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 3: Functions of Matrices and Model Reduction

- Toeplitz matrices
- Model Reduction: Balanced Truncation
- Nonnormality and Lyapunov Equations
- Model Reduction: Moment Matching

3(a) Toeplitz Matrices

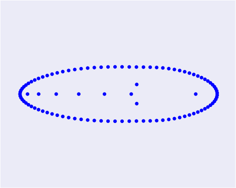
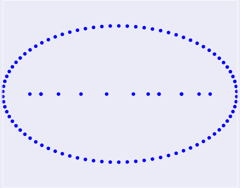
Pseudospectra

Recall the example that began our investigation of pseudospectra yesterday.

Example

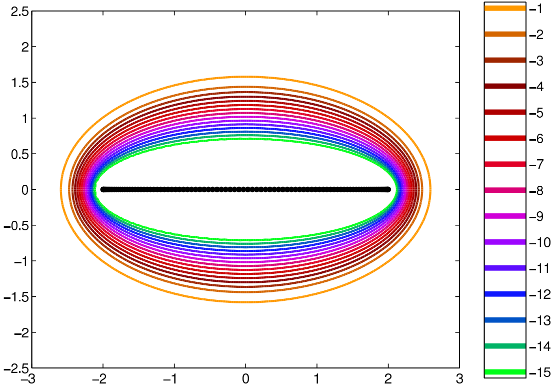
Compute eigenvalues of three similar 100×100 *matrices using MATLAB's* `eig`.

$$\begin{bmatrix} 0 & 1 & & \\ 1 & 0 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1/2 & & \\ 2 & 0 & \ddots & \\ & \ddots & \ddots & 1/2 \\ & & 2 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1/3 & & \\ 3 & 0 & \ddots & \\ & \ddots & \ddots & 1/3 \\ & & 3 & 0 \end{bmatrix}$$

Toeplitz Matrices

Consider the pseudospectra of the 100×100 matrix in the middle of the last slide, $\mathbf{A} = \text{tridiag}(2, 0, 1/2)$.

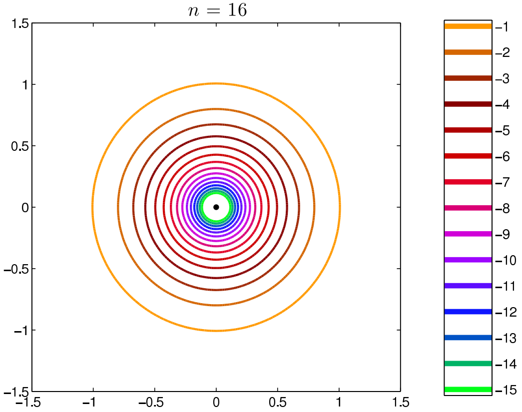


$\mathbf{A}$ is diagonalizable (it has distinct eigenvalues), but Bauer–Fike is useless here: $\kappa(\mathbf{V}) = 2^{99} \approx 6 \times 10^{29}$.

Jordan Blocks

We've already analyzed pseudospectra of Jordan blocks near λ for small $\varepsilon > 0$. Here we want to investigate the entire pseudospectrum for larger ε.

$$\mathbf{S}_n = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$$

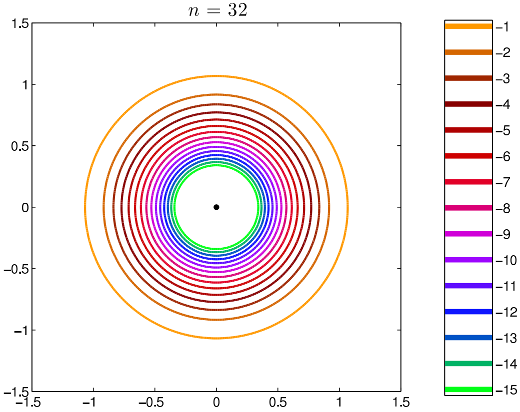


Near the eigenvalue, the resolvent norm grows with dimension n; outside the unit disk, the resolvent norm does not seem to get big. We would like to prove this.

Jordan Blocks

We've already analyzed pseudospectra of Jordan blocks near λ for small $\varepsilon > 0$. Here we want to investigate the entire pseudospectrum for larger ε.

$$\mathbf{S}_n = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$$

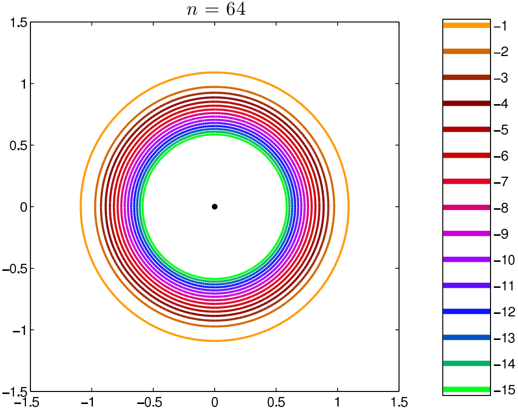


Near the eigenvalue, the resolvent norm grows with dimension n; outside the unit disk, the resolvent norm does not seem to get big. We would like to prove this.

Jordan Blocks

We've already analyzed pseudospectra of Jordan blocks near λ for small $\varepsilon > 0$. Here we want to investigate the entire pseudospectrum for larger ε.

$$\mathbf{S}_n = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$$



Near the eigenvalue, the resolvent norm grows with dimension *n*; outside the unit disk, the resolvent norm does not seem to get big. We would like to prove this.

Jordan Blocks/Shift Operator

Consider the generalization of the Jordan block to the domain

$$\ell^2(\mathbf{N}) = \{(x_1, x_2, \ldots) : \sum_{j=1}^{\infty} |x_j|^2 < \infty\}.$$

The shift operator $\mathbf{S}$ on $\ell^2(\mathbf{N})$ is defined as

$$\mathbf{S}(x_1, x_2, \ldots) = (x_2, x_3, \ldots).$$

Jordan Blocks/Shift Operator

Consider the generalization of the Jordan block to the domain

$$\ell^2(\mathbf{N}) = \{(x_1, x_2, \ldots) : \sum_{j=1}^{\infty} |x_j|^2 < \infty\}.$$

The shift operator $\mathbf{S}$ on $\ell^2(\mathbf{N})$ is defined as

$$\mathbf{S}(x_1, x_2, \ldots) = (x_2, x_3, \ldots).$$

In particular,

$$\begin{aligned} \mathbf{S}(1, z, z^2, \ldots) &= (z, z^2, z^3, \ldots) \\ &= z(1, z, z^2, \ldots). \end{aligned}$$

So if $(1, z, z^2, \ldots) \in \ell^2(\mathbf{N})$, then $z \in \sigma(\mathbf{S})$.

Jordan Blocks/Shift Operator

Consider the generalization of the Jordan block to the domain

$$\ell^2(\mathbf{N}) = \{(x_1, x_2, \ldots) : \sum_{j=1}^{\infty} |x_j|^2 < \infty\}.$$

The shift operator $\mathbf{S}$ on $\ell^2(\mathbf{N})$ is defined as

$$\mathbf{S}(x_1, x_2, \ldots) = (x_2, x_3, \ldots).$$

In particular,

$$\begin{aligned} \mathbf{S}(1, z, z^2, \ldots) &= (z, z^2, z^3, \ldots) \\ &= z(1, z, z^2, \ldots). \end{aligned}$$

So if $(1, z, z^2, \ldots) \in \ell^2(\mathbf{N})$, then $z \in \sigma(\mathbf{S})$.

If $|z| < 1$, then

$$\sum_{j=1}^{\infty} |z^{j-1}|^2 = \frac{1}{1 - |z|^2} < \infty.$$

So,

$$\{z \in \mathbf{C} : |z| < 1\} \subseteq \sigma(\mathbf{S}).$$

Jordan Blocks/Shift Operator

$$\mathbf{S}(x_1, x_2, \dots) = (x_2, x_3, \dots).$$

We have seen that

$$\{z \in \mathbf{C} : |z| < 1\} \subseteq \sigma(\mathbf{S}).$$

Observe that

$$\|\mathbf{S}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{Sx}\| = 1,$$

and so

$$\sigma(\mathbf{S}) \subseteq \{z \in \mathbf{C} : |z| \leq 1\}.$$

The spectrum is closed, so

$$\sigma(\mathbf{A}) = \{z \in \mathbf{C} : |z| \leq 1\}.$$

For any finite dimensional $n \times n$ Jordan block $\mathbf{S}_n$,

$$\sigma(\mathbf{S}_n) = \{0\}.$$

Jordan Blocks/Shift Operator

$$\mathbf{S}(x_1, x_2, \ldots) = (x_2, x_3, \ldots).$$

We have seen that

$$\{z \in \mathbf{C} : |z| < 1\} \subseteq \sigma(\mathbf{S}).$$

Observe that

$$\|\mathbf{S}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{S}\mathbf{x}\| = 1,$$

and so

$$\sigma(\mathbf{S}) \subseteq \{z \in \mathbf{C} : |z| \le 1\}.$$

The spectrum is closed, so

$$\sigma(\mathbf{A}) = \{z \in \mathbf{C} : |z| \le 1\}.$$

For any finite dimensional $n \times n$ Jordan block $\mathbf{S}_n$,

$$\sigma(\mathbf{S}_n) = \{0\}.$$

So the $\mathbf{S}_n \to \mathbf{S}$ strongly, but there is a discontinuity in the spectrum:

$$\sigma(\mathbf{S}_n) \not\to \sigma(\mathbf{S}).$$

Pseudospectra of Jordan Blocks

Pseudospectra resolve this unpleasant discontinuity.

Recall the eigenvectors $(1, z, z^2, \ldots)$ for $\mathbf{S}$.

Truncate this vector to length n, and apply it to $\mathbf{S}_n$:

$$\begin{bmatrix} 0 & 1 & & & \\ & 0 & \ddots & & \\ & & \ddots & 1 & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} = \begin{bmatrix} z \\ z^2 \\ \vdots \\ z^{n-1} \\ 0 \end{bmatrix}$$

Pseudospectra of Jordan Blocks

Pseudospectra resolve this unpleasant discontinuity.

Recall the eigenvectors $(1, z, z^2, \ldots)$ for $\mathbf{S}$.

Truncate this vector to length n, and apply it to $\mathbf{S}_n$:

$$\begin{bmatrix} 0 & 1 & & & \\ & 0 & \ddots & & \\ & & \ddots & 1 & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} = \begin{bmatrix} z \\ z^2 \\ \vdots \\ z^{n-1} \\ 0 \end{bmatrix} = z \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ z^n \end{bmatrix}.$$

Pseudospectra of Jordan Blocks

Pseudospectra resolve this unpleasant discontinuity.

Recall the eigenvectors $(1, z, z^2, \ldots)$ for $\mathbf{S}$.

Truncate this vector to length n, and apply it to $\mathbf{S}_n$:

$$\begin{bmatrix} 0 & 1 & & & \\ & 0 & \ddots & & \\ & & \ddots & 1 & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} = \begin{bmatrix} z \\ z^2 \\ \vdots \\ z^{n-1} \\ 0 \end{bmatrix} = z \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ z^n \end{bmatrix}.$$

Hence, $\|\mathbf{S}_n\mathbf{x} - z\mathbf{x}\| = |z|^n$, so for all $\varepsilon > \dfrac{|z|^n}{\|\mathbf{x}\|} = |z|^n \dfrac{\sqrt{1-|z|^{2n}}}{\sqrt{1-|z|^2}}$,

$$z \in \sigma_\varepsilon(\mathbf{S}_n).$$

We conclude that for fixed $|z| < 1$, the resolvent norm $\|(z - \mathbf{S}_n)^{-1}\|$ grows *exponentially* with n.

Upper Triangular Toeplitz Matrices

Consider an upper triangular Toeplitz matrix giving the matrix with constant diagonals containing the Laurent coefficients:

$$\mathbf{A}_n = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & \\ & a_0 & \ddots & \ddots & \vdots \\ & & \ddots & a_1 & a_2 \\ & & & a_0 & a_1 \\ & & & & a_0 \end{bmatrix} \in \mathbf{C}^{n\times n}.$$

Definition (Symbol, Symbol Curve)

Toeplitz matrices are described by their *symbol* a with Taylor expansion

$$a(z) = \sum_{k=0}^{\infty} a_k z^k.$$

Call the image of the unit circle $\mathbf{T}$ under a the *symbol curve*, $a(\mathbf{T})$.

Pseudospectra of Upper Triangular Toeplitz Matrices

Apply the same approximate eigenvector we used for the Jordan block:

$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ & a_0 & \ddots & \ddots & \vdots \\ & & \ddots & a_1 & a_2 \\ & & & a_0 & a_1 \\ & & & & a_0 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{n-1} a_k z^k \\ \sum_{k=0}^{n-1} a_k z^{k+1} \\ \vdots \\ a_0 z^{n-2} + a_1 z^{n-1} \\ a_0 z^{n-1} \end{bmatrix}.$$

If the matrix has fixed bandwidth $b \ll n$, (i.e., $a_k = 0$ for $k > b$), then

$$\begin{bmatrix} a_0 & \cdots & a_b & & \\ & a_0 & \ddots & \ddots & \\ & & \ddots & \ddots & a_b \\ & & & a_0 & \vdots \\ & & & & a_0 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{b} a_k z^k \\ \sum_{k=0}^{b} a_k z^{k+1} \\ \vdots \\ \sum_{k=0}^{b-1} a_k z^{k+n-b} \\ \vdots \\ a_0 z^{n-1} \end{bmatrix}.$$

Pseudospectra of Upper Triangular Toeplitz Matrices

$$\begin{bmatrix} a_0 & \cdots & a_b & & \\ & a_0 & \ddots & \ddots & \\ & & \ddots & \ddots & a_b \\ & & & a_0 & \vdots \\ & & & & a_0 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{b} a_k z^k \\ \sum_{k=0}^{b} a_k z^{k+1} \\ \vdots \\ \sum_{k=0}^{b-1} a_k z^{k+n-b} \\ \vdots \\ a_0 z^{n-1} \end{bmatrix}$$

$$= \sum_{k=0}^{b} a_k z^k \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ a_b z^n \\ \vdots \\ \sum_{k=1}^{b} a_k z^{n-1+k} \end{bmatrix}.$$

Pseudospectra of Upper Triangular Toeplitz Matrices

$$\begin{bmatrix} a_0 & \cdots & a_b & & \\ & a_0 & \ddots & \ddots & \\ & & \ddots & \ddots & a_b \\ & & & a_0 & \vdots \\ & & & & a_0 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{b} a_k z^k \\ \sum_{k=0}^{b} a_k z^{k+1} \\ \vdots \\ \sum_{k=0}^{b-1} a_k z^{k+n-b} \\ \vdots \\ a_0 z^{n-1} \end{bmatrix}$$

$$= \sum_{k=0}^{b} a_k z^k \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ a_b z^n \\ \vdots \\ \sum_{k=1}^{b} a_k z^{n-1+k} \end{bmatrix}.$$

Hence $\mathbf{A}_n\mathbf{x} - a(z)\mathbf{x}$ gets very small in size as $n \to \infty$.
In fact, this reveals the spectrum of the infinite Toeplitz operator on $\ell^2(\mathbf{N})\ldots.$

Spectrum of Toeplitz Operators on $\ell^2(\mathbb{Z})$

For the Toeplitz operator (semi-infinite matrix) $\mathbf{A}_\infty$ with the same symbol:

$$\begin{bmatrix} a_0 & \cdots & a_b & & & \\ & a_0 & \ddots & \ddots & & \\ & & \ddots & \ddots & a_b & \\ & & & a_0 & \ddots & \ddots \\ & & & & \ddots & \end{bmatrix} \begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{b} a_k z^k \\ \sum_{k=0}^{b} a_k z^{k+1} \\ \sum_{k=0}^{b} a_k z^{k+2} \\ \vdots \\ \vdots \end{bmatrix} = a(z) \begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \\ \vdots \end{bmatrix}$$

Spectrum of Toeplitz Operators on $\ell^2(\mathbb{Z})$

For the Toeplitz operator (semi-infinite matrix) $\mathbf{A}_\infty$ with the same symbol:

$$\begin{bmatrix} a_0 & \cdots & a_b & & & \\ & a_0 & \ddots & \ddots & & \\ & & \ddots & \ddots & a_b & \\ & & & a_0 & \ddots & \ddots \\ & & & & \ddots & \end{bmatrix} \begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{b} a_k z^k \\ \sum_{k=0}^{b} a_k z^{k+1} \\ \sum_{k=0}^{b} a_k z^{k+2} \\ \vdots \\ \vdots \end{bmatrix} = a(z) \begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \\ \vdots \end{bmatrix}$$

Thus $a(z) \in \sigma(\mathbf{A})$ for all $|z| < 1$. In fact, one can show that for this banded upper triangular symbol,

$$\sigma(\mathbf{A}_\infty) = \{a(z) : |z| \leq 1\}.$$

Spectrum of Toeplitz Operators on $\ell^2(\mathbb{Z})$

For the Toeplitz operator (semi-infinite matrix) $\mathbf{A}_\infty$ with the same symbol:

$$\begin{bmatrix} a_0 & \cdots & a_b & & & \\ & a_0 & \ddots & \ddots & & \\ & & \ddots & \ddots & a_b & \\ & & & a_0 & \ddots & \ddots \\ & & & & \ddots & \end{bmatrix} \begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{b} a_k z^k \\ \sum_{k=0}^{b} a_k z^{k+1} \\ \sum_{k=0}^{b} a_k z^{k+2} \\ \vdots \\ \vdots \end{bmatrix} = a(z) \begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \\ \vdots \end{bmatrix}$$

Thus $a(z) \in \sigma(\mathbf{A})$ for all $|z| < 1$. In fact, one can show that for this banded upper triangular symbol,

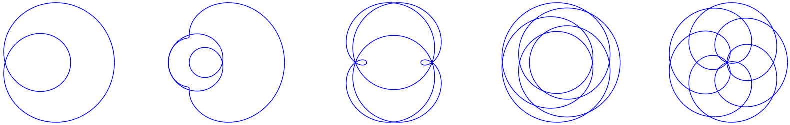
$$\sigma(\mathbf{A}_\infty) = \{a(z) : |z| \leq 1\}.$$

The calculation on the last slide guarantees that for any $\varepsilon > 0$, there exists $N > 0$ such that if $n > N$,

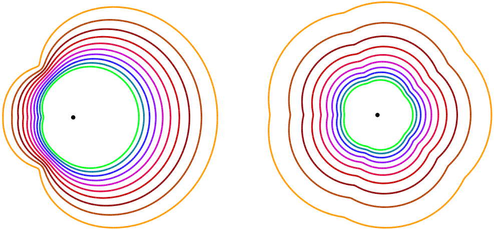
$$\sigma(\mathbf{A}_\infty) \subseteq \sigma_\varepsilon(\mathbf{A}_n).$$

Symbols of Upper Triangular Toeplitz Matrices

Symbol curves for banded upper triangular Toeplitz matrices.



- $a(z) = z + 2z^2$
- $a(z) = z + 2z^2 + z^3 + z^5$
- $a(z) = z + z^3 - z^5$
- $a(z) = z^2 + 3z^5$
- $a(z) = z^2 + z^7$



General Toeplitz Matrices

More generally, the dense Toeplitz matrix

$$\mathbf{A}_n = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & \\ a_{-1} & a_0 & \ddots & \ddots & \vdots \\ a_{-2} & \ddots & \ddots & a_1 & a_2 \\ \vdots & \ddots & a_{-1} & a_0 & a_1 \\ & \cdots & a_{-2} & a_{-1} & a_0 \end{bmatrix} \in \mathbf{C}^{n \times n}$$

follows the same terminology.

Definition (Symbol, Symbol Curve)

Toeplitz matrices are described by their *symbol* a with Laurent expansion

$$a(z) = \sum_{k=-\infty}^{\infty} a_k z^k.$$

Call the image of the unit circle $\mathbf{T}$ under a the *symbol curve*, $a(\mathbf{T})$.

Spectrum of a General Toeplitz Operator

Theorem (Spectrum of a Toeplitz Operator)

Suppose the Toeplitz operator $\mathbf{A}_\infty : \ell^2(N) \to \ell^2(N)$ *has a symbol that is continuous. Then*

$$\sigma(\mathbf{A}_\infty) = a(\boldsymbol{T}) \cup \{\textit{all points } a(\boldsymbol{T}) \textit{ encloses with nonzero winding number}\}.$$

Due variously to: Wintner; Gohberg; Krein; Calderón, Spitzer, & Widom.

See Albrecht Böttcher and colleagues for many more details.

Spectrum of a General Toeplitz Operator

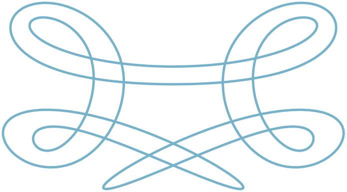
Theorem (Spectrum of a Toeplitz Operator)

Suppose the Toeplitz operator $\mathbf{A}_\infty : \ell^2(N) \to \ell^2(N)$ has a symbol that is continuous. Then

$$\sigma(\mathbf{A}_\infty) = a(\boldsymbol{T}) \cup \{\textit{all points } a(\boldsymbol{T}) \textit{ encloses with nonzero winding number}\}.$$

Due variously to: Wintner; Gohberg; Krein; Calderón, Spitzer, & Widom.

See Albrecht Böttcher and colleagues for many more details.



$a(\mathbf{T})$

Spectrum of a General Toeplitz Operator

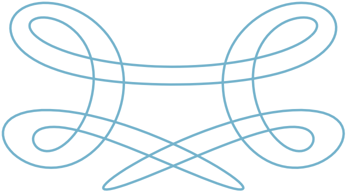
Theorem (Spectrum of a Toeplitz Operator)

Suppose the Toeplitz operator $\mathbf{A}_\infty : \ell^2(N) \to \ell^2(N)$ has a symbol that is continuous. Then

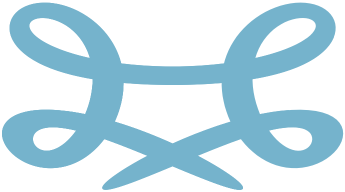
$$\sigma(\mathbf{A}_\infty) = a(\boldsymbol{T}) \cup \{\textit{all points } a(\boldsymbol{T}) \textit{ encloses with nonzero winding number}\}.$$

Due variously to: Wintner; Gohberg; Krein; Calderón, Spitzer, & Widom.

See Albrecht Böttcher and colleagues for many more details.



$a(\mathbf{T})$



$\sigma(\mathbf{A}_\infty)$

Spectrum of a General Toeplitz Matrix

What can be said of the eigenvalues of a finite-dimensional Toeplitz matrix?

Theorem (Limiting Spectrum of Finite Toeplitz Matrices)

Consider the family of banded Toeplitz matrices $\{\mathbf{A}_n\}_{n \in N}$ with upper bandwidth b and lower bandwidth d. For any fixed $\lambda \in C$, label the roots $\zeta_1, \ldots, \zeta_{b+d}$ of the polynomial $z^d(a(z) - \lambda)$ by increasing modulus.

If $|\zeta_d| = |\zeta_{d+1}|$, then $\lambda \in \lim_{n \to \infty} \sigma(\mathbf{A}_n)$.

This result, proved by [Schmidt & Spitzer, 1960], shows that in general:

$$\lim_{n \to \infty} \sigma(\mathbf{A}_n) \neq \sigma(\mathbf{A}_\infty).$$

Spectrum of a General Toeplitz Matrix

What can be said of the eigenvalues of a finite-dimensional Toeplitz matrix?

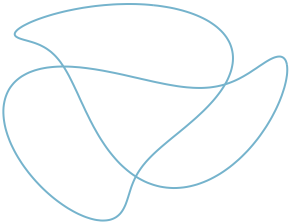
Theorem (Limiting Spectrum of Finite Toeplitz Matrices)

Consider the family of banded Toeplitz matrices $\{\mathbf{A}_n\}_{n\in N}$ with upper bandwidth b and lower bandwidth d. For any fixed $\lambda \in C$, label the roots $\zeta_1, \ldots, \zeta_{b+d}$ of the polynomial $z^d(a(z) - \lambda)$ by increasing modulus.

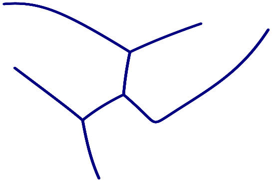
If $|\zeta_d| = |\zeta_{d+1}|$, then $\lambda \in \lim_{n\to\infty} \sigma(\mathbf{A}_n)$.

This result, proved by [Schmidt & Spitzer, 1960], shows that in general:

$$\lim_{n\to\infty} \sigma(\mathbf{A}_n) \neq \sigma(\mathbf{A}_\infty).$$

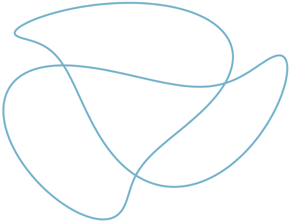


$a(\mathbf{T})$

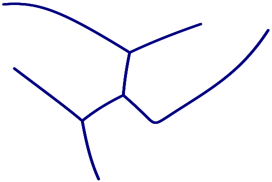


$\lim_{n\to\infty} \sigma(\mathbf{A}_n)$

Spectrum of a General Toeplitz Matrix: Example



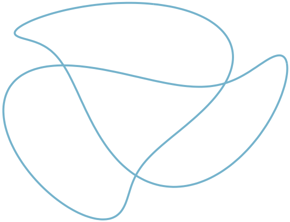
$a(\mathbb{T})$



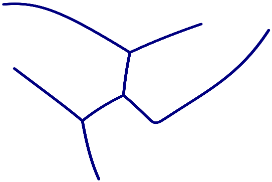
$\lim_{n \to \infty} \sigma(\mathbf{A}_n)$

$\sigma(\mathbf{A}_{50})$

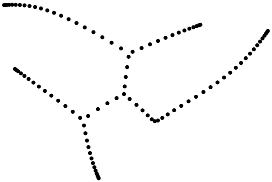
Spectrum of a General Toeplitz Matrix: Example



$a(\mathbb{T})$

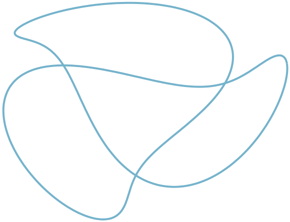


$\lim_{n\to\infty} \sigma(\mathbf{A}_n)$

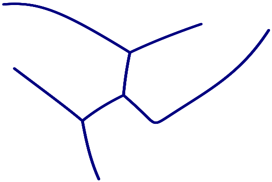


$\sigma(\mathbf{A}_{100})$

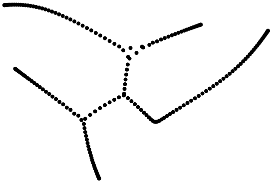
Spectrum of a General Toeplitz Matrix: Example



$a(\mathbb{T})$

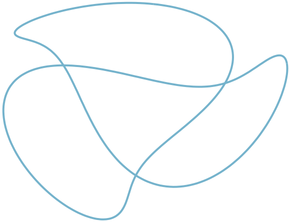


$\lim_{n\to\infty} \sigma(\mathbf{A}_n)$

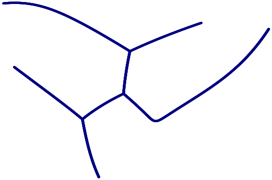


$\sigma(\mathbf{A}_{200})$

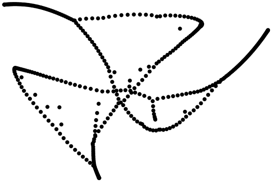
Spectrum of a General Toeplitz Matrix: Example



$a(\mathbb{T})$

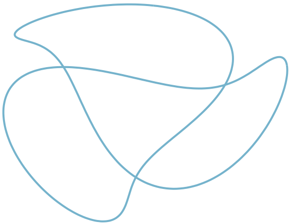


$\lim_{n\to\infty} \sigma(\mathbf{A}_n)$

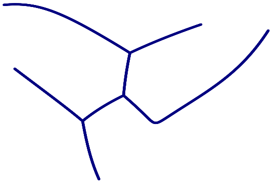


$\sigma(\mathbf{A}_{400})$

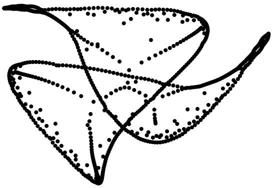
Spectrum of a General Toeplitz Matrix: Example



$a(\mathbb{T})$



$\lim_{n\to\infty} \sigma(\mathbf{A}_n)$



$\sigma(\mathbf{A}_{800})$

Pseudospectra of Toeplitz Matrices

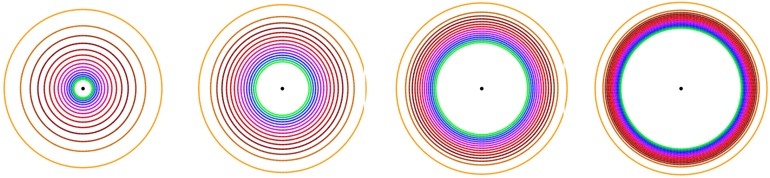
Theorem (Landau; Reichel and Trefethen; Böttcher)

Let $a(z) = \sum_{k=-M}^{M} a_k z^k$ be the symbol of a banded Toeplitz operator.

- *The pseudospectra of $\mathbf{A}_n$ converge to the pseudospectra of the Toeplitz operator on $\ell^2(\mathbf{N})$ as $n \to \infty$.*

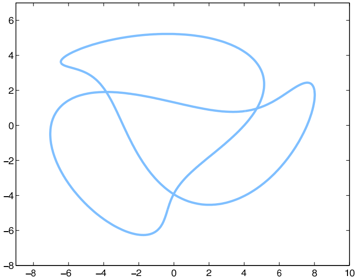
Let $z \in \boldsymbol{C}$ have nonzero winding number w.r.t. the symbol curve $a(\boldsymbol{T})$.

- *$\|(z - \mathbf{A}_n)^{-1}\|$ grows exponentially in n.*
- *For all $\varepsilon > 0$, $z \in \sigma_\varepsilon(\mathbf{A}_n)$ for all n sufficiently large.*

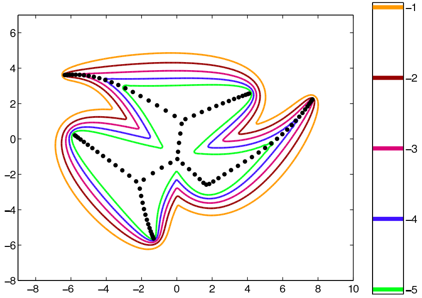


$n = 16$ $n = 32$ $n = 64$ $n = 128$

Pseudospectra of Toeplitz Matrices

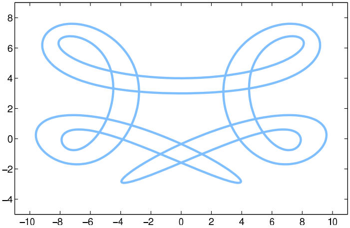


symbol curve

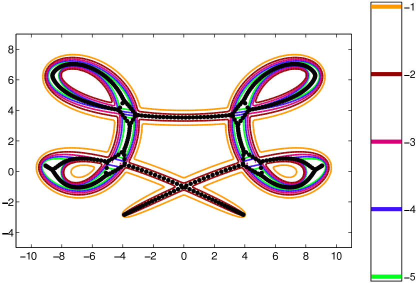


$\sigma_\varepsilon(\mathbf{A}_{100})$

Pseudospectra of Toeplitz Matrices



symbol curve



$\sigma_\varepsilon(\mathbf{A}_{500})$

Hermitian Toeplitz Matrices

We have seen “large” pseudospectra arise for generic Toeplitz matrices.

But what about Hermitian Toeplitz matrices?

For example, the matrix

$$\begin{bmatrix} 0 & 1 & & \\ 1 & 0 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 0 \end{bmatrix}$$

has symbol $a(z) = z^{-1} + z$, so

$$a(e^{i\theta}) = e^{-i\theta} + e^{i\theta} = 2\cos(\theta) \in [-2, 2] \subset \mathbf{R}.$$

Hermitian Toeplitz Matrices

We have seen "large" pseudospectra arise for generic Toeplitz matrices.

But what about Hermitian Toeplitz matrices?

For example, the matrix

$$\begin{bmatrix} 0 & 1 & & \\ 1 & 0 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 0 \end{bmatrix}$$

has symbol $a(z) = z^{-1} + z$, so

$$a(e^{i\theta}) = e^{-i\theta} + e^{i\theta} = 2\cos(\theta) \in [-2, 2] \subset \mathbf{R}.$$

In general, Hermitian symbols have $a_{-k} = \overline{a_k}$, so

$$\begin{aligned} a(e^{i\theta}) &= a_0 + \sum_{k=-\infty}^{-1} a_k e^{ik\theta} + \sum_{k=1}^{\infty} a_k e^{ik\theta} \\ &= a_0 + \sum_{k=1}^{\infty} \overline{a_k} e^{-ik\theta} + \sum_{k=1}^{\infty} a_k e^{ik\theta} = a_0 + \sum_{k=1}^{\infty} 2\mathrm{Re}(a_k e^{ik\theta}) \in \mathbf{R}. \end{aligned}$$

As $n \to \infty$, eigenvalues of $\mathbf{A}_n$ distribute according to Szegő's theorem.

Tridiagonal Toeplitz Matrices

Consider the case of a tridiagonal Toeplitz matrix,

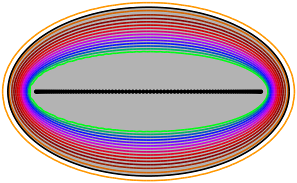
$$\mathbf{A}_n = \begin{bmatrix} \beta & \gamma & & \\ \alpha & \beta & \ddots & \\ & \ddots & \ddots & \gamma \\ & & \alpha & \beta \end{bmatrix} \in \mathbf{C}^{n\times n}$$

with symbol

$$a(z) = \frac{\alpha}{z} + \beta + \gamma z.$$

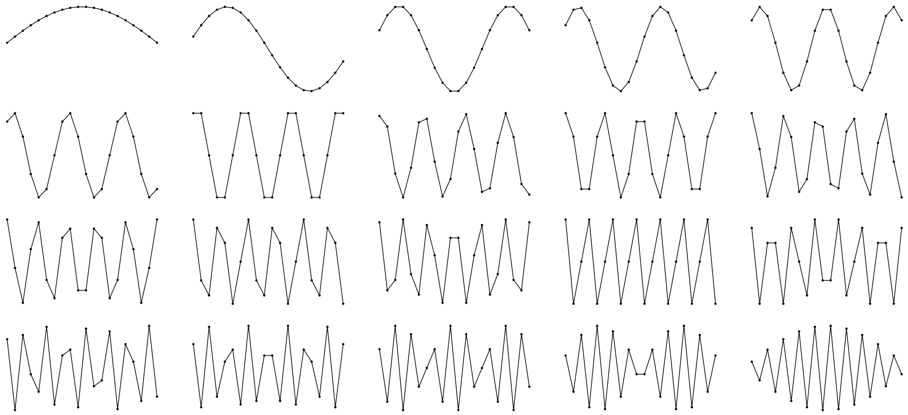
The symbol curve is the ellipse

$$a(\mathbf{T}) = \{z \in \mathbf{T} : \frac{\alpha}{z} + \beta + \gamma z\}.$$



Eigenvectors of Tridiagonal Toeplitz Matrices

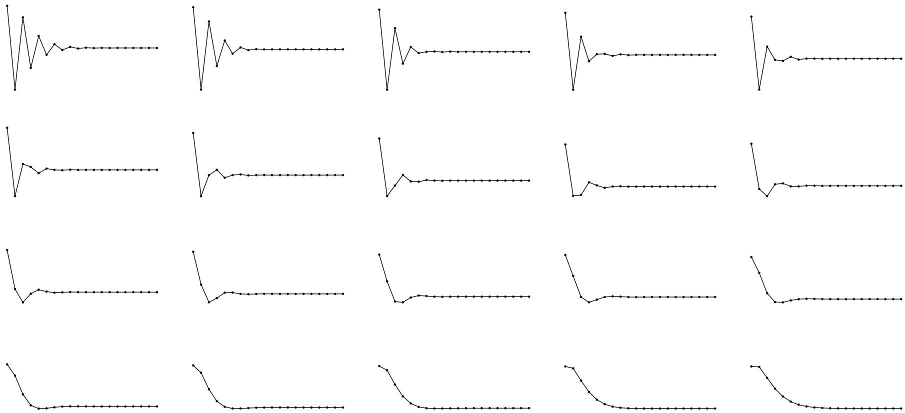
$$\mathbf{A} = \text{tridiag}(1, 0, 1),\ n = 20$$



Normal matrix: orthogonal eigenvectors

Eigenvectors of Tridiagonal Toeplitz Matrices

$$\mathbf{A} = \mathrm{tridiag}(2, 0, 1/2),\ n = 20$$



Nonnormal matrix: non-orthogonal eigenvectors

Circulant Matrices and Laurent Operators

In contrast, consider the *circulant matrix*

$$\mathbf{C}_n = \begin{bmatrix} a_0 & a_1 & \cdots & a_{-2} & a_{-1} \\ a_{-1} & a_0 & \ddots & \ddots & a_{-2} \\ \vdots & \ddots & \ddots & a_1 & \vdots \\ a_2 & \ddots & a_{-1} & a_0 & a_1 \\ a_1 & a_2 & \cdots & a_{-1} & a_0 \end{bmatrix} \in \mathbb{C}^{n\times n}$$

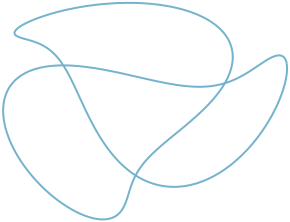
Circulant Matrices and Laurent Operators

In contrast, consider the *circulant matrix*

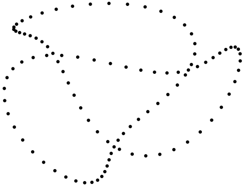
$$\mathbf{C}_n = \begin{bmatrix} a_0 & a_1 & \cdots & a_{-2} & a_{-1} \\ a_{-1} & a_0 & \ddots & \ddots & a_{-2} \\ \vdots & \ddots & \ddots & a_1 & \vdots \\ a_2 & \ddots & a_{-1} & a_0 & a_1 \\ a_1 & a_2 & \cdots & a_{-1} & a_0 \end{bmatrix} \in \mathbf{C}^{n\times n}$$

- $\mathbf{C}_n$ is *normal* for all symbols.
- $\mathbf{C}_n$ is diagonalized by the Discrete Fourier Transform matrix.
- $\sigma(\mathbf{C}_n) = \{a(z) : z \in \mathbf{T}_n\}$, where $\mathbf{T}_n := \{e^{2k\pi i/n}, k = 0, \ldots, n-1\}$: i.e., $\sigma(\mathbf{C}_n)$ comprises the image of the nth roots of unity under the symbol.
- Infinite dimensional generalization: Laurent operators $\mathbf{C}_\infty$ on $\ell^2(\mathbf{Z})$ (doubly-infinite matrices) with spectrum $\sigma(\mathbf{C}_\infty) = a(\mathbf{T})$.

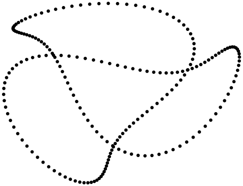
Circulant Matrix and Laurent Operators



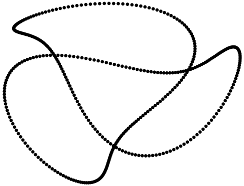
$\sigma(\mathbf{C}_\infty) = a(\mathbb{T})$



$\sigma(\mathbf{C}_{100})$



$\sigma(\mathbf{C}_{200})$



$\sigma(\mathbf{C}_{400})$

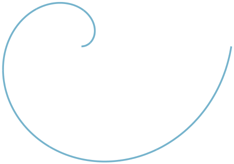
Piecewise Continuous Symbols

It is possible for the symbol

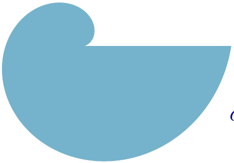
$$a(z) = \sum_{k=-\infty}^{\infty} a_k z^k$$

to, e.g., have a jump discontinuity. This compromises the exponential growth of the norm of the resolvent [Böttcher, E., Trefethen, 2002].

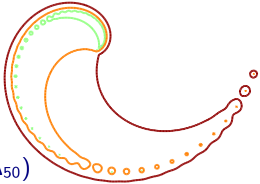
For example, take $a(e^{i\theta}) = \theta e^{i\theta}$, a symbol studied by [Basor & Morrison, 1994].



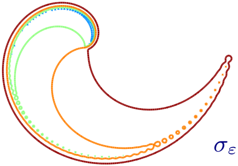
$a(\mathbf{T})$



$\sigma(\mathbf{A}_\infty)$



$\sigma_\varepsilon(\mathbf{A}_{50})$



$\sigma_\varepsilon(\mathbf{A}_{100})$

"Twisted" Toeplitz Matrices

A "twisted" Toeplitz matrix is a Toeplitz-like matrix with varying coefficients [Trefethen & Chapman, 2004].

For example, for $x_j = 2\pi j/n$, set

$$\mathbf{A}_n = \begin{bmatrix} x_1 & \frac{1}{2}x_1 & & \\ & x_2 & \ddots & \\ & & \ddots & \frac{1}{2}x_{n-1} \\ & & & x_n \end{bmatrix}.$$

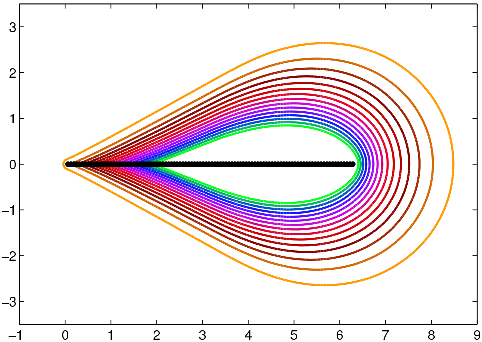
The "symbol" now depends on two variables: $a(x,z) = x + \frac{1}{2}xz$.

The pseudospectra resemble those of standard Toeplitz matrices, but the (pseudo)-eigenvectors have an entirely different character.

"Twisted" Toeplitz Matrices

For $x_j = 2\pi j/n$, set

$$\mathbf{A}_n = \begin{bmatrix} x_1 & \frac{1}{2}x_1 & & \\ & x_2 & \ddots & \\ & & \ddots & \frac{1}{2}x_{n-1} \\ & & & x_n \end{bmatrix}.$$

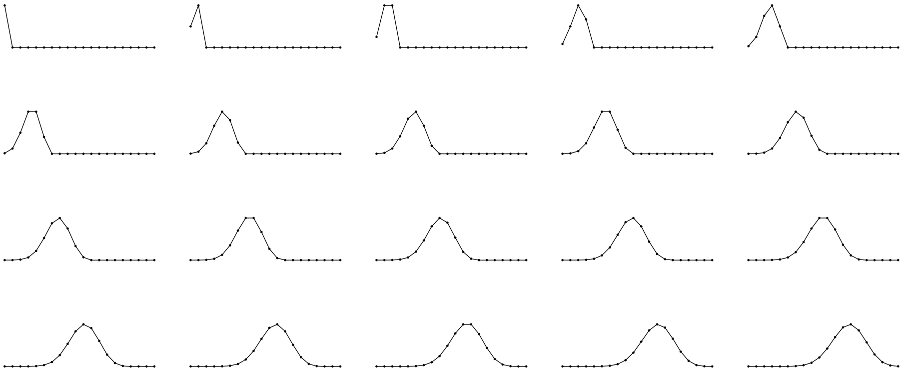


Pseudospectra of $\sigma_\varepsilon(\mathbf{A}_{100})$

Eigenvectors of a Twisted Toeplitz Matrix

Eigenvectors form *wave packets* for twisted Toeplitz matrices.

Pseudoeigenvectors for $z \in \sigma_\varepsilon(\mathbf{A})$ have a similar form.



Eigenvectors of $\mathbf{A}_{20}$ for $a(x,z) = x + \frac{1}{2}xz$

Looking Forward to Differential Operators

Where do Toeplitz (and twisted Toeplitz) matrices come from?

Numerous applications – be we will highlight just one of them: discretization of differential operators.

Consider the steady–state convection diffusion equation $Lu = f$, where

$$Lu = u'' + c\,u'$$

posed over for $x \in [0, 1]$ with $u(0) = u(1) = 0$.

- Fix a discretization parameter n
- Approximate the problem on a simple grid $\{x_j\}_{j=0}^{n+1}$ with spacing $h = 1/(n+1)$:
$$x_j = j\,h.$$
- Replace the first and second derivatives with second-order accurate formulas on the grid:

$$\begin{aligned} u'(x_j) &= \frac{u(x_{j+1}) - u(x_{j+1})}{2h} + \mathcal{O}(h^2) \\ u''(x_j) &= \frac{u(x_{j-1}) - 2u(x_j) + u(x_{j+1})}{h^2} + \mathcal{O}(h^2). \end{aligned}$$

Finite Difference Discretizations

Now let $u_j \approx u(x_j)$ with $u_0 = u_{n+1} = 0$, so that

$$\begin{aligned} u'(x_j) &\approx \frac{u_{j+1} - u_{j+1}}{2h} \\ u''(x_j) &\approx \frac{u_{j-1} - 2u_j + u_{j+1}}{h^2}. \end{aligned}$$

Approximate $Lu = f$, with

$$Lu = u'' + c\,u', \qquad u(0) = u(1) = 0,$$

via

$$\frac{1}{h^2}\begin{bmatrix} -2 & 1 + ch/2 & & \\ 1 - ch/2 & -2 & \ddots & \\ & \ddots & \ddots & 1 + ch/2 \\ & & 1 - ch/2 & -2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}.$$

Label these components: $\mathbf{A}_n \mathbf{u} = \mathbf{f}$.
$\mathbf{A}_n$ is Toeplitz for this constant-coefficient differential operator.

Finite Difference Discretizations

$$\mathbf{A}_n := \frac{1}{h^2}\begin{bmatrix} -2 & 1+ch/2 & & \\ 1-ch/2 & -2 & \ddots & \\ & \ddots & \ddots & 1+ch/2 \\ & & 1-ch/2 & -2 \end{bmatrix}$$

Notice that the symbol of $\mathbf{A}_n$ depends on n (recall $h = 1/(n+1)$):

$$a_n(z) = \left(\frac{1}{h^2} - \frac{c}{2h}\right) z^{-1} - \left(\frac{2}{h^2}\right) + \left(\frac{1}{h^2} + \frac{c}{2h}\right) z$$

Finite Difference Discretizations

$$\mathbf{A}_n := \frac{1}{h^2}\begin{bmatrix} -2 & 1+ch/2 & & \\ 1-ch/2 & -2 & \ddots & \\ & \ddots & \ddots & 1+ch/2 \\ & & 1-ch/2 & -2 \end{bmatrix}$$

Notice that the symbol of $\mathbf{A}_n$ depends on n (recall $h = 1/(n+1)$):

$$a_n(z) = \left(\frac{1}{h^2} - \frac{c}{2h}\right) z^{-1} - \left(\frac{2}{h^2}\right) + \left(\frac{1}{h^2} + \frac{c}{2h}\right) z$$

- For all n, the symbol curve $a_n(\mathbb{T})$ is an ellipse.

Finite Difference Discretizations

$$\mathbf{A}_n := \frac{1}{h^2}\begin{bmatrix} -2 & 1+ch/2 & & \\ 1-ch/2 & -2 & \ddots & \\ & \ddots & \ddots & 1+ch/2 \\ & & 1-ch/2 & -2 \end{bmatrix}$$

Notice that the symbol of $\mathbf{A}_n$ depends on n (recall $h = 1/(n+1)$):

$$a_n(z) = \left(\frac{1}{h^2} - \frac{c}{2h}\right) z^{-1} - \left(\frac{2}{h^2}\right) + \left(\frac{1}{h^2} + \frac{c}{2h}\right) z$$

- For all n, the symbol curve $a_n(\mathbb{T})$ is an ellipse.
- If $c = 0$ (no convection), then $a_n(\mathbb{T})$ is a real line segment: $\mathbf{A}_n$ and L are both self-adjoint, hence normal.
- If $c \neq 0$, the eigenvalues of $\mathbf{A}_n$ will be real, but the pseudospectra can be far from these eigenvalues.

Finite Difference Discretizations

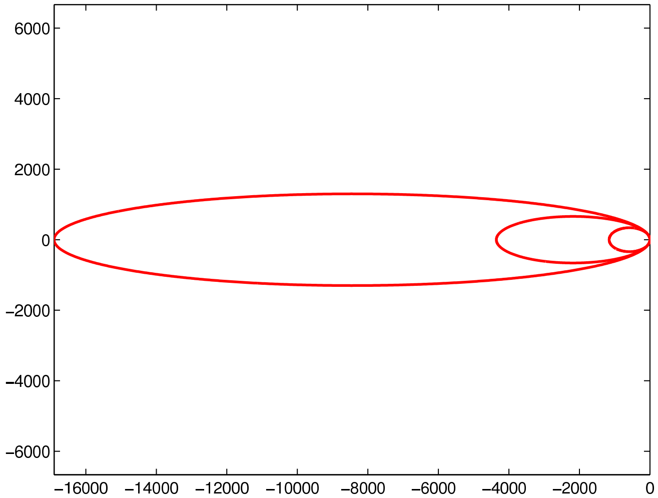
$$\mathbf{A}_n := \frac{1}{h^2}\begin{bmatrix} -2 & 1+ch/2 & & \\ 1-ch/2 & -2 & \ddots & \\ & \ddots & \ddots & 1+ch/2 \\ & & 1-ch/2 & -2 \end{bmatrix}$$

Notice that the symbol of $\mathbf{A}_n$ depends on n (recall $h = 1/(n+1)$):

$$a_n(z) = \left(\frac{1}{h^2} - \frac{c}{2h}\right) z^{-1} - \left(\frac{2}{h^2}\right) + \left(\frac{1}{h^2} + \frac{c}{2h}\right) z$$

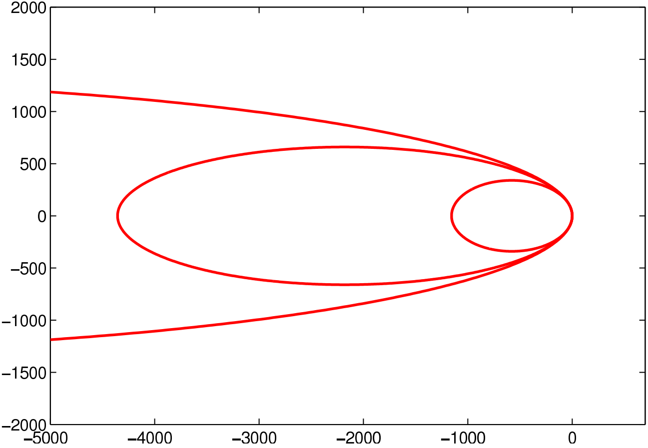
- For all n, the symbol curve $a_n(\mathbb{T})$ is an ellipse.
- If $c = 0$ (no convection), then $a_n(\mathbb{T})$ is a real line segment: $\mathbf{A}_n$ and L are both self-adjoint, hence normal.
- If $c \neq 0$, the eigenvalues of $\mathbf{A}_n$ will be real, but the pseudospectra can be far from these eigenvalues.
- Rightmost part of $\sigma_\varepsilon(\mathbf{A}_n)$ approximates the corresponding part of $\sigma_\varepsilon(L)$.

Finite Differences: Symbol Curves



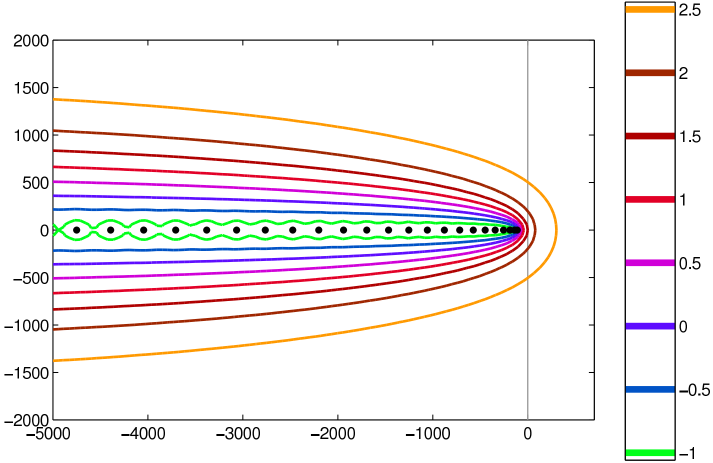
Symbol curves $a_n(\mathbf{T})$ for $n = 16, 32, 64$

Finite Differences: Symbol Curves



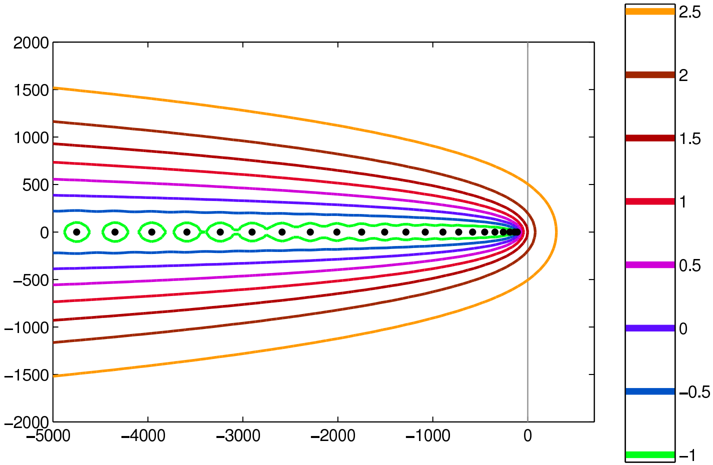
Symbol curves $a_n(\mathbf{T})$ for $n = 16, 32, 64$

Finite Differences: Approximate Pseudospectra



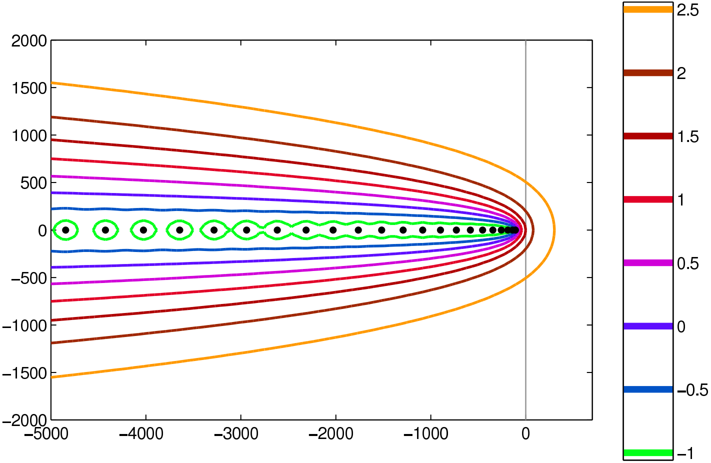
Rightmost part of $\sigma_\varepsilon(\mathbf{A}_n)$ for $n = 64$

Finite Differences: Approximate Pseudospectra



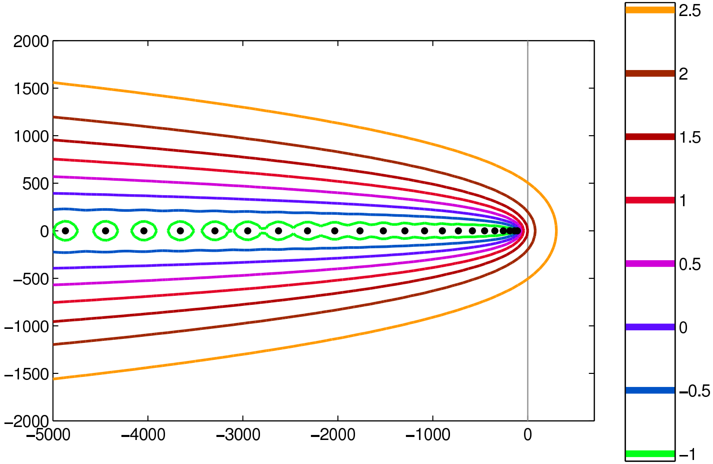
Rightmost part of $\sigma_\varepsilon(\mathbf{A}_n)$ for $n = 128$

Finite Differences: Approximate Pseudospectra



Rightmost part of $\sigma_\varepsilon(\mathbf{A}_n)$ for $n = 256$

Finite Differences: Approximate Pseudospectra



Rightmost part of $\sigma_\varepsilon(\mathbf{A}_n)$ for $n = 512$

3(b) Balanced Truncation Model Reduction

Balanced Truncation Model Reduction

Consider the single-input, single-output (SISO) linear dynamical system:

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

$\mathbf{A} \in \mathbb{C}^{n\times n}$, $\mathbf{b}, \mathbf{c}^{\mathsf{T}} \in \mathbb{C}^n$. We assume that $\mathbf{A}$ is stable: $\alpha(\mathbf{A}) < 0$.

We wish to reduce the dimension of the dynamical system by projecting onto well-chosen subspaces.

Balanced truncation: Change basis to match states that are easy to *reach* and easy to *observe*, then project onto that prominent subspace.

Controllability and Observability Gramians

To guage the *observability* of an initial state $\mathbf{x}_0 = \widehat{\mathbf{x}}$,
measure the energy in its output (when there is no input, $u = 0$):

$$y(t) = \mathbf{c}\mathrm{e}^{t\mathbf{A}}\widehat{\mathbf{x}}.$$

Then

$$\begin{aligned}\int_0^\infty |y(t)|^2\,\mathrm{d}t &= \int_0^\infty \widehat{\mathbf{x}}^*\mathrm{e}^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}\mathrm{e}^{t\mathbf{A}}\widehat{\mathbf{x}}\,\mathrm{d}t \\ &= \widehat{\mathbf{x}}^*\Big(\int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t\Big)\widehat{\mathbf{x}} =: \widehat{\mathbf{x}}^*\mathbf{Q}\widehat{\mathbf{x}}.\end{aligned}$$

Controllability and Observability Gramians

To guage the *observability* of an initial state $\mathbf{x}_0 = \widehat{\mathbf{x}}$,
measure the energy in its output (when there is no input, $u = 0$):

$$y(t) = \mathbf{c}\mathrm{e}^{t\mathbf{A}}\widehat{\mathbf{x}}.$$

Then

$$\begin{aligned}\int_0^\infty |y(t)|^2\,\mathrm{d}t &= \int_0^\infty \widehat{\mathbf{x}}^*\mathrm{e}^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}\mathrm{e}^{t\mathbf{A}}\widehat{\mathbf{x}}\,\mathrm{d}t \\ &= \widehat{\mathbf{x}}^*\Big(\int_0^\infty \mathrm{e}^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}\mathrm{e}^{t\mathbf{A}}\,\mathrm{d}t\Big)\widehat{\mathbf{x}} =: \widehat{\mathbf{x}}^*\mathbf{Q}\widehat{\mathbf{x}}.\end{aligned}$$

Similarly, we measure the *controllability* by the total input energy required to steer $\mathbf{x}_0 = \mathbf{0}$ to a target state $\widehat{\mathbf{x}}$ as $t \to \infty$. The special form of u that drives the system to $\widehat{\mathbf{x}}$ with minimal energy satisfies

$$\int_0^\infty |u(t)|^2\,\mathrm{d}t = \widehat{\mathbf{x}}^*\Big(\int_0^\infty \mathrm{e}^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*\mathrm{e}^{t\mathbf{A}^*}\,\mathrm{d}t\Big)^{-1}\widehat{\mathbf{x}} =: \widehat{\mathbf{x}}^*\mathbf{P}^{-1}\widehat{\mathbf{x}}.$$

Controllability and Observability Gramians

To guage the *observability* of an initial state $\mathbf{x}_0 = \widehat{\mathbf{x}}$,
measure the energy in its output (when there is no input, $u = 0$):

$$y(t) = \mathbf{c}e^{t\mathbf{A}}\widehat{\mathbf{x}}.$$

Then

$$\begin{aligned} \int_0^\infty |y(t)|^2\,\mathrm{d}t &= \int_0^\infty \widehat{\mathbf{x}}^* e^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}e^{t\mathbf{A}}\widehat{\mathbf{x}}\,\mathrm{d}t \\ &= \widehat{\mathbf{x}}^*\Big(\int_0^\infty e^{t\mathbf{A}^*}\mathbf{c}^*\mathbf{c}e^{t\mathbf{A}}\,\mathrm{d}t\Big)\widehat{\mathbf{x}} =: \widehat{\mathbf{x}}^*\mathbf{Q}\widehat{\mathbf{x}}. \end{aligned}$$

Similarly, we measure the *controllability* by the total input energy required to steer $\mathbf{x}_0 = \mathbf{0}$ to a target state $\widehat{\mathbf{x}}$ as $t \to \infty$. The special form of u that drives the system to $\widehat{\mathbf{x}}$ with minimal energy satisfies

$$\int_0^\infty |u(t)|^2\,\mathrm{d}t = \widehat{\mathbf{x}}^*\Big(\int_0^\infty e^{t\mathbf{A}}\mathbf{b}\mathbf{b}^*e^{t\mathbf{A}^*}\,\mathrm{d}t\Big)^{-1}\widehat{\mathbf{x}} =: \widehat{\mathbf{x}}^*\mathbf{P}^{-1}\widehat{\mathbf{x}}.$$

Thus we have the infinite controllability and observability gramians $\mathbf{P}$ and $\mathbf{Q}$:

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

See, e.g., [Antoulas, 2005].

Balanced Truncation Model Reduction

The 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) can be determined by solving the Lyapunov equations – see the next lecture.

If $\mathbf{x}_0 = \mathbf{0}$, the minimum energy of u required to drive $\mathbf{x}$ to state $\widehat{\mathbf{x}}$ is

$$\widehat{\mathbf{x}}^*\mathbf{P}^{-1}\widehat{\mathbf{x}}.$$

Starting from $\mathbf{x}_0 = \widehat{\mathbf{x}}$ with $u(t) \equiv 0$, the energy of output y is

$$\widehat{\mathbf{x}}^*\mathbf{Q}\,\widehat{\mathbf{x}}.$$

$\widehat{\mathbf{x}}^*\mathbf{P}^{-1}\widehat{\mathbf{x}}$: $\widehat{\mathbf{x}}$ is *hard to reach* if it is rich in the lowest modes of $\mathbf{P}$.
$\widehat{\mathbf{x}}^*\mathbf{Q}\,\widehat{\mathbf{x}}$: $\widehat{\mathbf{x}}$ is *hard to observe* if it is rich in the lowest modes of $\mathbf{Q}$.

Balanced truncation transforms the state space coordinate system to make these two gramians the same, then it truncates the lowest modes.

Balanced Truncation Model Reduction

Consider a generic coordinate transformation, for $\mathbf{S}$ invertible:

$$\begin{aligned}(\mathbf{S}\mathbf{x})'(t) &= (\mathbf{S}\mathbf{A}\mathbf{S}^{-1})(\mathbf{S}\mathbf{x}(t)) + (\mathbf{S}\mathbf{b})u(t)\\ y(t) &= (\mathbf{c}\mathbf{S}^{-1})(\mathbf{S}\mathbf{x}(t)) + du(t), \qquad (\mathbf{S}\mathbf{x})(0) = \mathbf{S}\mathbf{x}_0.\end{aligned}$$

With this transformation, the controllability and observability gramians are

$$\widehat{\mathbf{P}} = \mathbf{S}\mathbf{P}\mathbf{S}^*, \qquad \widehat{\mathbf{Q}} = \mathbf{S}^{-*}\mathbf{Q}\mathbf{S}^{-1}.$$

For balancing, we seek $\mathbf{S}$ so that $\widehat{\mathbf{P}} = \widehat{\mathbf{Q}}$ are diagonal.

Balanced Truncation Model Reduction

Consider a generic coordinate transformation, for $\mathbf{S}$ invertible:

$$\begin{aligned}(\mathbf{S}\mathbf{x})'(t) &= (\mathbf{S}\mathbf{A}\mathbf{S}^{-1})(\mathbf{S}\mathbf{x}(t)) + (\mathbf{S}\mathbf{b})u(t)\\ y(t) &= (\mathbf{c}\mathbf{S}^{-1})(\mathbf{S}\mathbf{x}(t)) + du(t), \qquad (\mathbf{S}\mathbf{x})(0) = \mathbf{S}\mathbf{x}_0.\end{aligned}$$

With this transformation, the controllability and observability gramians are

$$\widehat{\mathbf{P}} = \mathbf{S}\mathbf{P}\mathbf{S}^*, \qquad \widehat{\mathbf{Q}} = \mathbf{S}^{-*}\mathbf{Q}\mathbf{S}^{-1}.$$

For balancing, we seek $\mathbf{S}$ so that $\widehat{\mathbf{P}} = \widehat{\mathbf{Q}}$ are diagonal.

Observation (How does nonnormality affect balancing?)

- $\sigma_{\varepsilon/\kappa(\mathbf{S})}(\mathbf{S}\mathbf{A}\mathbf{S}^{-1}) \subseteq \sigma_\varepsilon(\mathbf{A}) \subseteq \sigma_{\varepsilon\kappa(\mathbf{S})}(\mathbf{S}\mathbf{A}\mathbf{S}^{-1})$.
- *The choice of internal coordinates will affect* $\mathbf{P}$, $\mathbf{Q}$, …

Balanced Truncation Model Reduction

Consider a generic coordinate transformation, for $\mathbf{S}$ invertible:

$$\begin{aligned}(\mathbf{S}\mathbf{x})'(t) &= (\mathbf{S}\mathbf{A}\mathbf{S}^{-1})(\mathbf{S}\mathbf{x}(t)) + (\mathbf{S}\mathbf{b})u(t)\\ y(t) &= (\mathbf{c}\mathbf{S}^{-1})(\mathbf{S}\mathbf{x}(t)) + du(t), \qquad (\mathbf{S}\mathbf{x})(0) = \mathbf{S}\mathbf{x}_0.\end{aligned}$$

With this transformation, the controllability and observability gramians are

$$\widehat{\mathbf{P}} = \mathbf{S}\mathbf{P}\mathbf{S}^*, \qquad \widehat{\mathbf{Q}} = \mathbf{S}^{-*}\mathbf{Q}\mathbf{S}^{-1}.$$

For balancing, we seek $\mathbf{S}$ so that $\widehat{\mathbf{P}} = \widehat{\mathbf{Q}}$ are diagonal.

Observation (How does nonnormality affect balancing?)

- $\sigma_{\varepsilon/\kappa(\mathbf{S})}(\mathbf{S}\mathbf{A}\mathbf{S}^{-1}) \subseteq \sigma_\varepsilon(\mathbf{A}) \subseteq \sigma_{\varepsilon\kappa(\mathbf{S})}(\mathbf{S}\mathbf{A}\mathbf{S}^{-1})$.
- *The choice of internal coordinates will affect* $\mathbf{P}$, $\mathbf{Q}$, ...
- *but not the Hankel singular values:* $\widehat{\mathbf{P}}\widehat{\mathbf{Q}} = \mathbf{S}\mathbf{P}\mathbf{Q}\mathbf{S}^{-1}$,
- *and not the transfer function:*
$$(\mathbf{c}\mathbf{S}^{-1})(z - \mathbf{S}\mathbf{A}\mathbf{S}^{-1})^{-1}(\mathbf{S}\mathbf{b}) = d + \mathbf{c}(z - \mathbf{A})^{-1}\mathbf{b},$$
- *and not the system moments:*
$$(\mathbf{c}\mathbf{S}^{-1})(\mathbf{S}\mathbf{b}) = \mathbf{c}\mathbf{b}, \quad (\mathbf{c}\mathbf{S}^{-1})(\mathbf{S}\mathbf{A}\mathbf{S}^{-1})(\mathbf{S}\mathbf{b}) = \mathbf{c}\mathbf{A}\mathbf{b}, \quad \ldots.$$