BIT manuscript No.
(will be inserted by the editor)

The structured distance to normality of banded Toeplitz matrices

Silvia Noschese · **Lothar Reichel**

Received: date / Accepted: date

Dedicated to Axel Ruhe on the occasion of his 65th birthday.

Abstract Spectral properties of normal $(2k+1)$-banded Toeplitz matrices of order n, with $k \leq \lfloor n/2 \rfloor$, are described. Formulas for the distance of $(2k+1)$-banded Toeplitz matrices to the algebraic variety of similarly structured normal matrices are presented.

Keywords banded Toeplitz matrix · matrix nearness problem · distance to normality

Mathematics Subject Classification (2000) 65F15 · 15A57

1 Introduction

Matrix nearness problems have been the focus of much research in linear algebra; see, e.g., [8, 9, 19, 32]. In particular, characterizations of the algebraic variety of normal matrices and distance measures to this variety have received considerable attention; see [10, 11, 13, 14, 16, 17, 23, 25–27]. Normal matrices are of interest because their eigenvalues are optimally conditioned and their singular values are the magnitude of the eigenvalues. Numerical methods for the computation of eigenvalues of Hermitian matrices are simpler than methods designed for the computation of eigenvalues of general matrices. In view of Corollary 2.1 below, the eigenvalues of a normal banded Toeplitz matrix of suitably restricted bandwidth can be determined by

Research supported by a grant from SAPIENZA Università di Roma.

S. Noschese
Dipartimento di Matematica "Guido Castelnuovo", SAPIENZA Università di Roma, P.le A. Moro, 2, I-00185 Roma, Italy.
E-mail: noschese@mat.uniroma1.it

L. Reichel
Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA.
E-mail: reichel@math.kent.edu

computing the eigenvalues of a Hermitian banded Toeplitz matrix followed by a simple transformation. Similarly, the eigenvectors of a normal banded Toeplitz matrix can be computed as eigenvectors of a Hermitian matrix. It therefore may be of interest to compute eigenvalues and eigenvectors of a nearby normal banded Toeplitz matrix instead of a nonnormal matrix. We also note that inverse iteration, which can be applied to determine the eigenvalue of smallest magnitude of a general square matrix, has been shown to perform the best when applied to normal matrices; see [22].

We remark that the eigenvalues of large banded nonnormal Toeplitz matrices often are poorly conditioned; the sensitivity of the eigenvalues of a banded Toeplitz matrix grows exponentially with the dimension n, except when the boundary of the spectrum of the associated Toeplitz operator is a curve with no interior. This curve is related to the ε-pseudospectrum of the matrix as $\varepsilon \to 0$ and $n \to \infty$; see [30, Theorem 3.2]. Beam and Warming [1] consider the eigenvalue problem for banded Toeplitz matrices to be "numerically unreliable." A recent treatment of asymptotic properties of the spectra of banded Toeplitz matrices is provided in [5].

Normal Toeplitz matrices are characterized in [13, 17, 23]. In particular, Gu and Patton [17, Theorem 3.4] show, by using properties of differences of products of Toeplitz matrices, that a Toeplitz matrix

$$T = \begin{bmatrix} \delta & \tau_1 & \tau_2 & & \dots & \tau_{n-1} \\ \sigma_1 & \delta & \tau_1 & \tau_2 & \dots & \tau_{n-2} \\ \sigma_2 & \sigma_1 & & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & & \\ & & & & \tau_1 & \tau_2 \\ \sigma_{n-2} & & \dots & \sigma_1 & \delta & \tau_1 \\ \sigma_{n-1} & & \dots & \sigma_2 & \sigma_1 & \delta \end{bmatrix} \in \mathbf{C}^{n\times n}$$

is normal if and only if $\sigma_\ell = e^{i\varphi}\bar{\tau}_\ell$, $1 \le \ell < n$, or $\sigma_\ell = e^{i\varphi}\tau_{n-\ell}$, $1 \le \ell < n$, for some $-\pi < \varphi \le \pi$. Here and below $i = \sqrt{-1}$ and the bar denotes complex conjugation. Thus, normal Toeplitz matrices are either modifications of Hermitian matrices or are so-called $\{e^{i\varphi}\}$-circulant matrices. Properties of the latter are discussed in, e.g., [7, Section 3.2.1]. Other proofs of the characterization of normal Toeplitz matrices are presented by Farenick et al. [13] and Ito [23].

This paper investigates the distance of $(2k+1)$-banded Toeplitz matrices of order n, with $k \le \lfloor n/2 \rfloor$, to the algebraic variety of normal Toeplitz matrices of the same bandwidth. Here $\lfloor \alpha \rfloor$ denotes the integer part of α. The distance is measured with the Frobenius norm $\|\cdot\|_F$. Since the normal and nonnormal matrices are required to have the same band- and Toeplitz structure, we refer to the distance as the *structured distance*. Whether the structured distance to normality is more meaningful than the (unstructured) distance to normality depends on the application. An example involving a Jordan block, which is a bidiagonal Toeplitz matrix, where this is the case is described in [29, Example 9.1]. This example illustrates that the structured distance to normality can be significantly larger than the (unstructured) distance to normality. We comment further on this at the end of Section 3.

It is the purpose of the present paper to illustrate that many properties of banded Toeplitz matrices can be shown in a simpler way than for general Toeplitz matrices by exploiting the bandedness. Our results complement and extend those in [29] for tridiagonal matrices. Banded Toeplitz matrices arise in many applications in signal processing, time-series analysis, and numerical methods for the solution of partial differential equations; see, e.g., [15,24,34]. Low-rank modifications of symmetric banded Toeplitz matrices are considered in [28].

In the numerical analysis community, the papers by Henrici [18] and Ruhe [33] on the distance to normality and the computation of the closest normal matrix, respectively, have received particular attention. The algorithm presented in the latter paper is iterative and computationally expensive for large matrices. The present paper illustrates that the closest normal matrix and the distance to normality easily can be determined if further structure is imposed.

This paper is organized as follows. Section 2 presents a simple proof of necessary and sufficient conditions for a $(2k+1)$-banded Toeplitz matrix of order n, with $k \leq \lfloor n/2 \rfloor$, to be normal. Spectral properties also are discussed. Section 3 is concerned with the structured distance of banded Toeplitz matrices to the algebraic variety of normal banded Toeplitz matrices of the same bandwidth. The remainder of this section introduces notation used in the sequel.

Toeplitz matrices in $\mathbf{C}^{n\times n}$ with bandwidth $2k+1$ are denoted by

$$T_{(k)} = (n;k;\sigma,\delta,\tau) = \begin{bmatrix} \delta & \tau_1 & \tau_2 & \dots & \tau_k & & & \mathbf{0} \\ \sigma_1 & \delta & \tau_1 & & & & & \\ \sigma_2 & \sigma_1 & \ddots & & & & \ddots & \\ \vdots & & & & \ddots & & & \tau_k \\ \sigma_k & & \ddots & \ddots & \ddots & & & \vdots \\ & & & & & & \tau_1 & \tau_2 \\ & & \ddots & & & \sigma_1 & \delta & \tau_1 \\ \mathbf{0} & & & \sigma_k & \dots & \sigma_2 & \sigma_1 & \delta \end{bmatrix}. \tag{1.1}$$

Some of the scalars σ_j, δ, and τ_j may vanish. We say that the matrix (1.1) has bandwidth $2k+1$, or equivalently, is $(2k+1)$-banded, even if σ_k or τ_k vanish.

Let $\mathscr{N}$ denote the algebraic variety of the normal matrices in $\mathbf{C}^{n\times n}$ and let $\mathscr{T}_{(k)}$ be the subspace of $\mathbf{C}^{n\times n}$ formed by $(2k+1)$-banded Toeplitz matrices. The following subsets of $\mathscr{N}$ and $\mathscr{T}_{(k)}$ are of particular interest,

$$\mathscr{N}_{\mathscr{T}_{(k)}} = \mathscr{N} \cap \mathscr{T}_{(k)}, \qquad \mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}} = \mathscr{N}_{\mathscr{T}_{(k)}} \cap \mathbf{R}^{n\times n}.$$

We measure distances with the Frobenius norm,

$$\begin{aligned} \Delta_F(A) &= \Delta_F(A, \mathscr{N}_{\mathscr{T}_{(k)}}) = \min\{\|E\|_F : A+E \in \mathscr{N}_{\mathscr{T}_{(k)}}\}, \\ \Delta_F^{\mathbf{R}}(A) &= \Delta_F(A, \mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}}) = \min\{\|E\|_F : A+E \in \mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}}\}. \end{aligned}$$

Let J be the *reversal matrix*, i.e., the matrix obtained by ordering the columns of the identity matrix I in reverse order. We use the superscript P to denote the persymmetric transpose,

$$A^P = JA^T J.$$

For brevity, we refer to A^P as the *perpose* of A. A matrix A is said to be *persymmetric* if it equals its perpose. In particular, Toeplitz matrices are persymmetric. The superscripts T and H denote transposition and transposition followed by complex conjugation, respectively.

2 Normal banded Toeplitz matrices

Theorem 2.1 *The $(2k+1)$-banded Toeplitz matrix $T_{(k)} = (n;k;\sigma,\delta,\tau)$, with $k \leq \lfloor n/2 \rfloor$, is normal if and only if there is an angle θ, such that*

$$\sigma_h = \overline{\tau}_h e^{i\theta}, \qquad h = 1:k. \tag{2.1}$$

Proof The theorem follows from the characterization of normal Toeplitz matrices shown in [13, 17, 23] and stated in the previous section. However, the bandedness makes it possible to show the result in a much simpler way than in these references. We therefore include the following elementary proof.

First note that $(2k+1)$-banded Toeplitz matrices, whose entries satisfy (2.1) are normal. Conversely, assume that $T_{(k)}$ is normal. Then

$$T_{(k)}^H T_{(k)} = T_{(k)} T_{(k)}^H.$$

The matrices in right-hand side and left-hand side are $(4k+1)$-banded and Hermitian. Equating the entries of the upper triangular parts of the leading $k \times k$ submatrices of the right-hand side and left-hand side yields

$$|\sigma_h| = |\tau_h|, \qquad h = 1:k, \tag{2.2}$$

and

$$\arg(\sigma_{i_1}) + \arg(\tau_{i_1}) = \arg(\sigma_{i_2}) + \arg(\tau_{i_2}) = \ldots = \arg(\sigma_{i_j}) + \arg(\tau_{i_j}),$$

where $\{i_1, i_2, \ldots, i_j\}$, for some $1 \leq j \leq k$, is the set of indices of the nonzero components of σ (and, due to (2.2), also of τ). This concludes the proof.

Remark 2.1 *The restriction on the bandwidth in Theorem 2.1 rules out non-Hermitian $\{e^{i\varphi}\}$-circulant Toeplitz matrices. Thus, the theorem applies to the Hermitian circulant*

$$T_{(2)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

but not to the non-Hermitian one

$$T_{(3)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Corollary 2.1 *Let the Toeplitz matrix $T_{(k)} = (n;k;\sigma,\delta,\tau)$, with $k \leq \lfloor n/2 \rfloor$, be normal. Then $T_{(k)} = \delta I + e^{i\theta/2} S_{(k)}$, where $S_{(k)} = (n;k;\sigma,0,\overline{\sigma})$ is a $(2k+1)$-banded Hermitian Toeplitz matrix.*

Proof The statement is a consequence of (2.1).

It follows from Corollary 2.1 that the eigenvalues of a normal banded Toeplitz matrix $T_{(k)} = (n;k;\sigma,\delta,\tau)$, $k \leq \lfloor n/2 \rfloor$, can be determined from the eigenvalues of the associated Hermitian Toeplitz matrix $S_{(k)}$ by an affine transformation and that the eigenvectors of $T_{(k)}$ are eigenvectors of $S_{(k)}$. Hence, it suffices to compute the eigenvalues and eigenvectors of the Hermitian Toeplitz matrix $S_{(k)}$ in order to determine the eigenvalues and eigenvectors of $T_{(k)}$.

Corollary 2.2 *A real $(2k+1)$-banded Toeplitz matrix is normal if and only if it is symmetric or shifted skew-symmetric.*

Proof The result follows from Theorem 2.1, since the angle θ may be either 0 or π.

We turn to the spectral properties of normal banded Toeplitz matrices. Introduce the symbol for the matrix $T_{(k)} - \delta I = (n;k;\sigma,0,\tau)$,

$$f(t) = \sum_{j=1}^{k} \left(\sigma_j e^{-ijt} + \tau_j e^{ijt} \right). \tag{2.3}$$

Theorem 2.2 *Let the $(2k+1)$-banded Toeplitz matrix $T_{(k)} = (n;k;\sigma,\delta,\tau)$, with $k \leq \lfloor n/2 \rfloor$, be normal. Then its eigenvalues are collinear. More precisely, let*

$$M = \max_{t\in\mathbf{R}} e^{-i\theta/2} f(t), \qquad m = \min_{t\in\mathbf{R}} e^{-i\theta/2} f(t),$$

where f is the symbol (2.3) and θ is determined by (2.1). Then the spectrum lives in the interval $\delta + e^{i\theta/2}[m,M]$.

Proof Let $\lambda_h(A)$, $h = 1:n$, denote the eigenvalues of the matrix $A \in \mathbf{C}^{n\times n}$. In view of Corollary 2.1, we have

$$\lambda_h(T_{(k)}) = \delta + e^{i\theta/2} \lambda_h(S_{(k)}), \qquad h = 1:n.$$

Since $S_{(k)} = (n;k;\sigma,0,\overline{\sigma})$ is Hermitian, the spectrum of $T_{(k)}$ lies on the straight line $\delta + e^{i\theta/2} t$, $t \in \mathbf{R}$. Moreover, $e^{-i\theta/2} f(t)$ is the symbol for $S_{(k)}$. By the connection between the eigenvalues and the symbol of a Toeplitz matrix, the spectrum of $S_{(k)}$ lives in the interval $[m,M]$; see, e.g., [15, Chapter 5] or [5]. This completes the proof.

Remark 2.2 *Faber and Manteuffel [12] investigated necessary and sufficient conditions on the matrix $A \in \mathbf{C}^{n\times n}$ for the existence of an iterative method of conjugate gradient-type with a three-term recurrence formula for the solution of linear systems of equations*

$$Ax = b, \qquad x, b \in \mathbf{C}^n. \tag{2.4}$$

Theorem 2.2 shows that normal $(2k+1)$-banded Toeplitz matrices $T_{(k)} = (n;k;\sigma,\delta,\tau)$, with $k \leq \lfloor n/2 \rfloor$, satisfy these conditions. Moreover, if the matrix A is a low-rank modification of such a Toeplitz matrix, then a slight extension of the development in [2] shows that there is a minimal residual method with short recursion formulas for the solution of (2.4) also in this situation.

Corollary 2.3 *The eigenvalues of a real normal $(2k+1)$-banded Toeplitz matrix $T_{(k)} = (n;k;\sigma,\delta,\tau)$, with $k \le \lfloor n/2 \rfloor$, live on the real axis or on the straight line parallel to the imaginary axis through δ.*

Proof The proof follows from Corollary 2.2. By using the symbol, more precise statements about the location of the spectrum can be made; cf. Theorem 2.2.

Many results on eigenvalues and eigenvectors of symmetric Toeplitz matrices follow from the persymmetry of these matrices; see [6]. Persymmetry also is employed in [31]. Related properties are shown in [4]. We conclude this section with some properties of eigenvalues and eigenvectors for skew-symmetric and normal Toeplitz matrices. The proofs explicitly utilize persymmetry.

Theorem 2.3 *Let T be a shifted skew-symmetric Toeplitz matrix of order $2m$, i.e.,*

$$T = \delta I + A,$$

where $\delta \in \mathbf{C}$ and A is a real skew-symmetric Toeplitz matrix. If $\delta+\lambda$ is an eigenvalue of T associated with the eigenvector v, then $\delta-\lambda$ is an eigenvalue associated with the eigenvector Jv.

Proof The proof follows easily from [31, Theorem 3.5]. We have

$$A = \begin{bmatrix} A_1 & -A_2^T \\ A_2 & A_1 \end{bmatrix} \in \mathbf{R}^{2m\times 2m},$$

where A_1 is skew-symmetric, and A_1 and A_2 are persymmetric. The eigenvalues λ of A are purely imaginary and occur in complex conjugate pairs. Moreover, if λ is an eigenvalue of A associated with the eigenvector v, then $-\lambda$ is an eigenvalue associated with the eigenvector Jv. These claims follow by fairly straightforward computations or from results in [31].

We turn to normal banded Toeplitz matrices.

Theorem 2.4 *Let $T_{(k)} \in \mathbf{C}^{2m\times 2m}$ be a normal $(2k+1)$-banded Toeplitz matrix, with $k \le m$. Then*

$$T_{(k)} = \delta I + e^{i\theta/2} S_{(k)}, \qquad S_{(k)} = (2m;k;\sigma,0,\overline{\sigma}) = \begin{bmatrix} S_1 & S_2^H \\ S_2 & S_1 \end{bmatrix}, \tag{2.5}$$

where $\delta \in \mathbf{C}$, $\theta \in \mathbf{R}$, S_1 is Hermitian, and S_1 and S_2 are persymmetric. Assume that $[(Jw)^T, v^T]^T$ is an eigenvector of $T_{(k)}$ associated with the eigenvalue $\delta + e^{i\theta/2}\lambda$. Then $x = v + \overline{w}$ satisfies the equation

$$(S_1 - \lambda I)x + S_2 J\overline{x} = 0. \tag{2.6}$$

Conversely, if x satisfies this equation, then $[(J\overline{x})^T, x^T]^T$ is an eigenvector of $T_{(k)}$ associated with the eigenvalue $\delta + e^{i\theta/2}\lambda$.

Proof The representation (2.5) is a consequence of Corollary 2.1. The relation between eigenvectors of $S_{(k)}$, which also are eigenvectors of $T_{(k)}$, and solutions of (2.6) follows by fairly straightforward computations; see [31, Theorem 3.6] for details.

3 Structured distance to normality

Theorem 3.1 *Let the* $(2k+1)$*-banded Toeplitz matrix* $T_{(k)} = (n;k;\sigma,\delta,\tau)$*, with* $k \leq \lfloor n/2 \rfloor$*, satisfy*

$$\sum_{h=1}^{k} (n-h)\sigma_h \tau_h \neq 0. \tag{3.1}$$

Then

$$T^*_{(k)} = (n;k;\sigma^*,\delta,\tau^*),$$

with

$$\sigma^* = \frac{\sigma + \overline{\tau} e^{i\theta^*}}{2}, \qquad \tau^* = \frac{\tau + \overline{\sigma} e^{i\theta^*}}{2}, \qquad \theta^* = \arg(\sum_{h=1}^{k} (n-h)\sigma_h \tau_h), \tag{3.2}$$

is the unique closest matrix, in the Frobenius norm, to $T_{(k)}$ *in* $\mathscr{N}_{\mathscr{T}_{(k)}}$*. Moreover, if* $(\sigma,\tau) \neq (0,0)$ *and (3.1) is violated, then there are infinitely many matrices* $T^*_{(k)}$*, depending on an arbitrary angle* $\theta \in \mathbf{R}$*, in* $\mathscr{N}_{\mathscr{T}_{(k)}}$ *at the same minimal distance from* $T_{(k)}$*, namely*

$$T^*_{(k)} = T^*_{(k)}(\theta) = (n;k;\frac{\sigma + \overline{\tau} e^{i\theta}}{2},\delta,\frac{\tau + \overline{\sigma} e^{i\theta}}{2}).$$

Proof It suffices to determine the closest matrix $T^*_{0,(k)} = (n;k;\sigma^*,0,\tau^*) \in \mathscr{N}_{\mathscr{T}_{(k)}}$ to the matrix $T_{0,(k)} = (n;k;\sigma,0,\tau)$. It follows from Theorem 2.1 that there is an angle θ, such that $\sigma^* = \overline{\tau}^* e^{i\theta}$. Substitute $\tau^* = z$ and $\sigma^* = \overline{z} e^{i\theta}$ into $T^*_{0,(k)}$. Then the squared distance

$$D(z,\theta) = \| T^*_{0,(k)} - T_{0,(k)} \|_F^2$$

can be expressed as

$$\begin{aligned} D(z,\theta) &= \sum_{h=1}^{k} (n-h)(|z-\tau_h|^2 + |\overline{z} e^{i\theta} - \sigma_h|^2) \\ &= \sum_{h=1}^{k} (n-h)(|z-\tau_h|^2 + |z - \overline{\sigma}_h e^{i\theta}|^2). \end{aligned}$$

We seek to determine a vector τ^* and an angle θ^*, such that

$$D(\tau^*,\theta^*) = \min_{\substack{z \in \mathbf{C}^k \\ -\pi < \theta \leq \pi}} D(z,\theta).$$

Since $z \rightarrow D(z,\theta)$ is convex and $\nabla_z D(z,\theta) = 0$ for

$$z = z(\theta) = \frac{\tau + \overline{\sigma} e^{i\theta}}{2},$$

this vector minimizes $D(z,\theta)$ for a given angle θ.

The desired value of $\theta \in \mathbf{R}$ is obtained by minimizing $d(\theta) = D(z(\theta), \theta)$. We have

$$\begin{aligned} d(\theta) &= \sum_{h=1}^{k} \frac{n-h}{2} |\tau_h - \overline{\sigma}_h e^{i\theta}|^2 \\ &= \frac{1}{2} \sum_{h=1}^{k} (n-h)(|\sigma_h|^2 + |\tau_h|^2) - \mathrm{Re}(e^{-i\theta} \sum_{h=1}^{k} (n-h)\sigma_h \tau_h). \end{aligned}$$

If (3.1) holds, then $d'(\theta) = 0$ if and only if

$$e^{2i\theta} = \frac{\sum_{h=1}^{k} (n-h)\sigma_h \tau_h}{\sum_{h=1}^{k} (n-h)\overline{\sigma_h \tau_h}}.$$

This yields $\theta^* = \arg(\sum_{h=1}^{k} (n-h)\sigma_h \tau_h)$ and the minimal value

$$d(\theta^*) = \frac{1}{2} \sum_{h=1}^{k} (n-h)(|\sigma_h|^2 + |\tau_h|^2) - \left| \sum_{h=1}^{k} (n-h)\sigma_h \tau_h \right| \tag{3.3}$$

of $d(\theta)$, which establishes (3.2).

We turn to the situation when (3.1) is violated. Then

$$d(\theta) = \frac{1}{2} \|T_{0,(k)}\|_F^2$$

for all values of $\theta \in \mathbf{R}$. It follows that there are infinitely many normal matrices $T^*_{0,(k)} = T^*_{0,(k)}(\theta)$, defined by

$$\sigma^* = \sigma^*(\theta) = \frac{\sigma + \overline{\tau} e^{i\theta}}{2}, \qquad \tau^* = \tau^*(\theta) = \frac{\tau + \overline{\sigma} e^{i\theta}}{2}, \qquad \theta \in \mathbf{R},$$

at the same minimal distance from $T_{0,(k)}$.

Corollary 3.1 *Let $T_{(k)} = (n;k;\sigma,\delta,\tau) \in \mathbf{R}^{n\times n}$ with $k \leq \lfloor n/2 \rfloor$. If the sum (3.1) is positive, then the projection of $T_{(k)}$ onto $\mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}}$ is the real symmetric $(2k+1)$-banded Toeplitz matrix*

$$T^*_{1,(k)} = (n;k;\frac{\sigma+\tau}{2},\delta,\frac{\sigma+\tau}{2}).$$

If, instead, the sum (3.1) is negative, then the projection of $T_{(k)}$ onto $\mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}}$ is the real shifted skew-symmetric $(2k+1)$-banded Toeplitz matrix

$$T^*_{2,(k)} = (n;k;\frac{\sigma-\tau}{2},\delta,\frac{\tau-\sigma}{2}).$$

Finally, if the sum (3.1) vanishes, then both matrices $T^*_{1,(k)}$ and $T^*_{2,(k)}$ are closest matrices to $T_{(k)}$ in $\mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}}$ in the Frobenius norm.

Corollary 3.2 *The structured distance to normality of a* $(2k+1)$*-banded Toeplitz matrix* $T_{(k)} = (n;k;\sigma,\delta,\tau)$*, with* $k \le \lfloor n/2 \rfloor$*, in the Frobenius norm, is equal to*

$$\Delta_F(T_{(k)}) = \sqrt{\frac{1}{2}\sum_{h=1}^{k}(n-h)(|\sigma_h|^2+|\tau_h|^2) - \left|\sum_{h=1}^{k}(n-h)\sigma_h\tau_h\right|}.$$

Proof The Frobenius norm of the difference between $T_{(k)}$ and the matrix $T^*_{(k)}$ determined by Theorem 3.1 gives the distance to $\mathscr{N}_{\mathscr{T}_{(k)}}$ of $T_{(k)}$; the distance is the square root of the quantity in (3.3). The distance reduces to $\frac{1}{\sqrt{2}}\|T_{0,(k)}\|_F$ if $\sum_{h=1}^{k}(n-h)\sigma_h\tau_h = 0$.

Remark 3.1 *Note that, though the matrices* $T_{(k)}$ *and* $T_{(k)} - \delta I$ *have the same structured distance to normality, they have different projections onto* $\mathscr{N}_{\mathscr{T}_{(k)}}$ *(at distance* $\sqrt{n}|\delta|$*).*

Corollary 3.3 *The structured distance to normality of a real* $(2k+1)$*-banded Toeplitz matrix* $T_{(k)} = (n;k;\sigma,\delta,\tau)$*, with* $k \le \lfloor n/2 \rfloor$*, in the Frobenius norm, is equal to*

$$\Delta_F^{\mathbf{R}}(T_{(k)}) = \sqrt{\frac{1}{2}\min\left\{\sum_{j=1}^{k}(n-j)(\sigma_j-\tau_j)^2, \sum_{j=1}^{k}(n-j)(\sigma_j+\tau_j)^2\right\}}.$$

This section has provided simple formulas for determining the closest normal Toeplitz band matrix, with suitably restricted bandwidth, to a given Toeplitz matrix of the same bandwidth. It is straightforward to extend these formulas to determine the closest normal banded Toeplitz matrix, with suitably restricted bandwidth, to an arbitrary matrix. We illustrate this and discuss an application. Introduce the $(2k+1)$-banded matrix

$$A_{(k)} = \begin{bmatrix} a_{0,0} & a_{0,1} & a_{0,2} & \cdots & a_{0,k} & & \mathbf{0} \\ a_{1,0} & a_{1,1} & a_{1,2} & & & \ddots & \\ a_{2,0} & a_{2,1} & & & & & \\ \vdots & & & \ddots & & & a_{n-k,n-1} \\ a_{k,0} & & & & & & \vdots \\ & \ddots & & & & a_{n-2,n-2} & a_{n-2,n-1} \\ \mathbf{0} & & a_{n-1,n-k} & \cdots & & a_{n-1,n-2} & a_{n-1,n-1} \end{bmatrix} \in \mathbf{R}^{n\times n}.$$

Then the closest symmetric matrix $T^*_{(k)} = (n;k;\sigma^*,\delta^*,\tau^*) \in \mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}}$ is given by

$$\delta^* = \frac{1}{n}\sum_{j=0}^{n-1} a_{j,j}, \quad \tau_h^* = \frac{1}{2(n-h)}\sum_{j=0}^{n-1-h}(a_{j,j+h}+a_{j+h,j}), \quad \sigma_h^* = \tau_h^*, \quad h = 1:k.$$

Similarly, the closest matrix $T^*_{(k)} \in \mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}}$ with $T^*_{(k)} - \delta^* I$ skew-symmetric is determined by

$$\tau^*_h = \frac{1}{2(n-h)} \sum_{j=0}^{n-1-h} (a_{j,j+h} - a_{j+h,j}), \quad \sigma^*_h = -\tau^*_h, \quad h = 1:k.$$

Remark 3.2 *The approximation of matrix functions $f(A)$, where f is a nonlinear function, such as the exponential function, a logarithm, or the square root, and $A \in \mathbf{C}^{n\times n}$, has many applications; see, e.g., [20] for a discussion. Numerical methods and error bounds for large-scale problems have recently been discussed in [3]. These bounds and some of the methods require a set $\mathscr{S}$ containing the field of values*

$$\mathscr{W}(A) = \left\{ \frac{x^H A x}{x^H x}, x \in \mathbf{C}^n \right\}$$

of A to be explicitly known; see, e.g., [21] for a thorough treatment of properties of the field of values.

When $A = A_{(k)} \in \mathbf{R}^{n\times n}$, such a set easily can be determined as follows. Let $T^*_{(k)} = (n;k;\sigma^*,\delta^*,\tau^*) \in \mathscr{N}^{\mathbf{R}}_{\mathscr{T}_{(k)}}$ be the closest matrix to $A_{(k)}$. Assume for ease of discussion that $T^*_{(k)}$ is symmetric. Then, in view of Theorem 2.2, we have $\mathscr{W}(T^*_{(k)}) \subset \delta^* + [m,M]$, where m and M are computed with the aid of the symbol for $T^*_{(k)}$. Moreover,

$$\mathscr{W}(A_{(k)}) \subset \mathscr{W}(T^*_{(k)}) + \mathscr{W}(A_{(k)} - T^*_{(k)}),$$
$$\mathscr{W}(A_{(k)} - T^*_{(k)}) \subset \{z \in \mathbf{C} : |z| \le \|A_{(k)} - T^*_{(k)}\|_F\}.$$

The evaluation of $\|A_{(k)} - T^*_{(k)}\|_F$ is straightforward. It follows that a rectangular region $\mathscr{S}$ containing $\mathscr{W}(A_{(k)})$, with the interval $\delta + [m,M]$ in its interior, can be computed with little effort.

Analogous formulas for sets containing $\mathscr{W}(A_{(k)})$ can be derived when $A_{(k)} \in \mathbf{C}^{n\times n}$. The determination of the closest matrix $T^*_{(k)} \in \mathscr{N}_{\mathscr{T}_{(k)}}$ to $A_{(k)}$ now requires the evaluation of an angle θ^*; cf. Theorem 2.1. This can be done similarly as in Theorem 3.1. A rectangular set $\mathscr{S}$, which contains $\mathscr{W}(A_{(k)})$, now can be determined in the same manner as above. We omit the details.

We conclude this section with a comment on the (unstructured) distance to normality from $A \in \mathbf{C}^{n\times n}$,

$$d_F(A) = \min_{N \in \mathscr{N}} \|A - N\|_F.$$

Let $B = (n;1;0,0,1)$ be a Jordan block of order n associated with the eigenvalue zero. Then it follows from László [25, Theorem 2] and [29, Example 9.1] that $d_F(B)/\|B\|_F = 1/\sqrt{n}$. Equality is achieved for a circulant approximant of B. The structured distance considered in this paper, which does not include circulants, is much larger for this example; we have $\Delta_F(B)/\|B\|_F = 1/\sqrt{2}$; see [29] for details. Thus, for a general matrix $A \in \mathbf{C}^{n\times n}$, there may be matrices in $\mathscr{N}$, which are much closer to A than the closest banded normal Toeplitz matrix. An attraction of the normal banded Toeplitz matrices considered in this paper, when compared to general normal matrices, is that their spectra live on line segments that easily can be determined; cf. Theorem 2.2.

Acknowledgements SN would like to thank Lionello Pasquini and Paolo Buttà for valuable discussions and comments. The authors would like to thank Mark Embree and the referees for comments.

References

1. R. M. Beam and R. F. Warming, The asymptotic spectra of banded Toeplitz and quasi-Toeplitz matrices, SIAM J. Sci. Comput., 14, 971–1006 (1993)
2. B. Beckermann and L. Reichel, The Arnoldi process and GMRES for nearly symmetric matrices, SIAM J. Matrix Anal. Appl., 30, 102–120 (2008),
3. B. Beckermann and L. Reichel, Error estimation and evaluation of matrix functions via the Faber transform, submitted for publication
4. D. Bini and M. Capovani, Spectral and computational properties of band symmetric Toeplitz matrices, Linear Algebra Appl., 52/53, 99–126 (1983)
5. A. Böttcher and S. Grudsky, Spectral Properties of Banded Toeplitz Matrices, SIAM, Philadelphia (2005)
6. A. Cantoni and P. Butler, Eigenvalues and eigenvectors of symmetric centrosymmetric matrices, Linear Algebra Appl., 13, 275–288 (1976)
7. P. J. Davis, Circulant Matrices, 2nd ed., Chelsea, New York (1994)
8. J. W. Demmel, On condition numbers and the distance to the nearest ill-posed problem, Numer. Math., 51, 251–289 (1987)
9. ———, 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
10. L. Elsner and Kh. D. Ikramov, Normal matrices: an update, Linear Algebra Appl., 285, 291–303 (1998)
11. L. Elsner and M. H. C. Paardekooper, On measures of nonnormality of matrices, Linear Algebra Appl., 92, 107–124 (1987)
12. V. Faber and T. Manteuffel, Necessary and sufficient conditions for the existence of a conjugate gradient method, SIAM J. Numer. Anal., 21, 352–362 (1984)
13. D. R. Farenick, M. Krupnik, N. Krupnik, and W. Y. Lee, Normal Toeplitz matrices, SIAM J. Matrix Anal. Appl., 17, 1037–1043 (1996)
14. S. Friedland, Normal matrices and the completion problem, SIAM J. Matrix Anal. Appl., 23, 896–902 (2002)
15. U. Grenander and G. Szegő, Toeplitz Forms and Their Applications, Chelsea, New York (1984)
16. R. Grone, C. R. Johnson, E. M. Sa, and H. Wolkowicz, Normal matrices, Linear Algebra Appl., 87, 213–225 (1987)
17. C. Gu and L. Patton, Commutation relations for Toeplitz and Hankel matrices, SIAM J. Matrix Anal. Appl., 24, 728–746 (2003)
18. P. Henrici, Bounds for iterates, inverses, spectral variation and field of values of non-normal matrices, Numer. Math., 4, 24–40 (1962)
19. 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.
20. N. J. Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelphia (2008)
21. R. A. Horn and C. R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge (1991)
22. I. Ipsen, Computing an eigenvector with inverse iteration, SIAM Rev., 39, 254–291 (1997)
23. T. Ito, Every normal Toeplitz matrix is either of type I or of type II, SIAM J. Matrix Anal. Appl., 17, 998–1006 (1996)
24. T. Kailath (ed.), Modern Signal Processing, Hemisphere Publishing, Washington (1985)
25. L. László, An attainable lower bound for the best normal approximation, SIAM J. Matrix Anal. Appl., 15, 1035–1043 (1994)
26. S. L. Lee, A practical upper bound for departure from normality, SIAM J. Matrix Anal. Appl., 16, 462–468 (1995)
27. ———, Best available bounds for departure from normality, SIAM J. Matrix Anal. Appl., 17, 984–991 (1996)
28. A. Luati and T. Proietti, On the spectral properties of matrices associated with trend filters, 2008, MPRA Paper No. 11502, `http://mpra.ub.uni-muenchen.de/11502`

29. S. Noschese, L. Pasquini, and L. Reichel, The structured distance to normality of an irreducible real tridiagonal matrix, Electron. Trans. Numer. Anal., 28, 65–77 (2007)
30. L. Reichel and L. N. Trefethen, Eigenvalues and pseudo-eigenvalues of Toeplitz matrices, Linear Algebra Appl., 162–164, 153–185 (1992)
31. R. M. Reid, Some eigenvalue properties of persymmetric matrices, SIAM Rev., 39, 313–316 (1997)
32. A. Ruhe, On the closeness of eigenvalues and singular values for almost normal matrices, Linear Algebra Appl., 11, 87–94 (1975)
33. ———, Closest normal matrix finally found!, BIT, 27, 585–598 (1987)
34. L. N. Trefethen and M. Embree, Spectra and Pseudospectra, Princeton University Press, Princeton (2005)