilon(\mathbf{A}) = \{z \in \boldsymbol{C} : \|(z - \mathbf{A})^{-1}\| > 1/\varepsilon\}$;
3. $\sigma_\varepsilon(\mathbf{A}) = \{z \in \boldsymbol{C} : z \in \sigma(\mathbf{A})$ *or* $\|\mathbf{A}\mathbf{v} - z\mathbf{v}\| < \varepsilon$ *for some unit vector* $\mathbf{v} \in \boldsymbol{C}^n\}$.

Stability of Bounded Invertibility

Basic properties follow from what Kato calls "stability of bounded invertibility".

Theorem

Suppose $\mathbf{A}$ is a closed operator on X with $\mathbf{A}^{-1} \in \mathcal{B}(X)$. Then for $\mathbf{E} \in \mathcal{B}(X)$ with $\|\mathbf{E}\| < 1/\|\mathbf{A}^{-1}\|$, $\mathbf{A}+\mathbf{E}$ has a bounded inverse that satisfies

$$\|(\mathbf{A}+\mathbf{E})^{-1}\| \leq \frac{\|\mathbf{A}^{-1}\|}{1-\|\mathbf{E}\|\|\mathbf{A}^{-1}\|}.$$

Proof: Since $1 > \|\mathbf{E}\|\|\mathbf{A}^{-1}\| \geq \|\mathbf{E}\mathbf{A}^{-1}\|$, $\mathbf{I}+\mathbf{E}\mathbf{A}^{-1}$ is invertible, with inverse given by the Neumann expansion

$$(\mathbf{I}+\mathbf{E}\mathbf{A}^{-1})^{-1} = \mathbf{I} - \mathbf{E}\mathbf{A}^{-1} + (\mathbf{E}\mathbf{A}^{-1})^2 - \cdots,$$

and hence

$$\|(\mathbf{I}+\mathbf{E}\mathbf{A}^{-1})^{-1}\| \leq \frac{1}{1-\|\mathbf{E}\mathbf{A}^{-1}\|} \leq \frac{1}{1-\|\mathbf{E}\|\|\mathbf{A}^{-1}\|}.$$

Since $\mathbf{A}$ also has a bounded inverse, so too does $\mathbf{A}+\mathbf{E} = (\mathbf{I}+\mathbf{E}\mathbf{A}^{-1})\mathbf{A}$, with

$$\|(\mathbf{A}+\mathbf{E})^{-1}\| = \|(\mathbf{I}+\mathbf{E}\mathbf{A}^{-1})^{-1}\mathbf{A}^{-1}\| \leq \frac{\|\mathbf{A}^{-1}\|}{1-\|\mathbf{E}\|\|\mathbf{A}^{-1}\|}. \qquad \square$$

Basic Properties of Pseudospectra on Banach Spaces

- $\|(z - \mathbf{A})^{-1}\| \geq 1/\mathrm{dist}(z, \sigma(\mathbf{A}))$.
- $(z - \mathbf{A})^{-1}$ is analytic on $\varrho(\mathbf{A})$.
- $\|(z - \mathbf{A})^{-1}\|$ is an unbounded, subharmonic function on $\varrho(\mathbf{A})$.
- For all $\varepsilon > 0$, $\sigma_\varepsilon(\mathbf{A})$ is a nonempty open subset of $\mathbf{C}$.
- Each bounded component of $\sigma_\varepsilon(\mathbf{A})$ intersects $\sigma(\mathbf{A})$.

Weak versus Strong Inequalities

For matrices (or operators on Hilbert space [Fink & Ehrhardt]),
$\sigma_\varepsilon(\mathbf{A})$ can be defined with weak or strong inequalities:
in both cases, definitions (1) and (2) are equivalent:

$$\{z \in \mathbf{C} : z \in \sigma(\mathbf{A}+\mathbf{E}) \text{ for some } \mathbf{E} \in \mathcal{B}(X) \text{ with } \|\mathbf{E}\| < \varepsilon\}$$
$$= \{z \in \mathbf{C} : \|(z-\mathbf{A})^{-1}\| > 1/\varepsilon\}.$$

and

$$\{z \in \mathbf{C} : z \in \sigma(\mathbf{A}+\mathbf{E}) \text{ for some } \mathbf{E} \in \mathcal{B}(X) \text{ with } \|\mathbf{E}\| \le \varepsilon\}$$
$$= \{z \in \mathbf{C} : \|(z-\mathbf{A})^{-1}\| \ge 1/\varepsilon\}.$$

Weak versus Strong Inequalities

For matrices (or operators on Hilbert space [Fink & Ehrhardt]), $\sigma_\varepsilon(\mathbf{A})$ can be defined with weak or strong inequalities: in both cases, definitions (1) and (2) are equivalent:

$$\{z \in \mathbf{C} : z \in \sigma(\mathbf{A}+\mathbf{E}) \text{ for some } \mathbf{E} \in \mathcal{B}(X) \text{ with } \|\mathbf{E}\| < \varepsilon\}$$
$$= \{z \in \mathbf{C} : \|(z-\mathbf{A})^{-1}\| > 1/\varepsilon\}.$$

and

$$\{z \in \mathbf{C} : z \in \sigma(\mathbf{A}+\mathbf{E}) \text{ for some } \mathbf{E} \in \mathcal{B}(X) \text{ with } \|\mathbf{E}\| \le \varepsilon\}$$
$$= \{z \in \mathbf{C} : \|(z-\mathbf{A})^{-1}\| \ge 1/\varepsilon\}.$$

Observation

Shargorodsky [2008,2009] has shown that the equivalence for weak inequalities can fail in Banach spaces:

$$\{z \in C : z \in \sigma(\mathbf{A}+\mathbf{E}) \textit{ for some } \mathbf{E} \in \mathcal{B}(X) \textit{ with } \|\mathbf{E}\| \le \varepsilon\}$$
$$\ne \{z \in C : \|(z-\mathbf{A})^{-1}\| \ge 1/\varepsilon\}.$$

4(e) Pseudospectra of Differential Operators

Pseudospectra of Differential Operators

As an illustrative example, we shall consider the pseudospectra of constant-coefficient differential operators in one dimension.
There is a close parallel to the pseudospectra of Toeplitz matrices in Lecture 3.

We pose these problems on the Hilbert space $X = L^2(0, \ell)$ for $\ell > 0$.
In general, consider operators of the form

$$\mathbf{A} = \sum_{j=0}^{d} a_j \frac{d^j}{dx^j}.$$

For example, we shall consider:

$$\begin{aligned}
&\mathbf{A}u = u', && a_0 = 0,\ a_1 = 1\\
&\mathbf{A}u = cu' + u'', && a_0 = 0,\ a_1 = c,\ a_2 = 1\\
&\mathbf{A}u = u + 3u' + 3u'' + u''', && a_0 = 1,\ a_1 = 3,\ a_2 = 3,\ a_3 = 1
\end{aligned}$$

Pseudospectra of Differential Operators

As an illustrative example, we shall consider the pseudospectra of constant-coefficient differential operators in one dimension.
There is a close parallel to the pseudospectra of Toeplitz matrices in Lecture 3.

We pose these problems on the Hilbert space $X = L^2(0, \ell)$ for $\ell > 0$.
In general, consider operators of the form

$$\mathbf{A} = \sum_{j=0}^{d} a_j \frac{d^j}{dx^j}.$$

For example, we shall consider:

$$\begin{aligned}
&\mathbf{A}u = u', && a_0 = 0,\ a_1 = 1\\
&\mathbf{A}u = cu' + u'', && a_0 = 0,\ a_1 = c,\ a_2 = 1\\
&\mathbf{A}u = u + 3u' + 3u'' + u''', && a_0 = 1,\ a_1 = 3,\ a_2 = 3,\ a_3 = 1
\end{aligned}$$

Definition (Symbol)

The *symbol* of the differential operator $\mathbf{A}$ is

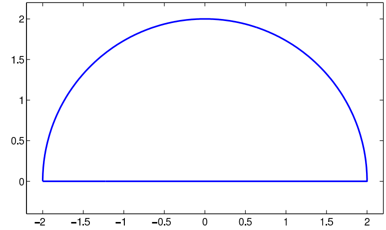
$$a(k) = \sum_{j=0}^{d} a_j(-i\,k)^j.$$

Symbol Curve

$$\mathbf{A} = \sum_{j=0}^{d} a_j \frac{d^j}{dx^j}, \qquad a(k) = \sum_{j=0}^{d} a_j(-i\,k)^j.$$

The *symbol curve* is more complicated than for the $a(\mathbb{T})$ for Toeplitz matrices.

For $R > 0$, define Γ_R to be the real interval $[-R, R]$ with endpoints joined by the semicircle of radius R centered at zero in the upper half-plane.



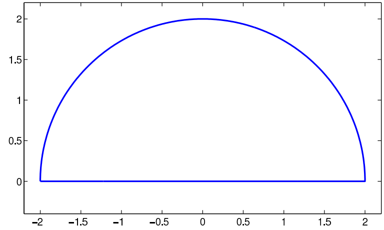
Γ_R for $R = 2$

Symbol Curve

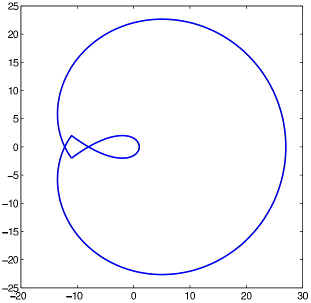
$$\mathbf{A} = \sum_{j=0}^{d} a_j \frac{d^j}{dx^j}, \qquad a(k) = \sum_{j=0}^{d} a_j(-i\,k)^j.$$

The *symbol curve* is more complicated than for the $a(\mathbb{T})$ for Toeplitz matrices.

For $R > 0$, define Γ_R to be the real interval $[-R, R]$ with endpoints joined by the semicircle of radius R centered at zero in the upper half-plane.
Then consider the image of Γ_R under a as $R \to \infty$.



Γ_R for $R = 2$



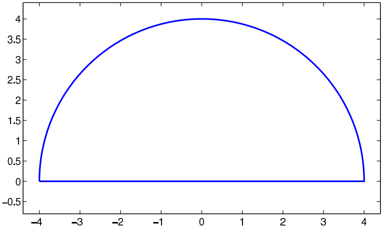
$a(\Gamma_R)$ for $R = 2$

Symbol Curve

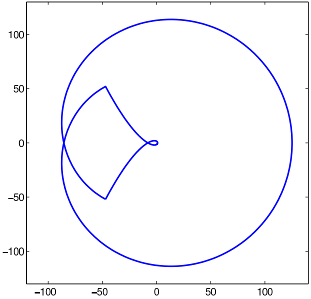
$$\mathbf{A} = \sum_{j=0}^{d} a_j \frac{d^j}{dx^j}, \qquad a(k) = \sum_{j=0}^{d} a_j(-i\,k)^j.$$

The *symbol curve* is more complicated than for the $a(\mathbb{T})$ for Toeplitz matrices.

For $R > 0$, define Γ_R to be the real interval $[-R, R]$ with endpoints joined by the semicircle of radius R centered at zero in the upper half-plane.
Then consider the image of Γ_R under a as $R \to \infty$.



Γ_R for $R = 4$



$a(\Gamma_R)$ for $R = 4$

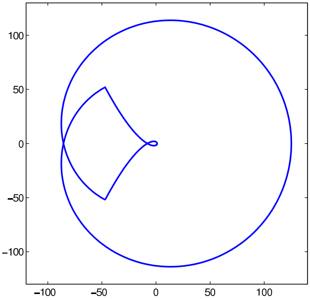
Symbol Curve: Winding Number

$$\mathbf{A} = \sum_{j=0}^{d} a_j \frac{d^j}{dx^j}, \qquad a(k) = \sum_{j=0}^{d} a_j(-ik)^j.$$

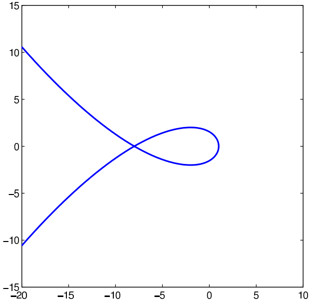
For $R > 0$, define Γ_R to be the real interval $[-R, R]$ with endpoints joined by the semicircle of radius R centered at zero in the upper half-plane.

Definition

For any $\lambda \notin a(\mathbf{R})$, let $I(a, \lambda)$ be the winding number of $a(\Gamma_R)$ about λ for sufficiently large R.



$a(\Gamma_R)$ for $R = 4$



$a(\Gamma_R)$ for $R = 4$ (zoom)

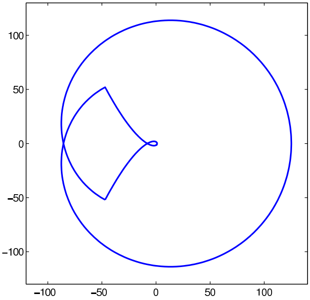
Symbol Curve: Winding Number

$$\mathbf{A} = \sum_{j=0}^{d} a_j \frac{d^j}{dx^j}, \qquad a(k) = \sum_{j=0}^{d} a_j(-ik)^j.$$

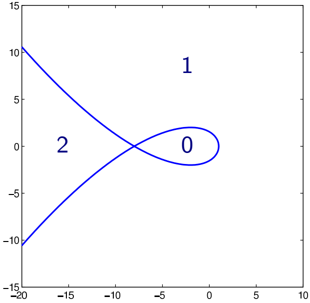
For $R > 0$, define Γ_R to be the real interval $[-R, R]$ with endpoints joined by the semicircle of radius R centered at zero in the upper half-plane.

Definition

For any $\lambda \notin a(\mathbb{R})$, let $I(a, \lambda)$ be the winding number of $a(\Gamma_R)$ about λ for sufficiently large R.



$a(\Gamma_R)$ for $R = 4$



$a(\Gamma_R)$ for $R = 4$ (zoom)

Pseudospectra of Differential Operators

Theorem (Reddy)

Consider the set of degree-d constant coefficient differential operators $\{\mathbf{A}_\ell\}$ *posed on* $L^2[0,\ell]$ *with symbol a and*

- β *homogeneous boundary conditions at* $x = 0$*;*
- γ *homogeneous boundary conditions at* $x = \ell$*.*

Let $z \in \boldsymbol{C}$*. If the winding number satisfies*

- $I(a, z) < d - \beta$*, or*
- $I(a, z) > \gamma$*,*

then there exists $M > 0$ *such that for all sufficiently large* ℓ*,*

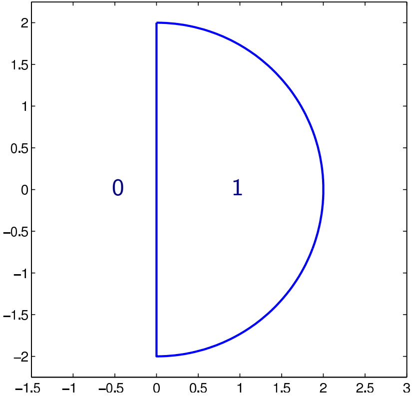
$$\|(z - \mathbf{A}_\ell)^{-1}\| \geq \mathrm{e}^{\ell M}.$$

Pseudospectra of Differential Operators: example 1

Consider the operator

$$\mathbf{A}u = u', \qquad u(\ell) = 0.$$

Since $a_0 = 0$ and $a_1 = 1$, the symbol is $a(k) = -ik$.



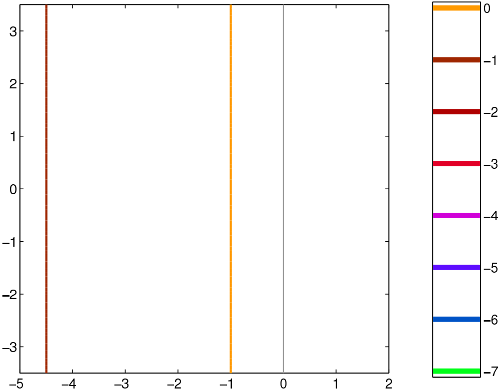
$a(\Gamma_R)$ for $R = 2$

Pseudospectra of Differential Operators: example 1

Consider the operator

$$\mathbf{A}u = u', \qquad u(\ell) = 0.$$

Since $a_0 = 0$ and $a_1 = 1$, the symbol is $a(k) = -ik$.



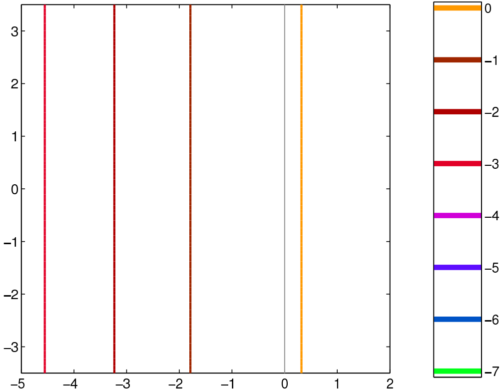
$\sigma_\varepsilon(\mathbf{A})$ for $\ell = 1$

Pseudospectra of Differential Operators: example 1

Consider the operator

$$\mathbf{A}u = u', \qquad u(\ell) = 0.$$

Since $a_0 = 0$ and $a_1 = 1$, the symbol is $a(k) = -ik$.



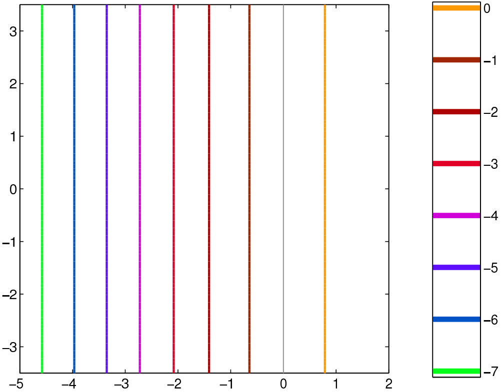
$\sigma_\varepsilon(\mathbf{A})$ for $\ell = 2$

Pseudospectra of Differential Operators: example 1

Consider the operator

$$\mathbf{A}u = u', \qquad u(\ell) = 0.$$

Since $a_0 = 0$ and $a_1 = 1$, the symbol is $a(k) = -ik$.



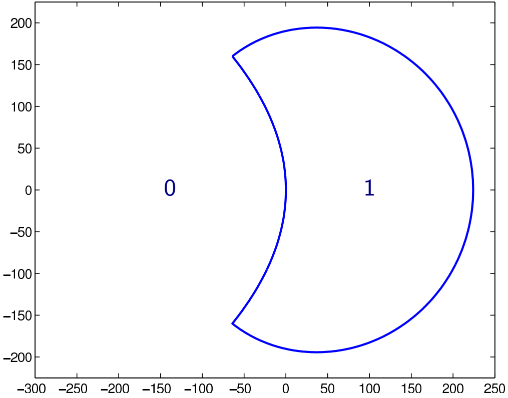
$\sigma_\varepsilon(\mathbf{A})$ for $\ell = 3$

Pseudospectra of Differential Operators: example 2

Consider the operator

$$\mathbf{A}u = cu' + u'', \qquad u(0) = u(\ell) = 0.$$

Since $a_0 = 0$, $a_1 = c = 20$, and $a_2 = 1$ the symbol is $a(k) = -cik - k^2$.



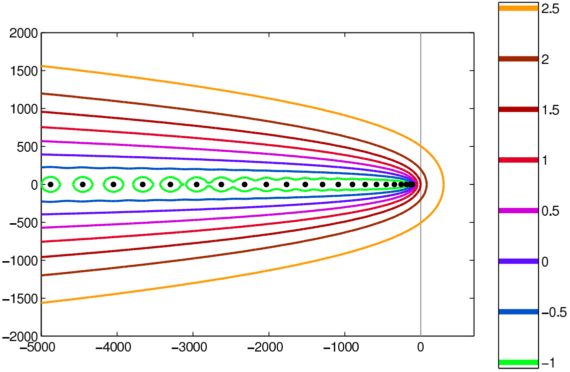
$a(\Gamma_R)$ for $R = 8$

Pseudospectra of Differential Operators: example 2

Consider the operator

$$\mathbf{A}u = cu' + u'', \qquad u(0) = u(\ell) = 0.$$

Since $a_0 = 0$, $a_1 = c = 20$, and $a_2 = 1$ the symbol is $a(k) = -cik - k^2$.



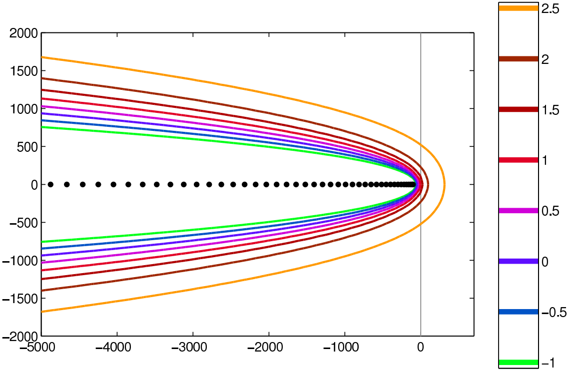
$\sigma_\varepsilon(\mathbf{A})$ for $\ell = 1$

Pseudospectra of Differential Operators: example 2

Consider the operator

$$\mathbf{A}u = cu' + u'', \qquad u(0) = u(\ell) = 0.$$

Since $a_0 = 0$, $a_1 = c = 20$, and $a_2 = 1$ the symbol is $a(k) = -cik - k^2$.



$\sigma_\varepsilon(\mathbf{A})$ for $\ell = 2$