# Tridiagonal Toeplitz Matrices: Properties and Novel Applications

Silvia Noschese[1] Lionello Pasquini[2] and Lothar Reichel[3*]

[1] *Dipartimento di Matematica "Guido Castelnuovo", SAPIENZA Università di Roma, P.le A. Moro, 2, I-00185 Roma, Italy. E-mail:* noschese@mat.uniroma1.it. *Research supported by a grant from SAPIENZA Università di Roma.*

[2] *Dipartimento di Matematica "Guido Castelnuovo", SAPIENZA Università di Roma, P.le A. Moro, 2, I-00185 Roma, Italy. E-mail:* Lionello.Pasquini@uniroma1.it.

[3] *Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail:* reichel@math.kent.edu. *Research supported in part by NSF grant DMS-1115385.*

*Dedicated to Biswa N. Datta on the Occasion of His 70th Birthday.*

KEY WORDS: Eigenvalues, conditioning, Toeplitz matrix, matrix nearness problem, distance to normality, inverse eigenvalue problem, Krylov subspace bases, Tikhonov regularization

SUMMARY

The eigenvalues and eigenvectors of tridiagonal Toeplitz matrices are known in closed form. This property is in the first part of the paper used to investigate the sensitivity of the spectrum. Explicit expressions for the structured distance to the closest normal matrix, the departure from normality, and the $\varepsilon$-pseudospectrum are derived. The second part of the paper discusses applications of the theory to inverse eigenvalue problems, the construction of Chebyshev polynomial-based Krylov subspace bases, and Tikhonov regularization. 

## 1. Introduction

Tridiagonal Toeplitz matrices and low-rank perturbations of such matrices arise in numerous applications, including the solution of ordinary and partial differential equations [12, 15, 37, 41], time series analysis [26], and as regularization matrices in Tikhonov regularization for the solution of discrete ill-posed problems [17, 33]. It is therefore important to understand properties of tridiagonal Toeplitz matrices relevant for computation.

The eigenvalues of real and complex tridiagonal Toeplitz matrices can be very sensitive to perturbations of the matrix. Using explicit formulas for the eigenvalues and eigenvectors of tridiagonal Toeplitz matrices, we derive explicit expressions that shed light on this sensitivity. Exploiting the Toeplitz and tridiagonal structures, we derive simple formulas for the distance to normality, the structured distance to normality, the departure from normality, and the $\varepsilon$-pseudospectrum, as well as for individual and global eigenvalue condition numbers. These quantities provide us with a thorough understanding of the sensitivity of the eigenvalues of tridiagonal Toeplitz matrices. In particular, we show that the sensitivity of the eigenvalues

Table I. Definitions of sets used in the paper.

| | |
|---|---|
| $\mathcal{T}$ | the subspace of $\mathbb{C}^{n\times n}$ formed by tridiagonal Toeplitz matrices |
| $\mathcal{N}$ | the algebraic variety of normal matrices in $\mathbb{C}^{n\times n}$ |
| $\mathcal{N}_\mathcal{T}$ | $\mathcal{N}\cap\mathcal{T}$ |
| $\mathcal{M}$ | the algebraic variety of matrices in $\mathbb{C}^{n\times n}$ with multiple eigenvalues |
| $\mathcal{M}_\mathcal{T}$ | $\mathcal{M}\cap\mathcal{T}$ |

grows exponentially with the ratio of the absolute values of the sub- and super-diagonal matrix entries; the sensitivity of the eigenvalues is independent of the diagonal entry and of the arguments of off diagonal entries. The distance to normality also depends on the difference between the absolute values of the sub- and super-diagonal entries.

Matrix nearness problems have received considerable attention in the literature; see, e.g., [11, 20, 25, 30, 31] and references therein. The $\varepsilon$-pseudospectra of banded Toeplitz matrices are analyzed in detail in [3, 34, 40]. Our interest in tridiagonal Toeplitz matrices stems from the possibility of deriving explicit formulas for quantities of interest and from the many applications of these matrices.

This paper is organized as follows. The eigenvalue sensitivity is investigated in Sections 2-6. Numerical illustrations also are provided. The latter part of this paper describes a few applications that are believed to be new. We consider an inverse eigenvalue problem in Section 7, where we also introduce a minimization problem, whose solution is a trapezoidal tridiagonal Toeplitz matrix. The latter matrices can be applied as regularization matrices in Tikhonov regularization. This application is described in Section 8. Section 9 is concerned with the construction of nonorthogonal Krylov subspace bases based on the recursion formulas for suitably chosen translated and scaled Chebyshev polynomials. The use of such bases in Krylov subspace methods for the solution of large linear systems of equations or for the computation of a few eigenvalues of a large matrix is attractive in parallel computing environments that do not allow efficient execution of the Arnoldi process for generating an orthonormal basis; see [21, 22, 32] for discussions. We describe how tridiagonal Toeplitz matrices can be applied to determine a suitable interval on which the translated and scaled Chebyshev polynomials are required to be orthogonal. Concluding remarks can be found in Section 10.

Several of the topics of this paper have been studied by Biswa Datta in the context of Control Theory. This includes inverse eigenvalue problems [1, 6, 7, 9] and Krylov subspace methods [8]. It is a pleasure to dedicate this paper to him.

We conclude this section by introducing notation to be used in the sequel. The Euclidean vector norm as well as the associated induced matrix norm are denoted by $\|\cdot\|_2$, and $\|\cdot\|_F$ stands for the Frobenius matrix or vector norms. Table I defines sets of interest. The distance to normality in the Frobenius norm of a matrix $A\in\mathbb{C}^{n\times n}$ is given by

$$d_F(A,\mathcal{N}) = \min_{A_\mathcal{N}\in\mathcal{N}} \|A-A_\mathcal{N}\|_F\,; \tag{1}$$

see, e.g., [13, 19, 20, 24, 30, 38] for results and discussions on the distance to normality. The

tridiagonal Toeplitz matrix

$$T = \begin{bmatrix} \delta & \tau & & & & O \\ \sigma & \delta & \tau & & & \\ & \sigma & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \tau \\ O & & & & \sigma & \delta \end{bmatrix} \in \mathbb{C}^{n\times n} \tag{2}$$

is denoted by $T = (n; \sigma, \delta, \tau)$, and we let

$$\alpha = \arg\sigma, \quad \beta = \arg\tau, \quad \gamma = \arg\delta. \tag{3}$$

The matrix $T_0 = (n; \sigma, 0, \tau)$ is of particular interest.

The quantity $d_F(T, \mathcal{N}_\mathcal{T})$ denotes the structured distance of $T \in \mathcal{T}$ to $\mathcal{N}_\mathcal{T}$ in the Frobenius norm, i.e.,

$$d_F(T, \mathcal{N}_\mathcal{T}) = \min_{T_\mathcal{N}\in\mathcal{N}_\mathcal{T}} \|T - T_\mathcal{N}\|_F .$$

Clearly, $d_F(T, \mathcal{N}_\mathcal{T}) \geq d_F(T, \mathcal{N})$ and for some matrices $T \in \mathcal{T}$, $d_F(T, \mathcal{N}_\mathcal{T})$ is much larger than $d_F(T, \mathcal{N})$. This is, for instance, the case for $T = (n; 0, \delta, \tau)$ when $\tau \neq 0$; see [30, Example 9.1].

For $T \in \mathcal{T}$, $d_F(T, \mathcal{M}_\mathcal{T})$ denotes the structured distance of $T$ to $\mathcal{M}_\mathcal{T}$ in the Frobenius norm, i.e.,

$$d_F(T, \mathcal{M}_\mathcal{T}) = \min_{T_\mathcal{M}\in\mathcal{M}_\mathcal{T}} \|T - T_\mathcal{M}\|_F .$$

## 2. Eigenvalues and eigenvectors

It is well known that the eigenvalues of $T = (n; \sigma, \delta, \tau)$ are given by

$$\lambda_h(T) = \delta + 2\sqrt{\sigma\tau}\,\cos\frac{h\pi}{n+1}, \quad h = 1 : n; \tag{4}$$

see, e.g., [37], and using (3), we obtain

$$\lambda_h(T) = \delta + 2\sqrt{|\sigma\tau|}\,e^{i(\alpha+\beta)/2}\,\cos\frac{h\pi}{n+1}\,, \quad h = 1 : n. \tag{5}$$

In particular, if $\sigma\tau \neq 0$, the matrix (2) has $n$ simple eigenvalues, which lie on the closed line segment

$$\mathcal{S}_{\lambda(T)} = \left\{ \delta + t\,e^{i(\alpha+\beta)/2} : t \in \mathbb{R},\, |t| \leq 2\sqrt{|\sigma\tau|}\,\cos\frac{\pi}{n+1} \right\} \subset \mathbb{C}. \tag{6}$$

The eigenvalues are allocated symmetrically with respect to $\delta$.

The spectral radius of the matrix (2) is given by

$$\rho(T) = \max\left\{ \left|\delta + 2\sqrt{|\sigma\tau|}e^{i(\alpha+\beta)/2}\cos\frac{\pi}{n+1}\right|, \left|\delta + 2\sqrt{|\sigma\tau|}\,e^{i(\alpha+\beta)/2}\cos\frac{n\pi}{n+1}\right| \right\}$$

and, if $T$ is nonsingular, i.e. $\lambda_h(T) \neq 0$ for all $h = 1 : n$, taking (5) into account, one has

$$\rho(T^{-1}) = \max_{h=1:n} \left|\delta + 2\sqrt{|\sigma\tau|}\,e^{i(\alpha+\beta)/2}\cos\frac{h\pi}{n+1}\right|^{-1}.$$

For $n$ odd, we have $\operatorname{rank}(T_0) = n - 1$.

When $\sigma\tau \neq 0$, the components of the right eigenvector $x_h = [x_{h,1}, x_{h,2}, \ldots, x_{h,n}]^T$ associated with the eigenvalue $\lambda_h(T)$ are given by

$$x_{h,k} = (\sigma/\tau)^{k/2} \sin \frac{hk\pi}{n+1}, \quad k = 1:n, \quad h = 1:n, \tag{7}$$

and the corresponding left eigenvector $y_h = [y_{h,1}, y_{h,2}, \ldots, y_{h,n}]^T$ has the components

$$y_{h,k} = (\bar{\tau}/\bar{\sigma})^{k/2} \sin \frac{hk\pi}{n+1}, \quad k = 1:n, \quad h = 1:n, \tag{8}$$

where the bar denotes complex conjugation. Throughout this paper the superscript $(\cdot)^T$ stands for transposition and the superscript $(\cdot)^H$ for transposition and complex conjugation.

If $\sigma = 0$ and $\tau \neq 0$ (or $\sigma \neq 0$ and $\tau = 0$), then the matrix (2) has the unique eigenvalue $\delta$ of geometric multiplicity one. The right and left eigenvectors are the first and last columns (or the last and first columns) of the identity matrix, respectively.

Note that, given the dimension of the matrix, knowing the ratio $\sigma/\tau$ is enough to uniquely determine all the right and left eigenvectors of $T$ up to a scaling factor.

## 3. Distance to and departure from normality

This section discusses the distance and structured distance of tridiagonal Toeplitz matrices to normality, as well as the departure and structured departure from normality.

**Theorem 3.1.** *The matrix (2) is normal if and only if*

$$|\sigma| = |\tau|. \tag{9}$$

**Proof:** The condition in (9) is equivalent to the equality $T^H T = T\,T^H$. ∎

The above theorem shows that a normal tridiagonal Toeplitz matrix can be written in the form

$$T' = (n; \rho e^{i\alpha'}, \delta, \rho e^{i\beta'}) = \begin{bmatrix} \delta & \rho e^{i\beta'} & & & & & O \\ \rho e^{i\alpha'} & \delta & \rho e^{i\beta'} & & & & \\ & \rho e^{i\alpha'} & \cdot & \cdot & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \rho e^{i\beta'} \\ O & & & & & \rho e^{i\alpha'} & \delta \end{bmatrix}, \tag{10}$$

where $\delta \in \mathbb{C}$, $\rho \geq 0$, and $\alpha', \beta' \in \mathbb{R}$. It follows from (5) that the eigenvalues of (10) are given by

$$\lambda_h(T') = \delta + 2\,\rho\, e^{i(\alpha'+\beta')/2} \cos \frac{h\pi}{n+1}, \qquad h = 1:n.$$

In particular, the eigenvalues lie on the closed line segment

$$\mathcal{S}_{\lambda(T')} = \left\{ \delta + t\, e^{i(\alpha'+\beta')/2} : t \in \mathbb{R},\ |t| \leq 2\,\rho \cos \frac{\pi}{n+1} \right\} \subset \mathbb{C}.$$

**Theorem 3.2.** *Let $T = (n; \sigma, \delta, \tau)$ be a matrix in $\mathcal{T}$. There is a unique matrix $T^* = (n; \sigma^*, \delta^*, \tau^*) \in \mathcal{N}_\mathcal{T}$ that minimizes $\|T_\mathcal{N} - T\|_F$ over $\mathcal{N}_\mathcal{T}$. This matrix is defined by*

$$\begin{aligned} \sigma^* &= \frac{|\sigma| + |\tau|}{2}\, e^{i\,\alpha}, \\ \delta^* &= \delta, \\ \tau^* &= \frac{|\sigma| + |\tau|}{2}\, e^{i\,\beta}, \end{aligned}$$

*where $\alpha$ and $\beta$ are given by (3).*

**Proof:** Theorem 3.1 gives the condition $|\sigma^*| = |\tau^*|$. Consequently, to minimize $\|T_\mathcal{N} - T\|_F$ over $T_\mathcal{N} \in \mathcal{N}_\mathcal{T}$, we must take

$$\delta^* = \delta, \quad \sigma^* = \rho^*\, e^{i\alpha}, \quad \tau^* = \rho^*\, e^{i\beta},$$

where $\rho^*$ denotes the common value of $|\sigma^*|$ and $|\tau^*|$. In addition, $\rho^*$ has to minimize the function $\rho \to (\rho - |\sigma|)^2 + (\rho - |\tau|)^2$. The unique minimum is $\rho^* = (|\sigma| + |\tau|)/2$. ∎

**Corollary 3.1.** *The eigenvalues of the normal tridiagonal Toeplitz matrix $T^* = (n; \sigma^*, \delta^*, \tau^*)$ closest to $T = (n; \sigma, \delta, \tau)$ are given by*

$$\lambda_h(T^*) = \delta + (|\sigma| + |\tau|)\; e^{i(\alpha+\beta)/2}\, \cos\frac{h\pi}{n+1}\,, \qquad h = 1:n, \tag{11}$$

*where as usual $\alpha$ and $\beta$ are defined by (3). The eigenvalues lie on the closed line segment*

$$\mathcal{S}_{\lambda(T^*)} = \left\{ \delta + t\, e^{i(\alpha+\beta)/2} : \, t \in \mathbb{R},\, |t| \le (|\sigma| + |\tau|)\, \cos\frac{\pi}{n+1} \right\}.$$

*Since*

$$|\sigma| + |\tau| - 2\sqrt{|\sigma\tau|} = \left(\sqrt{|\sigma|} - \sqrt{|\tau|}\right)^2,$$

*this line segment properly contains the line segment in (6) if and only if $T \notin \mathcal{N}_\mathcal{T}$. Moreover, $T^*$ has the spectral radius*

$$\rho(T^*) = \max\left\{ \left| \delta + (|\sigma| + |\tau|)\; e^{i(\alpha+\beta)/2} \cos\frac{\pi}{n+1} \right|, \left| \delta + (|\sigma| + |\tau|)\; e^{i(\alpha+\beta)/2} \cos\frac{n\,\pi}{n+1} \right| \right\}.$$

The following result provides a simple formula for the distance to normality of a tridiagonal Toeplitz matrix.

**Theorem 3.3.** *Let $T = (n; \sigma, \delta, \tau)$. Then*

$$d_F(T, \mathcal{N}_\mathcal{T}) = \sqrt{\frac{n-1}{2}}\, (\max\{|\sigma|, |\tau|\} - \min\{|\sigma|, |\tau|\}). \tag{12}$$

**Proof:** We obtain from Theorem 3.2 that

$$\begin{aligned} \|T - T^*\|_F^2 &= (n-1)(|\sigma - \sigma^*|^2 + |\tau - \tau^*|^2) \\ &= (n-1)\left(||\sigma| - |\sigma^*||^2 + ||\tau| - |\tau^*||^2\right) \\ &= (n-1)\left(||\sigma| - \rho^*|^2 + ||\tau| - \rho^*|^2\right) \\ &= \frac{n-1}{2}\, ||\sigma| - |\tau||^2. \end{aligned}$$

This proves the assertion. ∎

**Remark 3.1.** *The distance $d_F(T, \mathcal{N}_\mathcal{T})$ is independent of $\delta$, but the closest normal matrix $T^*$ to $T$ depends on $\delta$. In other words, matrices that differ only in $\delta$ have the same distance to the algebraic variety $\mathcal{N}_\mathcal{T}$, but they have different projections onto $\mathcal{N}_\mathcal{T}$. Also note that $T_1 = (n, \sigma, \delta_1, \tau)$ and $T_2 = (n, \sigma, \delta_2, \tau)$ yields*

$$\|T_1^* - T_2^*\|_F = \|T_1 - T_2\|_F = \sqrt{n}\,|\delta_1 - \delta_2|\,.$$

*3.1. The relation between the distance to and departure from normality*

The *departure from normality*

$$\Delta_F(A) = \left( \|A\|_F^2 - \sum_{h=1}^{n} |\lambda_h|^2 \right)^{\frac{1}{2}}, \quad A \in \mathbb{C}^{n\times n},$$

was introduced by Henrici [19] to measure the nonnormality of a matrix. It is easily shown, by using the trigonometric identity

$$\sum_{k=1}^{n} \cos^2\left(\frac{k\pi}{n+1}\right) = \frac{n-1}{2}, \tag{13}$$

that

$$\Delta_F(T_0) = \sqrt{n-1}\,(\max\{|\sigma|, |\tau|\} - \min\{|\sigma|, |\tau|\}).$$

It follows from (12) that $\Delta_F(T_0) = \sqrt{2}\,d_F(T_0, \mathcal{N}_\mathcal{T})$. László [24] has shown that for any $A \in \mathbb{C}^{n\times n}$,

$$\frac{\Delta_F(A)}{\sqrt{n}} \le d_F(A, \mathcal{N}) \le \Delta_F(A),$$

where $d_F(A, \mathcal{N})$ denotes the distance to normality (1). We conclude that

$$\frac{\sqrt{2}}{\sqrt{n}}\, d_F(T_0, \mathcal{N}_\mathcal{T}) \le d_F(T_0, \mathcal{N}) \le \sqrt{2}\, d_F(T_0, \mathcal{N}_\mathcal{T}). \tag{14}$$

*3.2. The distance between the spectra of $T$ and $T^*$*

We are in a position to bound the distance between the spectra of a tridiagonal Toeplitz matrix $T$ and of its closest normal tridiagonal Toeplitz matrix $T^*$.

**Theorem 3.4.** *Let $T^*$ be the closest normal tridiagonal Toeplitz matrix to $T = (n; \sigma, \delta, \tau)$. Define the eigenvalue vectors*

$$\lambda = [\lambda_1(T), \lambda_2(T), \ldots, \lambda_n(T)], \quad \lambda^* = [\lambda_1(T^*), \lambda_2(T^*), \ldots, \lambda_n(T^*)],$$

*where we assume that the eigenvalues of $T$ and $T^*$ are ordered in the same manner. Then*

$$\|\lambda - \lambda^*\|_2 = \sqrt{\frac{n-1}{2}}(\sqrt{|\sigma|} - \sqrt{|\tau|})^2.$$

**Proof:** We obtain from (4) and (11) that

$$|\lambda_h(T) - \lambda_h(T^*)| = \left(\sqrt{|\sigma|} - \sqrt{|\tau|}\right)^2 \left|\cos\frac{h\pi}{n+1}\right|, \qquad h = 1:n.$$

The theorem now follows from (13). ∎

The following result is a consequence of Theorems 3.3 and 3.4, and shows that

$$\lim_{T\to T^*} \frac{\|\lambda-\lambda^*\|_2}{d_F(T,\mathcal{N}_\mathcal{T})} = 0.$$

**Theorem 3.5.** *Let $T \notin \mathcal{N}_\mathcal{T}$. Using the notation of Theorems 3.3 and 3.4, we have*

$$\frac{\|\lambda-\lambda^*\|_2}{d_F(T,\mathcal{N}_\mathcal{T})} = \frac{\left|\sqrt{|\sigma|}-\sqrt{|\tau|}\right|}{\sqrt{|\sigma|}+\sqrt{|\tau|}}\,. \tag{15}$$

**Proof:** It follows from Theorems 3.3 and 3.4 that

$$\frac{\|\lambda-\lambda^*\|_2}{d_F(T,\mathcal{N}_\mathcal{T})} = \frac{\left(\sqrt{|\sigma|}-\sqrt{|\tau|}\right)^2}{||\sigma|-|\tau||} = \frac{\left|\sqrt{|\sigma|}-\sqrt{|\tau|}\right|}{\sqrt{|\sigma|}+\sqrt{|\tau|}}.$$

∎

*3.3. Normalized structured distance to normality*

We first consider the matrices $T_0$ with $(\sigma,\tau) \neq (0,0)$. Theorem 3.3 leads to the following observations:

- When $\sigma\,\tau \neq 0$, we have

  $$\frac{d_F(T_0,\mathcal{N}_\mathcal{T})}{\|T_0\|_F} = \frac{\sqrt{\frac{n-1}{2}}\,||\sigma|-|\tau||}{\sqrt{n-1}\,\sqrt{|\sigma|^2+|\tau|^2}} = \frac{||\sigma/\tau|-1|}{\sqrt{2}\sqrt{|\sigma/\tau|^2+1}} = \frac{||\tau/\sigma|-1|}{\sqrt{2}\sqrt{1+|\tau/\sigma|^2}}\,,$$

  and, therefore,

  $$0 \leq \frac{d_F(T_0,\mathcal{N}_\mathcal{T})}{\|T_0\|_F} < \frac{1}{\sqrt{2}}.$$

  Moreover, the normalized structured distance to normality decreases from $\sqrt{2}/2$ to 0 when one of the two ratios $|\sigma/\tau|$ or $|\tau/\sigma|$ increases from 0 to 1.

- $$\frac{d_F(T_0,\mathcal{N}_\mathcal{T})}{\|T_0\|_F} = 0, \quad \text{if and only if } |\sigma| = |\tau|.$$

- When $\sigma = 0,\ \tau \neq 0$ or $\sigma \neq 0,\ \tau = 0$, we have

  $$\frac{d_F(T_0,\mathcal{N}_\mathcal{T})}{\|T_0\|_F} = \frac{1}{\sqrt{2}}. \tag{16}$$

Remark 3.1 yields that $d_F(T,\mathcal{N}_\mathcal{T}) = d_F(T_0,\mathcal{N}_\mathcal{T})$. Therefore,

$$\begin{aligned} 0 \quad &\leq \quad \frac{d_F(T,\mathcal{N}_\mathcal{T})}{\|T\|_F} = \frac{d_F(T_0,\mathcal{N}_\mathcal{T})}{\|T_0\|_F}\,\frac{\|T_0\|_F}{\|T\|_F} \\ &= \quad \frac{d_F(T_0,\mathcal{N}_\mathcal{T})}{\|T_0\|_F}\sqrt{\frac{(n-1)\,(|\sigma|^2+|\tau|^2)}{(n-1)\,(|\sigma|^2+|\tau|^2)+n|\delta|^2}} \\ &\leq \quad \frac{d_F(T_0,\mathcal{N}_\mathcal{T})}{\|T_0\|_F} \leq \frac{1}{\sqrt{2}}. \end{aligned}$$

The upper bound is achieved if and only if $\delta = 0$ and $T$ is bidiagonal.

*3.4. Normalized departure and distance from normality*

It is straightforward to show that the upper bound for the normalized departure from normality of the matrix $T_0$ is one, and that the upper bound for the normalized distance to normality is $1/\sqrt{2}$. Moreover, the following result holds.

**Theorem 3.6.** *Let $T_0 = (n; \sigma, 0, \tau)$ with $\sigma = 0,\ \tau \neq 0$, or $\sigma \neq 0,\ \tau = 0$. Then*

$$\frac{d_F(T_0, \mathcal{N})}{\|T_0\|_F} = \frac{1}{\sqrt{n}}.$$

**Proof:** The inequality $d_F(T_0)/\|T_0\|_F \geq 1/\sqrt{n}$ follows from (14) and (16). To show equality, we construct a normal (circulant) matrix $N$ at normalized distance $1/\sqrt{n}$ from $T_0$. Specifically, if $\sigma \neq 0$ and $\tau = 0$, then we let

$$N = \frac{n-1}{n}(T_0 + \sigma e_1 e_n^T),$$

where $e_j$ denotes the $j$th axis vector, and if $\sigma = 0$ and $\tau \neq 0$, then we choose

$$N = \frac{n-1}{n}(T_0 + \tau e_n e_1^T).$$

■

## 4. Distance and structured distance to $\mathcal{M}_\mathcal{T}$

The matrices in $\mathcal{M}_\mathcal{T}$ are multiples of the identity, which are normal matrices, or bidiagonal matrices, which have the unique eigenvalue $\delta$ with geometric multiplicity 1. This observation leads to the following result.

**Theorem 4.1.** *Let $T = (n; \sigma, \delta, \tau)$. If $|\sigma| = \min\{|\sigma|, |\tau|\}$ (or $|\tau| = \min\{|\sigma|, |\tau|\}$), then $T^+ = (n; 0, \delta, \tau)$ (or $T^+ = (n; \sigma, \delta, 0)$) is the closest matrix to $T$ in $\mathcal{M}_\mathcal{T}$, when the distance is measured in the Frobenius norm.*

**Corollary 4.1.** *For any $T \in \mathcal{T}$, we have*

$$d_F(T, \mathcal{M}_\mathcal{T}) = \sqrt{n-1}\,\min\{|\sigma|, |\tau|\}.$$

*In particular, if $T \in \mathcal{N}_\mathcal{T}$, then*

$$d_F(T, \mathcal{M}_\mathcal{T}) = \sqrt{n-1}\,|\sigma| = \sqrt{n-1}\,|\tau|.$$

*Further, if $T \notin \mathcal{N}_\mathcal{T}$, then*

$$d_F(T^*, \mathcal{M}_\mathcal{T}) = \sqrt{n-1}\,\frac{|\sigma| + |\tau|}{2},$$

*where $T^*$ denotes the closest matrix to $T$ in $\mathcal{N}_\mathcal{T}$ in the Frobenius norm.*

We remark that for any $T \in \mathcal{T}$, it holds

$$\begin{aligned} d_F(T^*, \mathcal{M}_\mathcal{T}) - d_F(T, \mathcal{M}_\mathcal{T}) &= \sqrt{n-1}\left(\frac{|\sigma| + |\tau|}{2} - \min\{|\sigma|, |\tau|\}\right) \\ &= \sqrt{n-1}\,\frac{\max\{|\sigma|, |\tau|\} - \min\{|\sigma|, |\tau|\}}{2} \\ &= \frac{1}{\sqrt{2}} d_F(T, \mathcal{N}_\mathcal{T}). \end{aligned}$$

This shows that the larger the difference between $|\sigma|$ and $|\tau|$ is, the larger is the difference in the structured distances of $T$ and $T^*$ from $\mathcal{M}_{\mathcal{T}}$.

Introduce the ratio

$$r = \frac{\min\{|\sigma|, |\tau|\}}{\max\{|\sigma|, |\tau|\}}. \tag{17}$$

This ratio is used in the proof of the following theorem, which provides a bound for the normalized structured distance of $T_0$ to $\mathcal{M}_{\mathcal{T}}$.

**Theorem 4.2.**

$$\frac{d_F(T_0, \mathcal{M}_{\mathcal{T}})}{\|T_0\|_F} \leq \frac{1}{\sqrt{2}}.$$

*The upper bound is achieved if and only if $T_0$ is normal.*

**Proof:** Assume that $\min\{|\sigma|, |\tau|\} = |\sigma|$. Then

$$\frac{d_F(T_0, \mathcal{M}_{\mathcal{T}})}{\|T_0\|_F} = \frac{|\sigma|}{\sqrt{|\sigma|^2 + |\tau|^2}} = \frac{1}{\sqrt{1 + |\tau/\sigma|^2}} \leq \frac{1}{\sqrt{2}},$$

and it follows that the normalized structured distance decreases from $1/\sqrt{2}$ to 0 when the ratio (17) decreases from 1 to 0. The proof is analogous when $\min\{|\sigma|, |\tau|\} = |\tau|$. ∎

We conclude this section with a few observations:

$$\begin{aligned}
\frac{d_F(T_0, \mathcal{M}_{\mathcal{T}})}{\|T_0\|_F} &= \frac{1}{\sqrt{2}} && \text{if and only if } |\sigma| = |\tau|, \\
\lim_{|\sigma|\to 0} \frac{d_F(T_0, \mathcal{M}_{\mathcal{T}})}{\|T_0\|_F} &= 0 && \text{for } \tau \neq 0, \\
\lim_{|\tau|\to 0} \frac{d_F(T_0, \mathcal{M}_{\mathcal{T}})}{\|T_0\|_F} &= 0 && \text{for } \sigma \neq 0.
\end{aligned}$$

## 5. Eigenvalue sensitivity

We investigate the sensitivity of the eigenvalues of the matrices $T_0$ and $T$ in several ways, and begin by studying the sensitivity of the vector

$$\lambda(T_0) = [\lambda_1(T_0), \lambda_2(T_0), \ldots, \lambda_n(T_0)]$$

to perturbations in $\sigma$ and $\tau$. To this end, introduce the function

$$f : D \subset \mathbb{C}^2 \to f(D) \subset \mathbb{C}^n, \quad D = \{(\sigma, \tau) \in \mathbb{C}^2 : \sigma\tau \neq 0\} : \quad \lambda(T_0) = f(\sigma, \tau).$$

The sensitivity of $\lambda(T_0)$ to perturbations in $\sigma$ and $\tau$ is determined by the Jacobian of $f$. Using (4), we obtain the representation

$$J_f(\sigma, \tau) = \begin{bmatrix} \sqrt{\frac{\tau}{\sigma}} \cos\frac{\pi}{n+1} & \sqrt{\frac{\sigma}{\tau}} \cos\frac{\pi}{n+1} \\ \sqrt{\frac{\tau}{\sigma}} \cos\frac{2\pi}{n+1} & \sqrt{\frac{\sigma}{\tau}} \cos\frac{2\pi}{n+1} \\ \cdot & \cdot \\ \cdot & \cdot \\ \sqrt{\frac{\tau}{\sigma}} \cos\frac{n\pi}{n+1} & \sqrt{\frac{\sigma}{\tau}} \cos\frac{n\pi}{n+1} \end{bmatrix} \in \mathbb{C}^{n\times 2} \tag{18}$$

of the Jacobian matrix. Application of (13) yields

$$\|J_f(\sigma,\tau)\|_F = \sqrt{\frac{n-1}{2}}\sqrt{\left|\frac{\sigma}{\tau}\right| + \left|\frac{\tau}{\sigma}\right|} = \sqrt{\frac{n-1}{2}}\sqrt{\frac{|\sigma|^2+|\tau|^2}{|\sigma||\tau|}}. \tag{19}$$

If we instead consider relative errors in the data $\sigma, \tau$ and in $\lambda_h(T_0)$, then the analogue of (18) is the $n \times 2$ matrix

$$\Gamma_f(\sigma,\tau) = \begin{bmatrix} \frac{\sigma}{\lambda_1(T_0)}\left(J_f(\sigma,\tau)\right)_{1,1} & \frac{\tau}{\lambda_1(T_0)}\left(J_f(\sigma,\tau)\right)_{1,2} \\ \frac{\sigma}{\lambda_2(T_0)}\left(J_f(\sigma,\tau)\right)_{2,1} & \frac{\tau}{\lambda_2(T_0)}\left(J_f(\sigma,\tau)\right)_{2,2} \\ \cdot & \cdot \\ \cdot & \cdot \\ \frac{\sigma}{\lambda_n(T_0)}\left(J_f(\sigma,\tau)\right)_{n,1} & \frac{\tau}{\lambda n(T_0)}\left(J_f(\sigma,\tau)\right)_{n,2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \cdot & \cdot \\ \cdot & \cdot \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

We obtain

$$\Gamma_f(\sigma,\tau)^H \Gamma_f(\sigma,\tau) = \frac{n}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

and

$$\|\Gamma_f(\sigma,\tau)\|_2 = \|\Gamma_f(\sigma,\tau)\|_F = \sqrt{\frac{n}{2}}.$$

**Remark 5.1.** *The norm of $\Gamma_f$ is independent of $\sigma$ and $\tau$, but the norm of $J_f$ depends on the ratio $|\sigma/\tau|$. The norm of $J_f$ achieves its minimum, $\sqrt{n-1}$, if and only if $|\sigma| = |\tau|$, i.e., if and only if $T$ is normal. The norm of $J_f$ tends to $+\infty$ when the ratio (17) decreases.*

**Remark 5.2.** *The sensitivity of the eigenvalue $\lambda_h(T_0)$ to perturbations increases with its magnitude.*

**Theorem 5.1.** *Let $\sigma\tau \neq 0$. Then*

$$\|J_f(\sigma,\tau)\|_F = \sqrt{\frac{n-1}{1-2\frac{d_F(T_0,\mathcal{N}_\mathcal{T})^2}{\|T_0\|_F^2}}}.$$

**Proof:** If $\sigma\tau \neq 0$, then

$$\frac{d_F(T_0,\mathcal{N}_\mathcal{T})^2}{\|T_0\|_F^2} = \frac{\frac{n-1}{2}(|\sigma|^2+|\tau|^2-2|\sigma||\tau|)}{\|T_0\|_F^2} = \frac{1}{2}\left(1 - \frac{2|\sigma||\tau|}{|\sigma|^2+|\tau|^2}\right).$$

The last equality in (19) now gives

$$\frac{d_F(T_0,\mathcal{N}_\mathcal{T})^2}{\|T_0\|_F^2} = \frac{1}{2}\left(1 - \frac{n-1}{\|J_f(\sigma,\tau)\|_F^2}\right),$$

and the desired result follows. ∎

*5.1. Individual eigenvalue condition numbers*

Condition numbers for individual eigenvalues are discussed, e.g., in [16, 42, 43]. When $\sigma\tau \neq 0$, these condition numbers can be obtained from (7) and (8). Standard computations and the trigonometric identity

$$\sum_{k=1}^{n} \sin^2\left(\frac{hk\pi}{n+1}\right) = \frac{n+1}{2}, \quad h = 1:n,$$

yield, for $h = 1:n$,

$$\begin{aligned} \|x_h\|_2^2 &= \sum_{k=1}^{n} \left|\frac{\sigma}{\tau}\right|^k \sin^2\left(\frac{hk\pi}{n+1}\right), \\ \|y_h\|_2^2 &= \sum_{k=1}^{n} \left|\frac{\tau}{\sigma}\right|^k \sin^2\left(\frac{hk\pi}{n+1}\right), \\ |y_h^H x_h| &= \sum_{k=1}^{n} \sin^2\left(\frac{hk\pi}{n+1}\right) = \frac{n+1}{2}. \end{aligned}$$

Consequently, the individual condition numbers are, for $h = 1:n$, given by

$$\begin{aligned} \kappa(\lambda_h(T)) &= \frac{\|x_h\|_2 \|y_h\|_2}{|y_h^H x_h|} \\ &= \frac{2}{n+1} \sqrt{\sum_{k=1}^{n} \left|\frac{\sigma}{\tau}\right|^k \sin^2\left(\frac{hk\pi}{n+1}\right) \cdot \sum_{k=1}^{n} \left|\frac{\tau}{\sigma}\right|^k \sin^2\left(\frac{hk\pi}{n+1}\right)}. \end{aligned} \tag{20}$$

In the special case when $|\sigma| = |\tau|$, the matrix $T$ is normal, cf. Theorem 3.1, and

$$\|x_h\|_2^2 = \|y_h\|_2^2 = \sum_{k=1}^{n} \sin^2\left(\frac{hk\,\pi}{n+1}\right) = \frac{n+1}{2}, \quad h = 1:n.$$

It follows that

$$\kappa(\lambda_h(T)) = \frac{\|x_h\|_2 \, \|y_h\|_2}{|y_h^H x_h|} = 1.$$

In the general case when $|\sigma| \neq |\tau|$, we obtain from (20) the expressions

$$\kappa(\lambda_h(T)) = \frac{1 - r^{n+1}}{r^{n/2}(n+1)} \sqrt{S_{n,r}(h) S_{n,1/r}(h)}, \quad h = 1:n,$$

where $r$ is defined by (17) and

$$\begin{aligned} S_{n,r}(h) &= \frac{1}{1-r} - \frac{1 - r\cos\frac{2h\pi}{n+1}}{(1 - r\cos\frac{2h\pi}{n+1})^2 + r^2 \sin^2\left(\frac{2h\pi}{n+1}\right)}, \\ S_{n,1/r}(h) &= \frac{1}{1-r} - \frac{\cos\frac{2nh\pi}{n+1} - r}{(\cos\frac{2nh\pi}{n+1} - r)^2 + \sin^2\left(\frac{2nh\pi}{n+1}\right)}. \end{aligned}$$

A straightforward computation yields

$$\kappa(\lambda_h(T)) = \frac{(1-r^{n+1})(1+r)(1-\cos\frac{2h\pi}{n+1})}{r^{(n-1)/2}(n+1)(1-r)(1+r^2-2r\cos\frac{2h\pi}{n+1})}, \quad h = 1:n, \tag{21}$$

where the factor that depends on $h$ satisfies the bounds

$$\frac{1}{2} \leq \frac{1-\cos\frac{2h\pi}{n+1}}{1+r^2-2r\cos\frac{2h\pi}{n+1}} \leq 2. \tag{22}$$

This factor is the largest for $h = \lfloor n/2 \rfloor$, where $\lfloor t \rfloor$ denotes the largest integer smaller than or equal to $t$. It follows that the eigenvalues in the middle of the spectrum are the worst conditioned. Moreover, for $0 < r < 1$, $\kappa(\lambda_h(T))$ grows exponentially with $n$. Further, $\kappa(\lambda_h(T)) \to 1$ as $r \to 1$, and $\kappa(\lambda_h(T)) \to \infty$ as $r \to 0$. In the latter case, we have the estimates

$$\kappa(\lambda_h(T)) \approx \frac{1-\cos\frac{2h\pi}{n+1}}{n+1}\left(\frac{1}{r}\right)^{\frac{n-1}{2}}, \quad h = 1:n.$$

*5.2. The global eigenvalue condition number*

Properties of the global condition number

$$\kappa_F(\lambda) = \sum_{h=1}^{n} \kappa(\lambda_h(T))$$

are discussed by Stewart and Sun [39]. It can be evaluated by summing the individual condition numbers. We would like to determine a simple explicit approximation that provides insight into the conditioning. Using (21) and (22), we obtain for any diagonalizable matrix $T = (n; \sigma, \delta, \tau)$ with $|\sigma| \neq |\tau|$ the bounds

$$\frac{K_{n,r}}{2} \leq \kappa_F(\lambda) \leq 2K_{n,r},$$

where

$$K_{n,r} = \frac{1}{r^{(n-1)/2}}\frac{1-r^{n+1}}{1-r}(1+r)\frac{n}{n+1}, \quad 0 < r < 1, \tag{23}$$

and $r$ is given by (17).

*5.3. The $\varepsilon$-pseudospectrum*

For a given $\varepsilon > 0$, the $\varepsilon$-pseudospectrum of $A \in \mathbb{C}^{n\times n}$ is the set

$$\Lambda_\varepsilon(A) = \left\{z : \left\|(zI-A)^{-1}\right\|_2 \geq \varepsilon^{-1}\right\};$$

see, e.g., Trefethen and Embree [40]. The following alternative definition will be used in Section 7:

$$\Lambda_\varepsilon(A) = \left\{z : \exists\, u \in \mathbb{C}^n, \|u\|_2 = 1, \text{ such that } \|(zI-A)u\|_2 \leq \varepsilon\right\}. \tag{24}$$

The vectors $u$ in the above definition are referred to as $\varepsilon$-pseudoeigenvectors.

The $\varepsilon$-pseudospectrum $\Lambda_\varepsilon(T)$ of $T = (n; \sigma, \delta, \tau)$ approximates the spectrum of the Toeplitz operator $T_\infty = (\infty; \sigma, \delta, \tau)$ as $\varepsilon \searrow 0$ and $n \to \infty$; see [34, 40]. Introduce the *symbol* of the matrix $T$,

$$f(z) = \tau z + \delta + \sigma z^{-1}.$$

Then the ellipse

$$f(S) = \{f(z) : z \in \mathbb{C}, |z| = 1\} \tag{25}$$

is the boundary of the spectrum of $T_\infty$. The major axis of $f(S)$ is

$$\mathcal{S}_{\text{major\_axis}} = \left\{\delta + t\, e^{i(\alpha+\beta)/2},\ t \in \mathbb{R},\ |t| \le |\sigma| + |\tau|\right\} \tag{26}$$

and the interval between the foci of $f(S)$ is given by

$$\mathcal{S}_{\text{foci}} = \left\{\delta + t\, e^{i(\alpha+\beta)/2},\ t \in \mathbb{R},\ |t| \le 2\sqrt{|\sigma\tau|}\right\}. \tag{27}$$

According to (6), the spectrum $T = (n; \sigma, \delta, \tau)$ lives in the interval $\mathcal{S}_{\text{foci}}$ for every finite $n \ge 1$ and there is no shorter interval with this property. Moreover, by (11), the spectrum of the normal matrix $T^*$ closest to $T$ lives in the interval (26).

### *5.4. Structured perturbations*

Let $|\sigma| = \min\{|\sigma|, |\tau|\}$ and consider the tridiagonal perturbation $E_s = (n; -s, 0, 0)$ of the matrix $T = (n; \sigma, \delta, \tau)$. For $s = \upsilon\sigma$ with $0 < \upsilon < 1$, we obtain a family of diagonalizable matrices $T + E_s$ with simple eigenvalues. The matrices $T + E_s$ converge to the defective matrix $T^+ = (n; 0, \delta, \tau)$ when $\upsilon \nearrow 1$. The latter matrix has the unique eigenvalue $\delta$ of geometric multiplicity one. Thus, the structured perturbation

$$E_\sigma = (n; -\sigma, 0, 0), \quad \|E_\sigma\|_F = \sqrt{n-1}|\sigma|,$$

moves all the eigenvalues to $\delta$. The rate of change for the $h$th eigenvalue of $T$ is, for $0 < |\sigma| \le |\tau|$, given by

$$\frac{|\lambda_h(T + E_\sigma) - \lambda_h(T)|}{\|E_\sigma\|_F} = \frac{2\sqrt{|\sigma\tau|}\left|\cos\frac{h\pi}{n+1}\right|}{\sqrt{n-1}|\sigma|} = \frac{2}{\sqrt{(n-1)r}}\left|\cos\frac{h\pi}{n+1}\right| \tag{28}$$

with $r$ defined by (17). The closer $r$ is to unity, the smaller is the rate of change (28) of the eigenvalues. This rate is minimal when $r = 1$ and $T$ is normal.

Analogously, let $E_{s,t} = (n; -s, 0, -t)$ with $s = \upsilon\sigma$ and $t = \upsilon\tau$ for $0 < \upsilon < 1$. Then

$$\lim_{\nu \to 1}(T + E_{s,t}) = \delta I,$$

where $I$ denotes the identity matrix. Thus, the limit matrix is normal. The structured perturbation

$$E_{\sigma,\tau} = (n; -\sigma, 0, -\tau), \quad \|E_{\sigma,\tau}\|_F = \sqrt{n-1}\,\sqrt{|\sigma|^2 + |\tau|^2},$$

gives the limit matrix. The rate of change of the eigenvalue under this perturbation is given by

$$\frac{|\lambda_h(T + E_{\sigma,\tau}) - \lambda_h(T)|}{\|E_{\sigma,\tau}\|_F} = \frac{2\sqrt{|\sigma\tau|}\left|\cos\frac{h\pi}{n+1}\right|}{\sqrt{n-1}\,\sqrt{|\sigma|^2 + |\tau|^2}} = \frac{\sqrt{2}}{\|J_f(\sigma,\tau)\|_F}\left|\cos\frac{h\pi}{n+1}\right|.$$

Thus, the rate is inversely proportional to the norm of the Jacobian matrix (18); cf. (19). The rate is the largest when $T$ is normal; see Remark 5.1. Also note that the further the eigenvalues of $T$ are from $\delta$, the higher is their sensitivity to the structured perturbation; cf. Remark 5.2.

| $\lambda$ | $\kappa(\lambda(T))$ | $\kappa_{\mathcal{T}}(\lambda(T))$ |
|---|---|---|
| $\lambda_1$ | $7.0463 \cdot 10^4$ | $8.7215 \cdot 10^{-1}$ |
| $\lambda_2$ | $2.5759 \cdot 10^5$ | $8.2610 \cdot 10^{-1}$ |
| $\lambda_3$ | $5.0517 \cdot 10^5$ | $7.5194 \cdot 10^{-1}$ |
| $\lambda_4$ | $7.5633 \cdot 10^5$ | $6.5374 \cdot 10^{-1}$ |
| $\lambda_5$ | $9.7209 \cdot 10^5$ | $5.3790 \cdot 10^{-1}$ |
| $\lambda_6$ | $1.1325 \cdot 10^6$ | $4.1511 \cdot 10^{-1}$ |
| $\lambda_7$ | $1.2300 \cdot 10^6$ | $3.0680 \cdot 10^{-1}$ |
| $\lambda_8$ | $1.2626 \cdot 10^6$ | $2.5820 \cdot 10^{-1}$ |
| $\lambda_9$ | $1.2300 \cdot 10^6$ | $3.0680 \cdot 10^{-1}$ |
| $\lambda_{10}$ | $1.1325 \cdot 10^6$ | $4.1511 \cdot 10^{-1}$ |
| $\lambda_{11}$ | $9.7209 \cdot 10^5$ | $5.3790 \cdot 10^{-1}$ |
| $\lambda_{12}$ | $7.5633 \cdot 10^5$ | $6.5374 \cdot 10^{-1}$ |
| $\lambda_{13}$ | $5.0517 \cdot 10^5$ | $7.5194 \cdot 10^{-1}$ |
| $\lambda_{14}$ | $2.5759 \cdot 10^5$ | $8.2610 \cdot 10^{-1}$ |
| $\lambda_{15}$ | $7.0463 \cdot 10^4$ | $8.7215 \cdot 10^{-1}$ |

Table II. Traditional and structured individual eigenvalue condition numbers, $\kappa(\lambda_h(T))$ and $\kappa_{\mathcal{T}}(\lambda_h(T))$, respectively, for the matrix $T = (15; -i, 11 - 2i, 6 + 8i)$.

In order to be able to discuss the sensitivity of the eigenvalues to structured perturbations, we introduce the right and left eigenvectors of unit length,

$$\widetilde{x}_h = \frac{x_h}{\|x_h\|}, \quad \widetilde{y}_h = \frac{y_h}{\|y_h\|}, \quad h = 1 : n,$$

where $x_h$ and $y_h$ are defined by (7) and (8), respectively. The smaller $|\sigma/\tau| < 1$ is, the larger is the first component of $\widetilde{x}_h$ and the last component of $\widetilde{y}_h$. Similarly, the larger $|\sigma/\tau| > 1$ is, the larger is the last component of $\widetilde{x}_h$ and the first component of $\widetilde{y}_h$.

Consider the Wilkinson perturbation,

$$W_h = \widetilde{y}_h \widetilde{x}_h^H,$$

associated with $\lambda_h$. This is a unit-norm perturbation of $T$ that yields the largest perturbation in $\lambda_h$; see, e.g., [43]. The entries of largest magnitude of $W_h$ are in the bottom-left corner when $|\sigma/\tau| < 1$ and in the top-right corner when $|\sigma/\tau| > 1$. In particular, the largest entries are not in $W_h|_{\mathcal{T}}$, the orthogonal projection of $W_h$ in the subspace $\mathcal{T}$ of tridiagonal Toeplitz matrices. The (tridiagonal Toeplitz) structured condition number of the eigenvalue $\lambda_h$ of the tridiagonal Toeplitz matrix $T$ is given by

$$\kappa_{\mathcal{T}}(\lambda_h(T)) = \kappa(\lambda_h(T))\|W_h|_{\mathcal{T}}\|_F;$$

see [23, 28, 29]. It follows that a large (traditional) condition number $\kappa(\lambda_h(T))$ does not imply that the structured condition number is large. Thus, an eigenvalue $\lambda_h(T)$ may be much more sensitive to a general perturbation of $T$ than to a structured perturbation. This is illustrated in the following example.

Example 5.1. Let $T = (15; \sigma, \delta, \tau)$ for $\sigma = -i$, $\delta = 11 - 2i$, and $\tau = 6 + 8i$. The ratio (17) for this matrix is $r = 1/10$. Table II shows traditional and structured individual eigenvalue

| $r$ | $d_F(T_{(r)}, \mathcal{N}_\mathcal{T})$ | $K_{50,r}$ | $\|\lambda(T_{(r)}) - \lambda(T^*_{(r)})\|_2$ |
|---|---|---|---|
| 0.1 | $2.23 \cdot 10^1$ | $3.79 \cdot 10^{24}$ | $1.16 \cdot 10^1$ |
| 0.3 | $1.73 \cdot 10^1$ | $1.18 \cdot 10^{13}$ | $5.06 \cdot 10^0$ |
| 0.5 | $1.24 \cdot 10^1$ | $6.98 \cdot 10^7$ | $2.12 \cdot 10^0$ |
| 0.9 | $2.47 \cdot 10^0$ | $2.45 \cdot 10^2$ | $6.52 \cdot 10^{-2}$ |

Table III. Quantities related to the matrices $T_{(r)}$ defined by (29) and the closest normal matrices $T^*_{(r)}$.

condition numbers, $\kappa(\lambda_h(T))$ and $\kappa_\mathcal{T}(\lambda_h(T))$, respectively, for all eigenvalues. These condition numbers are independent of $\delta$, as well as of $\sigma$ and $\tau$ that correspond to the same ratio $r$. The structured condition numbers are seen to be much smaller than the traditional ones. □

## 6. Illustrations of eigenvalue sensitivity

This section presents computations that illustrate properties of tridiagonal Toeplitz matrices and their eigenvalues discussed in the previous sections. All computations shown in this paper were carried out in MATLAB with about 16 significant decimal digits.

Table III displays quantities associated with matrices of the form

$$T_{(r)} = (50; (4+3i)r, 16-3i, -5) \tag{29}$$

for several values of the parameter $0 < r < 1$, which is the ratio (17). Note that $T_{(0)}$ is defective and $T_{(1)}$ is normal. The latter property follows from the fact that $|4+3i| = |-5|$; cf. Theorem 3.1. The distance $d_F(T_{(r)}, \mathcal{N}_\mathcal{T})$ is computed using (12). The quantity $K_{50,r}$, defined by (23), is an indicator of the sensitivity of the eigenvalues. We use the formula (15) to measure the distance between the spectra of $T_{(r)}$ and of the closest normal matrix $T^*_{(r)}$, i.e.,

$$\|\lambda(T_{(r)}) - \lambda(T^*_{(r)})\|_2 = \frac{1-\sqrt{r}}{1+\sqrt{r}} d_F(T, \mathcal{N}_\mathcal{T}).$$

Figures 1-4 show the eigenvalues of the matrices $T_{(r)}$ and $T^*_{(r)}$ considered in Table III. The eigenvalues are computed with the formulas (4) and (11). The figures also display the image of the unit circle under the symbol for the matrices $T_{(r)}$; see (25). These images are ellipses, each of which is the boundary of the spectrum of the Toeplitz operators $T_\infty = (\infty; (4+3i)r, 16-3i, -5)$.

If, instead of using formula (4), the eigenvalues of $T_{(0.1)}$ were computed with the QR algorithm, then Figure 1 would look quite different. This is illustrated by Figure 5, which displays the computed spectra of the matrices $T^T_{(0.1)}$ and $(T^T_{(0.1)})^*$ using the QR algorithm as implemented by the MATLAB function eig. The fact that the matrices $T_{(0.1)}$ and $T^T_{(0.1)}$ have the same eigenvalues is not apparent from Figures 1 and 5. Indeed the spectrum of the matrix $T^T_{(0.1)}$ in Figure 5 is close to the boundary of the $\varepsilon$-pseudospectrum for $\varepsilon$ equal to machine epsilon $2 \cdot 10^{-16}$.

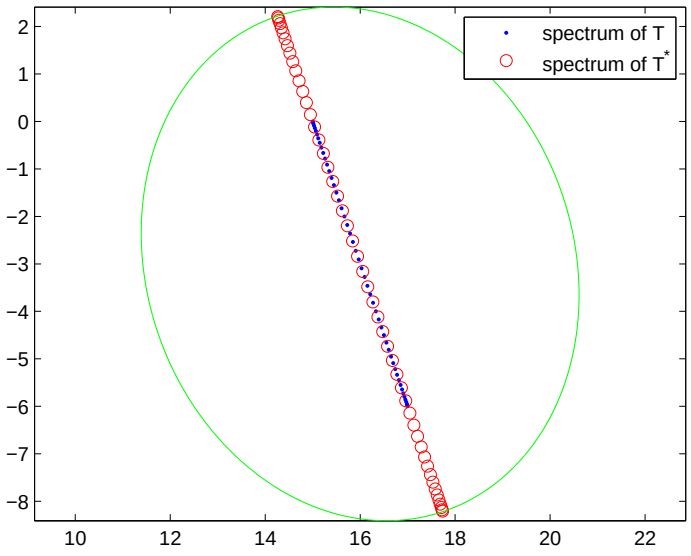



Figure 1. Spectra of the matrix $T_{(r)}$ and of the closest normal tridiagonal matrix $T^*_{(r)}$, as well as the image of the unit circle under the symbol for $T_{(r)}$ for $r = 0.1$. The horizontal axis shows the real part and the vertical axis the imaginary part of the eigenvalues.

## 7. Inverse problems for tridiagonal Toeplitz matrices

This section first discusses an inverse eigenvalue problem for tridiagonal Toeplitz matrices, and then considers an inverse vector problem for tridiagonal Toeplitz matrices. The latter problem determines a trapezoidal tridiagonal Toeplitz matrix by minimizing the norm of the matrix-vector product with a given vector. The solution of this problem finds application to Tikhonov regularization. Details about this application are discussed in Section 8.

**Inverse problem 1**: *Given two distinct complex numbers $a$ and $b$, and a natural number $n$, determine a tridiagonal Toeplitz matrix $T = (n; \sigma, \delta, \tau)$ with extreme eigenvalues $a$ and $b$.* Results of Sections 2-4 shed light on this problem. We note that the problem does not have a unique solution. However, all eigenvalues of $T$ are uniquely determined by the data. The following discussion shows how constraints can be added to achieve unicity. It follows from

$$\lambda_1 = a = \delta + 2\sqrt{\sigma\tau}\cos\frac{\pi}{n+1}, \quad \lambda_n = b = \delta + 2\sqrt{\sigma\tau}\cos\frac{n\pi}{n+1},$$

that the diagonal entry $\delta$ and the product of the sub- and super-diagonal entries, $\sigma\tau$, are uniquely determined by

$$\sqrt{\sigma\tau} = \frac{a-b}{2(\cos\frac{\pi}{n+1} - \cos\frac{n\pi}{n+1})}, \quad \delta = \frac{b\cos\frac{\pi}{n+1} - a\cos\frac{n\pi}{n+1}}{\cos\frac{\pi}{n+1} - \cos\frac{n\pi}{n+1}}.$$

Thus, the absolute value $|\sigma\tau|$ and the angle $\arg(\sigma)+\arg(\tau)$ are determined by the data. We may arbitrarily choose the angle of the sub- or super-diagonal entries as well as the ratio $0 < r \leq 1$

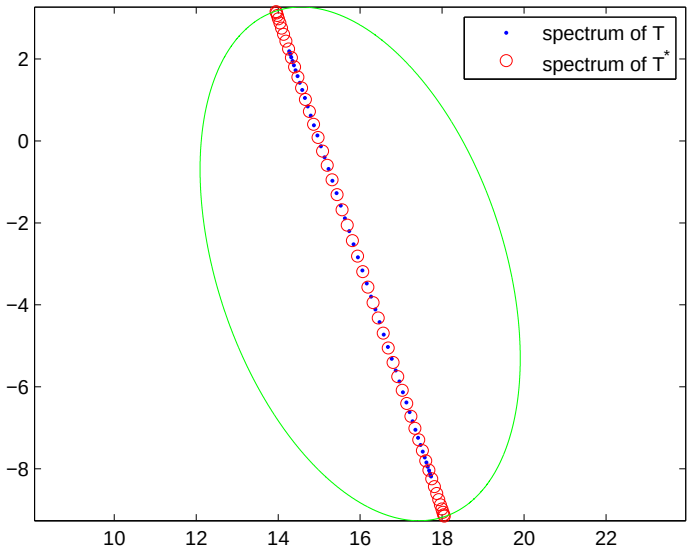



Figure 2. Spectra of the matrix $T_{(r)}$ and of the closest normal tridiagonal matrix $T^*_{(r)}$, as well as the image of the unit circle under the symbol for $T_{(r)}$ for $r = 0.3$. The horizontal axis shows the real part and the vertical axis the imaginary part of the eigenvalues.

defined by (17). The closer $r$ is to zero, the more the ill-conditioned are the eigenvalues. The choice $r = 1$, i.e., $|\sigma| = |\tau|$, yields a normal matrix. Since we may choose the angle of the sub- or super-diagonal entries, the normal matrix is not unique. Unicity can be achieved, e.g., by also prescribing $\arg(\sigma)$ or $\arg(\tau)$.

**Inverse problem 2:** *Given a vector $x \in \mathbb{C}^n$, determine an upper trapezoidal Toeplitz matrix $T \in \mathbb{C}^{(n-2)\times n}$ with first row $[\sigma, 1, \tau, 0, \ldots, 0]$ such that $T$ solves*

$$\min_{\sigma,\tau} \|Tx\|_2. \tag{30}$$

Let $x = [\xi_1, \xi_2, \ldots, \xi_n]^T$. Then the minimization problem (30) can be expressed as

$$\min_{\sigma,\tau} \left\| \begin{bmatrix} \xi_1 & \xi_3 \\ \xi_2 & \xi_4 \\ \cdot & \cdot \\ \cdot & \cdot \\ \xi_{n-2} & \xi_n \end{bmatrix} \begin{bmatrix} \sigma \\ \tau \end{bmatrix} + \begin{bmatrix} \xi_2 \\ \xi_3 \\ \cdot \\ \cdot \\ \xi_{n-1} \end{bmatrix} \right\|_2 . \tag{31}$$

This least-squares problem has a unique solution unless the matrix has linearly dependent columns. The columns are linearly dependent if and only if the components of $x$ satisfy

$$\xi_{k+2} = \alpha \xi_k, \quad k = 1 : n-2,$$

for some $\alpha \in \mathbb{C}$. In this case, we determine the unique solution of minimal Euclidean norm. Note that when

$$\xi_{k+1} = \alpha \xi_k, \quad k = 1 : n-1,$$

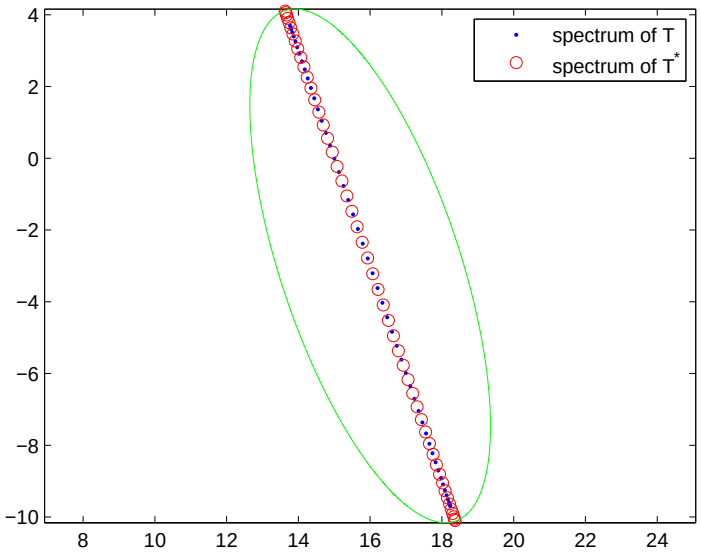



Figure 3. Spectra of the matrix $T_{(r)}$ and of the closest normal tridiagonal matrix $T^*_{(r)}$, as well as the image of the unit circle under the symbol for $T_{(r)}$ for $r = 0.5$. The horizontal axis shows the real part and the vertical axis the imaginary part of the eigenvalues.

for some $\alpha \in \mathbb{C}$, the least-squares problem (31) is consistent.

Having determined the solution $T$ of (30), it is interesting to investigate for which unit vectors $x$ the norm $\|Tx\|_2$ is small. Let $\hat{T} \in \mathbb{C}^{n\times n}$ denote the tridiagonal Toeplitz matrix obtained by prepending and appending suitable rows to T. It follows from definition (24) that the $\varepsilon$-pseudoeigenvectors of $\hat{T}$ associated with $z = 0$ form a subset of

$$\{u : \|Tu\|_2 \leq \varepsilon, \|u\|_2 = 1\}.$$

If zero is in the $\varepsilon$-pseudospectrum of $\hat{T}$, then the corresponding $\varepsilon$-pseudoeigenvectors will be essentially undamped in the Tikhonov regularization method below.

## 8. Tikhonov regularization

This section considers the computation of an approximate solution of the minimization problem

$$\min_{x\in\mathbb{C}^n} \|Ax - b\|_2, \tag{32}$$

where $A \in \mathbb{C}^{m\times n}$ is a matrix with many singular values of different orders of magnitude close to the origin. Minimization problems (32) with a matrix of this kind are commonly referred to as discrete ill-posed problems. They arise, for example, from the discretization of linear ill-posed problems, such as Fredholm integral equations of the first kind. The vector $b \in \mathbb{C}^m$ in (32) represents error-contaminated data. We will for notational simplicity assume that $m \geq n$.

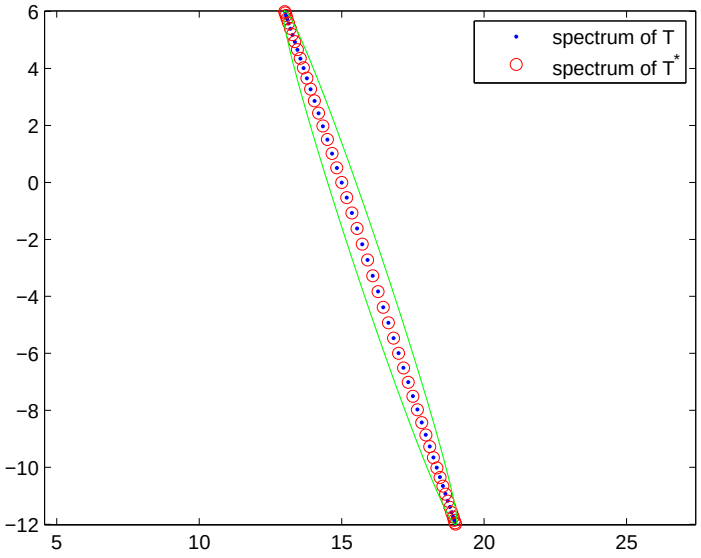



Figure 4. Spectra of the matrix $T_{(r)}$ and of the closest normal tridiagonal matrix $T^*_{(r)}$, as well as the image of the unit circle under the symbol for $T_{(r)}$ for $r = 0.9$. The horizontal axis shows the real part and the vertical axis the imaginary part of the eigenvalues.

Let $e \in \mathbb{C}^m$ denote the (unknown) error in $b$, and let $\hat{b} \in \mathbb{C}^m$ be the error-free vector associated with $b$, i.e.,

$$b = \hat{b} + e.$$

The unavailable linear system of equations with error-free right-hand side,

$$Ax = \hat{b}, \tag{33}$$

is assumed to be consistent. Let $A^\dagger$ denote the Moore-Penrose pseudoinverse of $A$. We are interested in computing an approximation of the solution $\hat{x} = A^\dagger \hat{b}$ of minimal Euclidean norm of the unavailable linear system (33) by determining an approximate solution of the available least-squares problem (32). Note that the solution of (32),

$$\breve{x} = A^\dagger b = A^\dagger(\hat{b} + e) = \hat{x} + A^\dagger e,$$

typically is dominated by the propagated error $A^\dagger e$ and therefore is meaningless.

Tikhonov regularization seeks to determine a useful approximation of $\hat{x}$ by replacing the minimization problem (32) by a penalized least-squares problem of the form

$$\min_{x\in\mathbb{C}^n} \{\|Ax - b\|_2^2 + \mu\|Lx\|_2^2\}, \tag{34}$$

where the matrix $L \in \mathbb{C}^{k\times n}$, $k \leq n$, is referred to as the regularization matrix. It is commonly chosen to be a square or trapezoidal Toeplitz matrix, such as the identity matrix, the $(n-1)\times n$ matrix $T'$ obtained by removing the first row from $T = (n; 0, 1, -1)$, or the $(n-2)\times n$ matrix

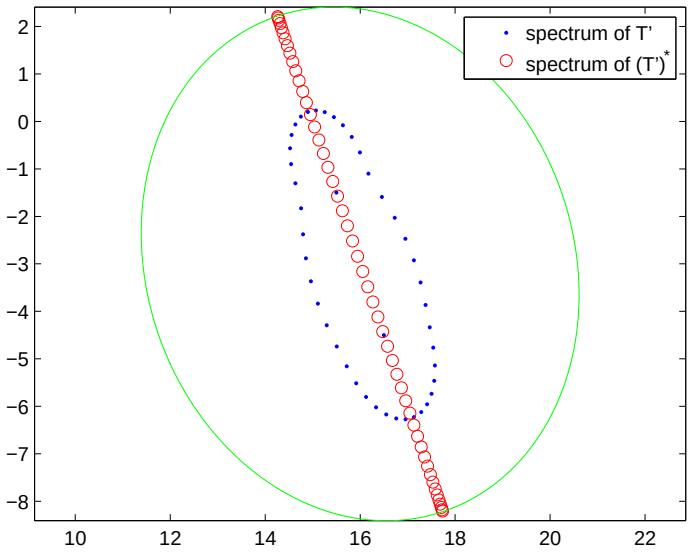



Figure 5. Spectra of the matrices $T_{(0.1)}^T$ and $(T_{(0.1)}^T)^*$ (denoted by $T'$ and $(T')^*$, respectively, in the legend) computed with the QR algorithm as implemented by the MATLAB function eig. The horizontal axis shows the real part and the vertical axis the imaginary part of the eigenvalues.

$T''$ determined by removing the first and last rows from $T = (n; -1, 2, -1)$. The regularization matrices $T'$ and $T''$ are finite difference approximations of the first and second derivatives in one space-dimension, respectively. The scalar $\mu > 0$ is the regularization parameter. In many discrete ill-posed problems (32), the matrix $A$ has a numerical null space of dimension larger than zero. It is the purpose of the regularization term $\mu\|Lx\|_2^2$ in (34) to damp unwanted behavior of the computed solution; see, e.g., [5, 17, 27, 33] and references therein for discussions on Tikhonov regularization and the choice of regularization matrices.

Let $L$ be such that the null spaces of $A$ and $L$ intersect trivially. Then the minimization problem (34) has the unique solution

$$x_{L,\mu} = (A^T A + \mu L^T L)^{-1} A^T b,$$

The size of $\mu$ determines how well the vector $x_{L,\mu}$ approximates $\hat{x}$ and how sensitive $x_{L,\mu}$ is to the error $e$ in $b$. The quality of $x_{L,\mu}$ also depends on the choice of regularization matrix $L$. This is illustrated below.

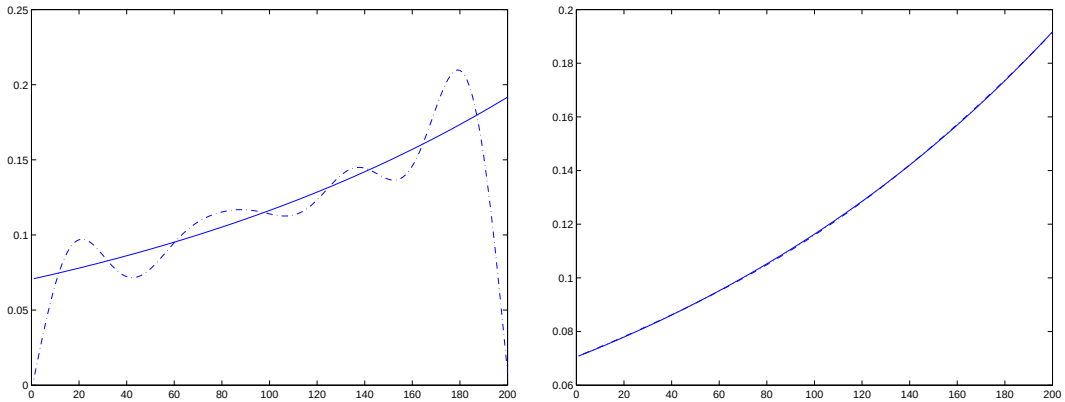
It is the purpose of this section to show that the solution $T \in \mathbb{C}^{(n-2)\times n}$ of Inverse Problem 2 of Section 7 with $x$ an available approximate solution of (32), such as $x = x_{I,\mu}$, can be a suitable regularization matrix for (34). The rationale for using the regularization matrix $L = T$ is that we do not want the regularization matrix to damp important features of the desired solution $\hat{x}$ when solving (34). Ideally, we would like to solve (30) for $L = T$ with $x = \hat{x}$; however, since $\hat{x}$ is not known, we let $x$ in (30) be the best available approximation of $\hat{x}$. Example 8.1 below illustrates application of this approach in an iterative fashion.

We assume that an estimate $\delta$ of $\|e\|$ is available. This allows us to determine the regularization parameter $\mu$ with the aid of the discrepancy principle. Specifically, we choose $\mu > 0$ so that

$$\|Ax_{L,\mu} - b\|_2 = \delta; \tag{35}$$

however, we remark that other approaches to determine $\mu$ also can be used, such as the L-curve and generalized cross validation; see, e.g., [17].

We will solve (34) for a general matrix $L$ by using the generalized singular value decomposition (GSVD) of the matrix pair $\{A, L\}$. It is then easy to determine $\mu$ from the nonlinear equation (35). When $L = I$, the generalized singular value decomposition can be replaced by the (standard) singular value decomposition (SVD); see, e.g., Hansen [17] for details on the applications of the GSVD or SVD to the solution of (34).




(a) (b)

Figure 6. Solution $\hat{x}$ to the error-free problem (33) (solid curves) and computed approximations (dash-dotted curves); the approximate solutions are $x_{I,\mu}$ in (a) and $x_4$ in (b). Note the different scalings of the vertical axes.

Example 8.1. Consider the Fredholm integral equation of the first kind

$$\int_0^1 k(s,t)x(t)dt = e^s + (1-e)s - 1, \qquad 0 \le s \le 1, \tag{36}$$

where

$$k(s,t) = \begin{cases} s(t-1), & s < t, \\ t(s-1), & s \ge t. \end{cases}$$

This equation is discussed, e.g., by Delves and Mohamed [10, p. 315]. We discretize the integral equation by a Galerkin method with orthonormal box functions as test and trial functions using the MATLAB function deriv2 from Regularization Tools [18]. The function yields a symmetric indefinite matrix $A \in \mathbb{R}^{200\times 200}$ and a scaled discrete approximation $\hat{x} \in \mathbb{R}^{200}$ of the solution $x(t) = e^t$ of (36). The error-free right-hand side vector in (33) is computed as $\hat{b} = A\hat{x}$. The entries of the error $e$ in $b$ are normally distributed with zero mean, and they are scaled to correspond to 1% error.

We first compute the approximate solution $x_{I,\mu}$ of (32) by solving (34) with $L = I$, and with $\mu > 0$ determined by the discrepancy principle. Figure 6(a) displays $x_{I,\mu}$ (dash-dotted curve)

as well as the desired solution $\hat{x}$ (solid curve) of the error-free system (33). The error $x_{I,\mu} - \hat{x}$ is seen to be quite large; we have $\|x_{I,\mu} - \hat{x}\|_2 = 2.42 \cdot 10^{-1}$.

Next we determine a trapezoidal tridiagonal Toeplitz regularization matrix $T \in \mathbb{R}^{198\times 200}$ by solving Inverse Problem 2 with $x = x_{I,\mu}$. The regularization matrix $L = T$ so obtained is used in (34) to compute a new approximate solution, $x_1$, of (32) with the aid of the discrepancy principle. The vector $x_1$ is a better approximation of $\hat{x}$ than $x_{I,\mu}$; we have $\|x_1 - \hat{x}\|_2 = 8.98 \cdot 10^{-2}$. We now can solve (30) with $x = x_1$ to determine a new trapezoidal tridiagonal Toeplitz regularization matrix $L = T$. Using this regularization matrix in (34) yields an improved approximate solution, $x_2$, of $\hat{x}$ with $\|x_2 - \hat{x}\|_2 = 4.08 \cdot 10^{-2}$. Similarly, we compute $x_3$ and $x_4$ with errors $\|x_3 - \hat{x}\|_2 = 2.53 \cdot 10^{-2}$ and $\|x_4 - \hat{x}\|_2 = 1.74 \cdot 10^{-3}$. Figure 6(b) displays $x_4$. The values of the regularization parameters $\mu$ are determined by the discrepancy principle for all solutions $x_j$.

The regularization matrix obtained by solving (30) generally is of better quality, the better the vector $x$ in (30) approximates $\hat{x}$. For instance, when $x = \hat{x}$, solution of (30) gives a regularization matrix $L = T$ such that the error in the subsequently computed Tikhonov solution $x_{L,\mu}$ is $\|x_{L,\mu} - \hat{x}\|_2 = 1.19 \cdot 10^{-3}$.

Commonly used regularization matrices $L$ in (34) include the rectangular bidiagonal Toeplitz matrix $T' \in \mathbb{R}^{(n-1)\times n}$ and the rectangular tridiagonal Toeplitz matrix $T'' \in \mathbb{R}^{(n-2)\times n}$ introduced above; see, e.g., [5, 17, 33]. When using $L = T'$ with $n = 200$ in (34) for the present example, and determining $\mu$ by the discrepancy principle, we obtain the approximate solution $x'$ with error $\|x' - \hat{x}\|_2 = 3.05 \cdot 10^{-2}$. Similarly, solving (34) with $L = T''$ yields the approximate solution $x''$ with $\|x'' - \hat{x}\|_2 = 5.79 \cdot 10^{-3}$. Thus, $x_4$ is a better approximation of $\hat{x}$ than $x'$ and $x''$.

We remark that determining a regularization matrix by solving the minimization problem (30) obviates the need to guess the appropriate form of the regularization matrix. □

## 9. Generation of Krylov subspace bases

Restarted GMRES is one of the most popular iterative methods for the solution of linear systems of equations

$$Ax = b, \quad A \in \mathbb{C}^{m\times m}, \quad x, b \in \mathbb{C}^m, \tag{37}$$

with a large sparse nonsymmetric and nonsingular matrix; see [35]. The method is based on repeatedly projecting the system (37) into Krylov subspaces of smaller size and solving the sequence of reduced problems so obtained.

Let $x_0$ be an available approximate solution of (37) and define the associated residual error $r = b - Ax_0$. GMRES computes an improved approximation $x_1 = x_0 + \Delta x_0$ by determining a correction $\Delta x_0$ in a Krylov subspace

$$\mathcal{K}_n(A, r) = \text{span}\{r, Ar, A^2r, \ldots, A^{n-1}r\} \tag{38}$$

of dimension $n \ll m$. The standard GMRES implementation uses the Arnoldi process to compute an orthonormal basis for (38). Application of $n < m$ steps of the Arnoldi process to $A$ with initial vector $r \in \mathbb{C}^m$ yields the decompositions

$$AV_n = V_{n+1}H_{n+1,n} = V_nH_n + \alpha_n v_{n+1}e_n^T, \tag{39}$$

where the columns of $V_n$ form an orthonormal basis for (38) and $H_{n+1,n} \in \mathbb{C}^{(n+1)\times n}$ is an upper Hessenberg matrix. The matrix $H_n \in \mathbb{C}^{n\times n}$ is obtained by removing the last row of $H_{n+1,n}$ and the vector $v_{n+1}$ is the last columns of $V_{n+1}$.

The correction $\Delta x_0 = V_n y$ of $x_0$ is the solution of the least-squares problem

$$\min_{\Delta x_0 \in \mathcal{K}_n(A,r)} \|A\Delta x_0 - r\|_2 = \min_{y\in\mathbb{C}^n} \|H_{n+1,n}y - e_1\|b\|_2\,\|_2.$$

Due to storage and work considerations, $n$ generally is chosen much smaller than $m$; in many applications $20 \le n \le 50$. Therefore, the computed approximate solution $x_1$ of (37) typically is not of desired accuracy. One then seeks to determine an improved approximate solution $x_2 = x_1 + \Delta x_1$ by determining a correction $\Delta x_1$ in (38) with $r = b - Ax_1$. The vector $\Delta x_1$ can be computed similarly as $\Delta x_0$, i.e., by application of $n$ steps of the Arnoldi process. Generally, several corrections $\Delta x_j$ have to be computed until a sufficiently accurate approximate solution of (37) has been found.

The Arnoldi process determines one column of the matrix $V_n$ at a time. Each new column is orthogonalized against all already available columns by the modified Gram-Schmidt method. This makes it difficult to achieve high performance on parallel computers. Therefore, the use of nonorthogonal Krylov subspace bases, that circumvent the sequential orthogonalization of the Arnoldi process and lend themselves better to efficient implementation on parallel computers, has received considerable attention; see, e.g., [2, 14, 21, 22, 32, 36]. We remark that the basis in (38) generally cannot be used, because for many matrices $A$ it is very ill-conditioned; in fact the vectors $A^j b$ in (38) may be numerically linearly dependent already for $n$ of modest size.

We would like to use a Krylov subspace basis that is easy to construct and is numerically linearly independent in finite precision arithmetic. Krylov subspace bases based on translated and scaled Chebyshev polynomials $p_0, p_1, p_2, \ldots$ of the first kind, that are orthogonal with respect to an inner product on some interval in the complex plane,

$$\mathcal{S} = \{tz_1 + (1-t)z_2 : 0 \le t \le 1\}, \quad z_1, z_2 \in \mathbb{C}, \quad z_1 \neq z_2, \tag{40}$$

are convenient to use; see [21, 22, 32] and references therein. Here $p_j$ is a polynomial of degree $j$. One can evaluate the basis

$$\{p_0(A)r, p_1(A)r, \ldots, p_{n-1}(A)r\} \tag{41}$$

for (38) without sequential orthogonalization, by using the three-term recursion formula for the $p_j$. Subsequent orthogonalization of the basis (41) by QR factorization of the matrix with columns $p_j(A)r$, $0 \le j < n$, can be carried out efficiently on a parallel computer; see [4, 21, 22, 32] for discussions. The computations require the vectors (41) to be numerically linearly independent. This is typically satisfied with an appropriate choice of the interval (40); see [21, 32] for analyses. The polynomials are scaled so that the vectors $p_j(A)r$ are of unit length.

A suitable interval (40) for defining the translated and scaled Chebyshev polynomials often can be determined from the spectrum of the matrix $H_n$ computed by the Arnoldi process (39) when computing the initial correction $\Delta x_0$. A common approach described in the literature, see, e.g., [21, 22, 32] and references therein, is to determine the smallest ellipse that contains the spectrum of $H_n$, and let $z_1$ and $z_2$ be the foci of this ellipse. The translated Chebyshev polynomials associated with the interval (40), suitable scaled, are used in all subsequent restarts until a sufficiently accurate approximate solution of (37) has been found. The use of bases of the

form (41) sidesteps the need to apply the Arnoldi process in restarts and yields an algorithm that is well suited for implementation on parallel computers; see, e.g., [22, 32] for discussions.

However, the determination of the smallest ellipse that contains a given point set is a fairly complicated computational task. We describe two ways, based on properties of tridiagonal Toeplitz matrices, to simplify the computations. First we transform $H_n$ to a similar non-Hermitian tridiagonal matrix $T_n$ by application of the non-Hermitian Lanczos process to $H_n$ with initial vectors $e_1$. Our first approach to determine a suitable interval (40) is to solve the minimization problem

$$\min_{T\in\mathcal{T}} \|T - T_n\|_F \tag{42}$$

for the matrix $\hat{T} = (n; \sigma, \delta, \tau)$. We then let (40) be the line segment (6) determined by $\hat{T}$. These computations are very simple. Since the spectrum of $\hat{T}$ is explicitly known, the smallest interval containing all eigenvalues can be determined accurately also when $\hat{T}$ is highly nonnormal.

Alternatively, we may determine the interval (40) by using the field of values of $T_n$, defined by

$$\mathcal{W}(T_n) = \left\{ \frac{x^H T_n x}{x^H x}, \quad x \in \mathbb{C}^n \backslash \{0\} \right\}.$$

Let $\hat{T} = (n; \sigma, \delta, \tau)$ be the solution of (42). We now determine a region in $\mathbb{C}$ that contains $\mathcal{W}(T_n)$ as follows; see [31] for further details. The closest normal tridiagonal Toeplitz matrix to $T_n$, denoted by $T^*$, is the normal tridiagonal Toeplitz matrix closest to $\hat{T}$. Therefore,

$$\mathcal{W}(T^*) = \left\{ \delta + t\, e^{i(\arg\sigma + \arg\tau)/2} : \, t \in \mathbb{R}, \, |t| \le (|\sigma| + |\tau|) \cos\frac{\pi}{n+1} \right\}; \tag{43}$$

cf. Corollary 3.1. Moreover,

$$\begin{aligned} \mathcal{W}(T_n) &\subset \mathcal{W}(T^*) + \mathcal{W}(T_n - T^*), \\ \mathcal{W}(T_n - T^*) &\subset \{z \in \mathbb{C} : \, |z| \le \|T_n - T^*\|_F\}. \end{aligned}$$

The evaluation of $\|T_n - T^*\|_F$ is straightforward and so is the computation of a sports field-shaped region $\mathcal{R}$ that contains $\mathcal{W}(T_n)$. We may let (40) be the interval between the foci of the largest ellipse that can be inscribed in $\mathcal{R}$ or, simpler, the interval (43).

Example 9.1. We illustrate the first approach. Consider the elliptic boundary value problem

$$\begin{aligned} -\Delta u + \gamma \frac{\partial u}{\partial s} &= f \ \text{ in } \ \Omega, \\ u &= 0 \ \text{ on } \ \partial\Omega, \end{aligned} \tag{44}$$

where $\Omega$ is the unit square in the $(s,t)$-plane with boundary $\partial\Omega$ and $\gamma = 60$. We approximate $\Delta$ and $\partial/\partial s$ by standard 2nd order finite differences, using 38 equidistant interior grid points in both the $s$- and $t$-directions. This yields a nonsymmetric nonsingular matrix $A \in \mathbb{R}^{1444\times 1444}$, which can be expressed as $I\otimes T_1 + T_2 \otimes I$, where $T_1$ and $T_2$ are tridiagonal Toeplitz matrices and $\otimes$ denotes Kronecker product. Using (4), one can derive explicit expressions for the eigenvalues of $A$; they are allocated in a rectangle that is symmetric with respect to the real axis in $\mathbb{C}$. We let $f \equiv 1$.

Figure 7 displays the computed spectrum of the matrix $A$ (blue dots) in the complex plane; the horizontal and vertical axes are the real and imaginary axes, respectively. The computed eigenvalues are not very accurate, because one of the tridiagonal matrices $T_j$ that determine

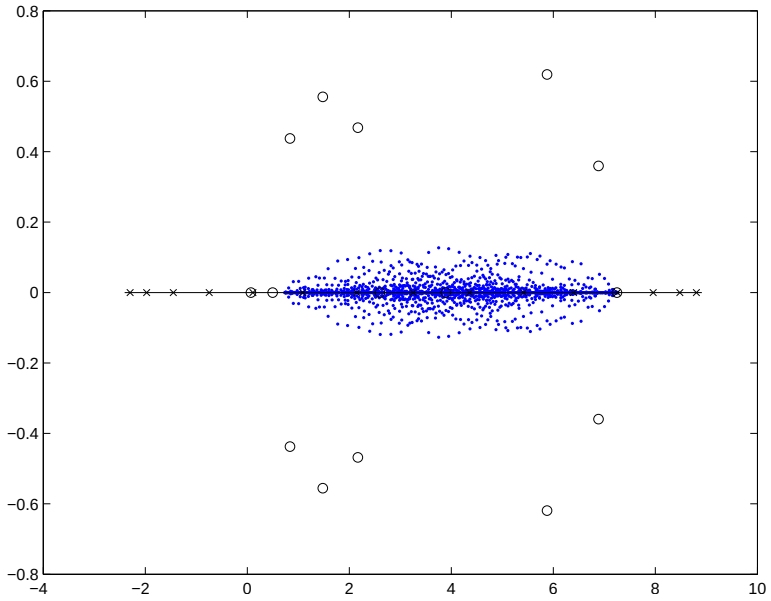



Figure 7. Computed spectra in the complex plane $\mathbb{C}$ of the matrices $A$ (blue dots), $H_{15}$ and $T_{15}$ (black circles), and of the tridiagonal Toeplitz matrix $T$ closest to $T_{15}$ (black crosses). The horizontal black line segment displays the interval between the foci of the ellipse associated with $T$. The horizontal axis marks the real part and the vertical axis the imaginary part of the eigenvalues.

$A$ is far from normal. The eigenvalues are computed with the MATLAB function eig. The difficulty of eig to compute accurate eigenvalue approximations already has been illustrated by Figure 5.

The black circles in Figure 7 mark 15 Ritz values, i.e., the 15 eigenvalues of the matrix $H_{15}$ in (39) determined by 15 steps of the Arnoldi process applied to $A$ with the initial vector a multiple of $[1, 1, \ldots, 1]^T$. A common approach to determine an interval that defines a family of Chebyshev polynomials $p_j$ is to compute the smallest ellipse that contains these Ritz values.

We instead proceed to determine a nonsymmetric tridiagonal matrix $T_n$ that is similar to $H_n$ by the nonsymmetric Lanczos process, and then compute the tridiagonal Toeplitz matrix $\hat{T}$ that satisfies (42). The spectrum of the latter matrix is marked by black crosses in Figure 7, which also shows the interval between the foci associated with $\hat{T}$; cf. (27). This interval contains all the eigenvalues of $\hat{T}$. We propose to use a scaled and translated Chebyshev polynomial basis associated with this interval.

We have $\|\hat{T} - T_n\|_F = 4.15 \cdot 10^1$. Moreover, $\|\hat{T} - T^*\|_F = 6.17$, where $T^*$ denotes the closest matrix to $\hat{T}$ in $\mathcal{N}_{\mathcal{T}}$, which shows that $\hat{T}$ is quite close to normal.

Since the coefficient $\gamma$ in (44) is large, the solution displays a steep transient. Figure 8 shows the solution of the discretized problem at interior and boundary grid points. We remark that similar results are obtained for other discretizations of the boundary value problem (44). □

Example 9.2. The boundary value problem and discretization are the same as in Example 9.1, except that the coefficient in (44) is $\gamma = 6$. This makes the spectrum of the nonsymmetric matrix $A \in \mathbb{R}^{1444\times 1444}$ real; the smallest and largest eigenvalues of $A$ are $1.89 \cdot 10^{-2}$ and 7.98,

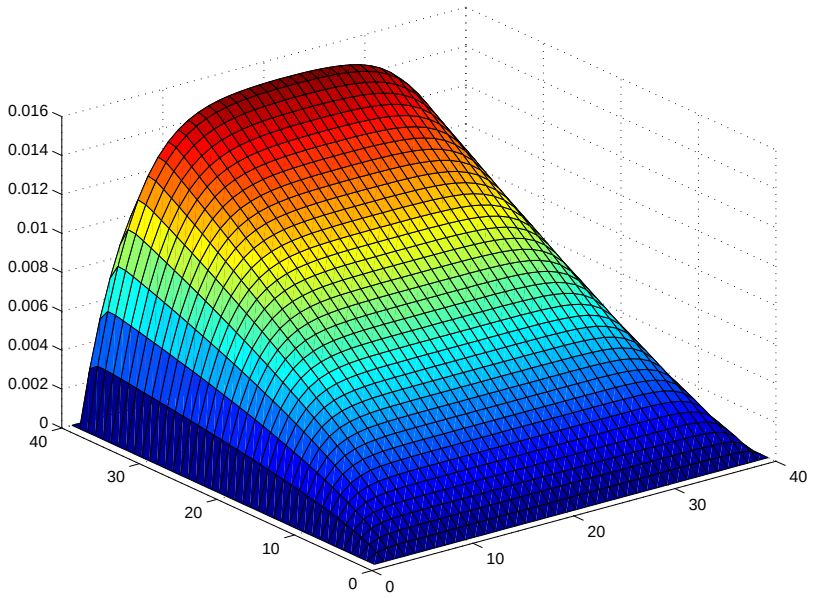



Figure 8. The solution of the discretized boundary value problem (44) with $\gamma = 60$ at interior and boundary grid points.

respectively.

Figure 9 shows 15 Ritz values of $A$, i.e., the spectra of the matrices $H_{15}$ in (39) and of the nonsymmetric tridiagonal matrix $T_{15}$ (black circles). All Ritz values are seen to be real. The spectrum of the closest tridiagonal Toeplitz matrix $\hat{T}$, i.e., the solution of (42), is displayed by black crosses. The figure also shows the interval between the foci associated with $\hat{T}$; cf. (27). This interval contains all the eigenvalues of $\hat{T}$. We may use a scaled and translated Chebyshev polynomial basis associated with this interval. Finally, Figure 9 depicts the eigenvalues of the closest normal tridiagonal Toeplitz matrix $T^*$ to $\hat{T}$; they are marked by red plus signs. We also can use the interval between the foci of $T^*$ to define the translated and scaled Chebyshev polynomials $p_j$ in (41). We have $\|\hat{T} - T_n\|_F = 4.91$ and $\|\hat{T} - T^*\|_F = 1.46 \cdot 10^{-1}$.

Figure 10 shows the solution of the discretized problem at interior and boundary grid points. □

## 10. Conclusion

This paper discusses the conditioning of eigenvalues of tridiagonal Toeplitz matrices. The simple structure of these matrices makes it possible to derive simple expressions and bounds for the individual, global, traditional, and structured condition numbers. This led us to discuss several applications, including an inverse eigenvalue problem. New applications of tridiagonal Toeplitz matrices to the construction of regularization matrices for Tikhonov regularization and to the construction of Krylov subspace bases are described. These applications are very promising and will be investigated in more detail in forthcoming work.

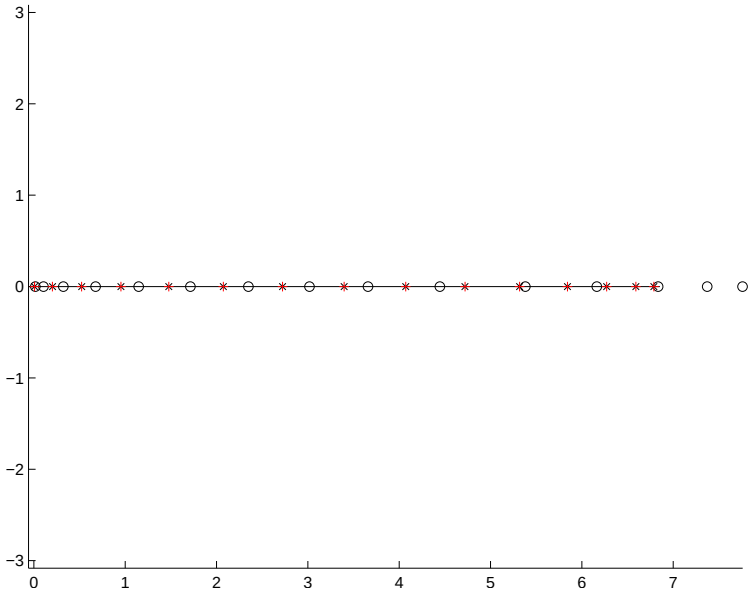



Figure 9. Spectra of the matrices $H_{15}$ and $T_{15}$ (black circles), of the tridiagonal Toeplitz matrix $\hat{T}$ closest to $T_{15}$ (black crosses), and of $T^*$, the closest matrix in $\mathcal{N}_{\mathcal{T}}$ to $\hat{T}$ (red pluses). The horizontal black line segment displays the interval between the foci of the ellipse associated with $\hat{T}$. The eigenvalues are shown in $\mathbb{C}$, but they are all real.

## Acknowledgement

We would like to thank the referees for comments.

## REFERENCES

1. M. Arnold and B. N. Datta, Single-input eigenvalue assignment algorithms: a close look, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 444–467.
2. Z. Bai, D. Hu, and L. Reichel, A Newton basis GMRES implementation, IMA J. Numer. Anal., 14 (1994), pp. 563–581.
3. A. Böttcher and S. Grudsky, Spectral Properties of Banded Toeplitz Matrices, SIAM, Philadelphia, 2005.
4. D. Calvetti, J. Petersen, and L. Reichel, A parallel implementation of the GMRES algorithm, in Numerical Linear Algebra, eds. L. Reichel, A. Ruttan, and R. S. Varga, de Gruyter, Berlin, 1993, pp. 31–46
5. D. Calvetti, L. Reichel, and A. Shuibi, Invertible smoothing preconditioners for linear discrete ill-posed problems, Appl. Numer. Math., 54 (2005), pp. 135–149.
6. B. N. Datta, An algorithm to assign eigenvalues in a Hessenberg matrix: single input case, IEEE Trans Autom. Control, AC-32, (1987), pp. 414–417.
7. B. N. Datta, W.-W. Lin, and J.-N. Wang, Robust partial pole assignment for vibrating systems with aerodynamic effects, IEEE Trans. Autom. Control, 51 (2006), pp. 1979–1984.
8. B. N. Datta and Y. Saad, Arnoldi methods for large Sylvester-like observer matrix equations, and an associated algorithm for partial spectrum assignment, Linear Algebra Appl., 154-156 (1991), pp. 225–244.
9. B. N. Datta and V. Sokolov, A solution of the affine quadratic inverse eigenvalue problem, Linear Algebra Appl., 434 (2011), pp. 1745–1760.
10. L. M. Delves and J. L. Mohamed, Computational Methods for Integral Equations, Cambridge University Press, Cambridge, 1985.

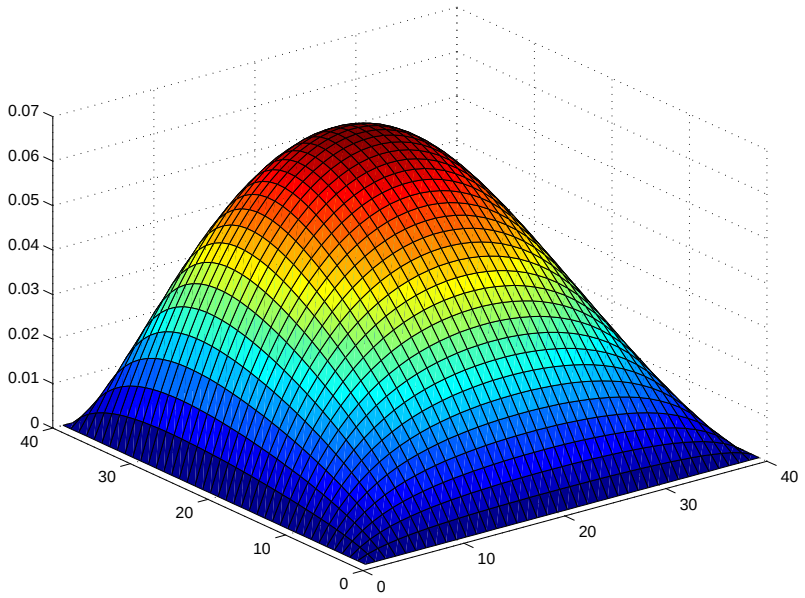



Figure 10. The solution of the discretized boundary value problem (44) with $\gamma = 6$ at interior and boundary grid points.


11. J. W. Demmel, Nearest defective matrices and the geometry of ill-conditioning, in Reliable Numerical Computation, M. G. Cox and S. Hammarling, eds., Clarendon Press, Oxford, 1990, pp. 35–55.
12. F. Diele and L. Lopez, The use of the factorization of five-diagonal matrices by tridiagonal Toeplitz matrices, Appl. Math. Lett., 11 (1998), pp. 61–69.
13. L. Elsner and M. H. C. Paardekooper, On measures of nonnormality of matrices, Linear Algebra Appl., 92 (1987), pp. 107–124.
14. J. Erhel, A parallel GMRES version for general sparse matrices, Electron. Trans. Numer. Anal., 3 (1995), pp. 160–176.
15. D. Fischer, G. Golub, O. Hald, C. Leiva, and O. Widlund, On Fourier-Toeplitz methods for separable elliptic problems, Math. Comp., 28 (1974), pp. 349–368.
16. G. H. Golub and J. H. Wilkinson, Ill-conditioned eigensystems and the computation of the Jordan canonical form, SIAM Rev., 18 (1976), pp. 578–619.
17. P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.
18. P. C. Hansen, Regularization tools version 4.0 for MATLAB 7.3, Numer. Algorithms, 46 (2007), pp. 189–194.
19. P. Henrici, Bounds for iterates, inverses, spectral variation and field of values of non-normal matrices, Numer. Math., 4 (1962), pp. 24–40.
20. N. J. Higham, Matrix nearness problems and applications, in Applications of Matrix Theory, M. J. C. Gover and S. Barnett, eds., Oxford University Press, Oxford, 1989, pp. 1–27.
21. W. D. Joubert and G. F. Carey, Parallelizable restarted iterative methods for nonsymmetric linear systems. Part I: Theory, Intern. J. Computer Math., 44 (1992), pp. 243–267.
22. W. D. Joubert and G. F. Carey, Parallelizable restarted iterative methods for nonsymmetric linear systems. Part II: Parallel implementation, Intern. J. Computer Math., 44 (1992), pp. 269–290.
23. M. Karow, D. Kressner, and F. Tisseur, Structured eigenvalue condition numbers, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1052–1068.
24. L. László, An attainable lower bound for the best normal approximation, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 1035–1043.
25. S. L. Lee, Best available bounds for departure from normality, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 984–991.
26. A. Luati and T. Proietti, On the spectral properties of matrices associated with trend filters, Econometric

Theory, 26 (2010), pp. 1247–1261.
27. S. Morigi, L. Reichel, and F. Sgallari, A truncated projected SVD method for linear discrete ill-posed problems, Numer. Algorithms, 43 (2006), pp. 197–213.
28. S. Noschese and L. Pasquini, Eigenvalue condition numbers: zero-structured versus traditional, J. Comput. Appl. Math., 185 (2006), pp. 174–189.
29. S. Noschese and L. Pasquini, Eigenvalue patterned condition numbers: Toeplitz and Hankel cases, J. Comput. Appl. Math., 206 (2007), pp. 615–624.
30. S. Noschese, L. Pasquini, and L. Reichel, The structured distance to normality of an irreducible real tridiagonal matrix, Electron. Trans. Numer. Anal., 28 (2007), pp. 65–77.
31. S. Noschese and L. Reichel, The structured distance to normality of banded Toeplitz matrices, BIT, 49 (2009), pp. 629–640.
32. B. Philippe and L. Reichel, On the generation of Krylov subspace bases, Appl. Numer. Math., in press.
33. L. Reichel and Q. Ye, Simple square smoothing regularization operators, Electron. Trans. Numer. Anal., 33 (2009), pp. 63–83.
34. L. Reichel and L. N. Trefethen, Eigenvalues and pseudo-eigenvalues of Toeplitz matrices, Linear Algebra Appl., 162-164 (1992), pp. 153–185.
35. Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed., SIAM, Philadelphia, 2003.
36. R. B. Sidje, Alternatives to parallel Krylov subspace basis computation, Numer. Linear Algebra Appl., 4 (1997), pp. 305–331.
37. G. D. Smith, Numerical Solution of Partial Differential Equations, 2nd ed., Clarendon Press, Oxford, 1978.
38. L. Smithies, The structured distance to nearly normal matrices, Electron. Trans. Numer. Anal., 36 (2010), pp. 99–112.
39. G. W. Stewart and J. Sun, Matrix Perturbation Theory, Academic Press, London, 1990.
40. L. N. Trefethen and M. Embree, Spectra and Pseudospectra, Princeton University Press, Princeton, 2005.
41. W.-C. Yueh and S. S. Cheng, Explicit eigenvalues and inverses of tridiagonal Toeplitz matrices with four perturbed corners, ANZIAM J., 49 (2008), pp. 361–387.
42. J. H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965.
43. J. H. Wilkinson, Sensitivity of eigenvalues II, Util. Math., 30 (1986), pp. 243–286.