Electronic Transactions on Numerical Analysis.
Volume 28, pp. 65-77, 2007.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

THE STRUCTURED DISTANCE TO NORMALITY OF AN IRREDUCIBLE REAL TRIDIAGONAL MATRIX*

S. NOSCHESE†, L. PASQUINI†, AND L. REICHEL‡

Dedicated to Gene Golub on the occasion of his 75th birthday

Abstract. The problem of computing the distance in the Frobenius norm of a given real irreducible tridiagonal matrix T to the algebraic variety of real normal irreducible tridiagonal matrices is solved. Simple formulas for computing the distance and a normal tridiagonal matrix at this distance are presented. The special case of tridiagonal Toeplitz matrices also is considered.

Key words. matrix nearness problem, distance to normality, real tridiagonal matrix, eigenvalue conditioning, Toeplitz matrix

AMS subject classifications. 65F30, 65F50, 15A57, 65F35

1. Introduction. Matrix nearness problems have received considerable attention in the literature; see, e.g., [3, 4, 12, 13, 19] and references therein. It is the purpose of the present paper to investigate the structured distance of a real irreducible tridiagonal matrix to the algebraic variety of real normal irreducible tridiagonal matrices, which we denote by $\mathcal{I}$. We present a simple formula for determining this distance measured in the Frobenius norm. Moreover, given a real irreducible tridiagonal matrix T of distance d from the set $\mathcal{I}$, we provide formulas for computing a real normal tridiagonal matrix of distance d to T. The latter formulas are easy to evaluate. The special case when the tridiagonal matrix is of Toeplitz form also is considered. The simplicity of our formulas contrasts with the rather cumbersome task of determining the closest normal matrix of a general square matrix; a Jacobi-type algorithm for the latter endeavor has been described by Ruhe [20].

Many authors have contributed to our understanding of the distance to normality; see, e.g., [6, 7, 12, 14, 15, 16, 19, 20]. A large number of characteristic properties of normal matrices can be found in [5, 11]. The distance to normality is of interest because the eigenvalue problem is perfectly conditioned for normal matrices; a small distance to normality may make it feasible to replace the given matrix by a closest real normal tridiagonal matrix and compute the eigenvalues of the latter. This replacement can be attractive because in the setting of the present paper, the closest normal matrices are symmetric or shifted skew-symmetric, and fast reliable algorithms are available for the computation of eigenvalues of these types of matrices.

Gene Golub has made many significant contributions to matrix nearness problems and to the development of algorithms for structured eigenvalue problems, including the Golub-Kahan and Golub-Welsch algorithms; see, e.g., [1, 2, 8, 9, 10, 22].

This paper is organized as follows. Section 2 introduces notation used throughout the paper, section 3 presents an upper bound for the distance in the Frobenius norm of a real tridiagonal matrix to the algebraic variety of real normal tridiagonal matrices. A characterization of the real normal tridiagonal matrices is given in section 4, and a formula for the

*Received March 9, 2007. Accepted for publication April 20, 2007. Recommended by L. N. Trefethen. The work of the first and second author was supported by a grant from Università degli Studi di Roma “La Sapienza”. The work of the third author was supported in part by an OBR Research Challenge Grant.

†Dipartimento di Matematica “Guido Castelnuovo”, Università degli Studi di Roma “La Sapienza”, P.le A. Moro, 2, I-00185 Roma, Italy ({noschese, lionello.pasquini}@uniroma1.it).

‡Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA (reichel@math.kent.edu).

ETNA
Kent State University
etna@mcs.kent.edu

distance in the Frobenius norm of a tridiagonal matrix to the set of symmetric tridiagonal matrices or the set of shifted skew-symmetric tridiagonal matrices is provided in section 5. The latter formula is a key result of the paper, since we show in section 6 that this formula also yields the distance to the set of real normal irreducible tridiagonal matrices. Section 7 considers the special case of tridiagonal Toeplitz matrices, section 8 reviews eigenvalue conditioning, and section 9 presents a few computed examples. The examples indicate that for interesting classes of real irreducible matrices T close to the algebraic variety of real normal irreducible tridiagonal matrices, the eigenvalue problem for T is well conditioned. Finally, section 10 contains concluding remarks.

TABLE 2.1
Definitions of sets used in the paper.

$\mathcal{N}$	the algebraic variety of the normal real matrices in $\mathbf{R}^{n\times n}$
$\mathcal{S}$	the subspace of $\mathcal{N}$ formed by the symmetric matrices
$\mathcal{A}$	the subspace of $\mathcal{N}$ formed by the antisymmetric matrices
$\mathcal{A}^+$	the subspace of $\mathcal{N}$ formed by the shifted antisymmetric matrices
$\mathcal{T}$	the subspace of $\mathbf{R}^{n\times n}$ formed by the real tridiagonal matrices
$\mathcal{N}_{\mathcal{T}}$	$\mathcal{N}\cap\mathcal{T}$
$\mathcal{S}_{\mathcal{T}}$	$\mathcal{S}\cap\mathcal{T}$
$\mathcal{A}_{\mathcal{T}}$	$\mathcal{A}\cap\mathcal{T}$
$\mathcal{A}^+_{\mathcal{T}}$	$\mathcal{A}^+\cap\mathcal{T}$
$\mathcal{I}$	the subset of $\mathcal{N}_{\mathcal{T}}$ formed by the irreducible matrices

2. Notation. This section defines notation used in the sequel. We let $T = (n;\sigma,\delta,\tau)$ denote the real tridiagonal matrix

$$T = \begin{bmatrix} \delta_1 & \tau_1 & & & & \mathbf{0} \\ \sigma_1 & \delta_2 & \tau_2 & & & \\ & \sigma_2 & \delta_3 & \ddots & & \\ & & \ddots & \ddots & \tau_{n-2} & \\ & & & \sigma_{n-2} & \delta_{n-1} & \tau_{n-1} \\ \mathbf{0} & & & & \sigma_{n-1} & \delta_n \end{bmatrix} \in \mathbf{R}^{n\times n} \tag{2.1}$$

and introduce the inner product

$$(A,B) = \text{trace}\,(B^TA), \qquad A,B\in\mathbf{R}^{n\times n},$$

which induces the Frobenius norm

$$\|A\|_F = (A,A)^{1/2}, \qquad A\in\mathbf{R}^{n\times n}.$$

Let $\mathcal{X}$ denote a subset of the set of real normal matrices. We define the distance of a matrix A to this subset by

$$d_F(A,\mathcal{X}) = \inf\{\|E\|_F : A+E\in\mathcal{X}\}.$$

Further notation used in the paper is summarized in Table 2.1. We use the terminology antisymmetric synonymously with skew-symmetric.

3. An upper bound for $d_F(T, \mathcal{N}_\mathcal{T})$. We first present some auxiliary results used to determine the desired bound. The following result is well known and easy to prove.

PROPOSITION 3.1. *Let $\mathcal{S}$ be the subspace of $\mathbf{R}^{n\times n}$ formed by the symmetric matrices and let $\mathcal{A}$ be the subspace of $\mathbf{R}^{n\times n}$ formed by the antisymmetric matrices. Then $\mathcal{S}$ and $\mathcal{A}$ are orthogonal.*

It is worth noticing that the matrices $\frac{1}{2}(A + A^T)$ and $\frac{1}{2}(A - A^T)$ are the projections of A onto $\mathcal{S}$ along $\mathcal{A}$, and of A onto $\mathcal{A}$ along $\mathcal{S}$, respectively.

THEOREM 3.2.

$$d_F(A, \mathcal{N}) \leq \min\{d_F(A_0, \mathcal{S}), d_F(A_0, \mathcal{A})\} \leq \frac{1}{\sqrt{2}}\|A_0\|_F, \qquad \forall A \in \mathbf{R}^{n\times n}, \tag{3.1}$$

where $A_0 = A - \frac{1}{n}\mathrm{trace}\,(A)I$, and I denotes the identity matrix.

Proof. We first note that $d_F(A, \mathcal{N}) = d_F(A + cI, \mathcal{N})$ for any $c \in \mathbf{R}$. In particular, $d_F(A, \mathcal{N}) = d_F(A_0, \mathcal{N})$. The inequality

$$d_F(A_0, \mathcal{N}) \leq \min\{d_F(A_0, \mathcal{S}), d_F(A_0, \mathcal{A})\}$$

follows from the possibility of the existence of normal matrices closer to A_0 than the projections of A_0 onto $\mathcal{S}$ and $\mathcal{A}$. Since $\mathcal{S}$ and $\mathcal{A}$ are orthogonal, a geometric argument shows the right-hand side inequality in (3.1). ◻

REMARK 3.1. The inequalities (3.1) also hold with A_0 replaced by A everywhere. The matrix A_0 satisfies

$$\|A_0\|_F = \min_{c\in\mathbf{R}} \|A + cI\|_F.$$

REMARK 3.2. A simple argument shows that equality in the right-hand side inequality in (3.1) is achieved if and only if $A_0 = [a_{ij}^{(0)}]$ satisfies

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^{(0)} a_{ji}^{(0)} = 0. \tag{3.2}$$

COROLLARY 3.3.

$$d_F(T, \mathcal{N}_\mathcal{T}) \leq \min\{d_F(T_0, \mathcal{S}_\mathcal{T}), d_F(T_0, \mathcal{A}_\mathcal{T})\} \leq \frac{1}{\sqrt{2}}\|T_0\|_F, \qquad \forall T \in \mathcal{T}, \tag{3.3}$$

where

$$T_0 = T - \frac{1}{n}\mathrm{trace}\,(T)I. \tag{3.4}$$

Proof. The result follows immediately from Theorem 3.2 and the observation that the projections of a tridiagonal matrix onto $\mathcal{S}$ and $\mathcal{A}$ also are tridiagonal matrices. ◻

REMARK 3.3. For real tridiagonal matrices $T = (n; \sigma, \delta, \tau)$, the condition (3.2) simplifies to

$$\sum_{i=1}^{n}(\delta_i - s)^2 + 2\sum_{i=1}^{n-1}\sigma_i\tau_i = 0, \tag{3.5}$$

where $s = \frac{1}{n}\sum_{j=1}^{n}\delta_j$, i.e., we obtain equality in the right-hand side inequality of (3.3) if and only if (3.5) holds. Equality in the left-hand side bound in (3.3) also can be achieved; see Example 9.1 of section 9 below.

ETNA
Kent State University
etna@mcs.kent.edu

4. Real normal tridiagonal matrices.

THEOREM 4.1. *A real tridiagonal matrix is normal if and only if it is a direct sum of symmetric and shifted skew-symmetric matrices.*

Proof. The statement of the theorem can be expressed in the form that a real tridiagonal matrix $T = (n; \sigma, \delta, \tau)$ is normal if and only if it is block diagonal, with each block either a diagonal block or an irreducible tridiagonal block $\hat{T} = (\nu; \hat{\sigma}, \hat{\delta}, \hat{\tau})$, whose entries satisfy one of the following conditions:

i) $\hat{\sigma}_h = \hat{\tau}_h, \quad h = 1 : \nu - 1,$
ii) $\hat{\sigma}_h = -\hat{\tau}_h, \quad h = 1 : \nu - 1 \quad \text{and} \quad \hat{\delta}_1 = \hat{\delta}_2 = \ldots = \hat{\delta}_\nu.$

We first note that tridiagonal matrices T that satisfy the above conditions are normal. Conversely, assume that T is normal. Then

$$T^T T = T T^T. \tag{4.1}$$

Both the right-hand side and left-hand side matrices are pentadiagonal and symmetric. Therefore (4.1) is equivalent to the conditions

$$\begin{array}{rcll} (T^T T)_{h,h} & = & (T T^T)_{h,h}, & h = 1 : n, \\ (T^T T)_{h+1,h} & = & (T T^T)_{h+1,h}, & h = 1 : n-1, \\ (T^T T)_{h+2,h} & = & (T T^T)_{h+2,h}, & h = 1 : n-2. \end{array}$$

These conditions imply, in order,

$$\tau_h^2 = \sigma_h^2, \qquad h = 1 : n-1, \tag{4.2}$$

$$\tau_h \delta_h + \sigma_h \delta_{h+1} = \tau_h \delta_{h+1} + \sigma_h \delta_h, \qquad h = 1 : n-1, \tag{4.3}$$

$$\tau_{h+1} \sigma_h = \tau_h \sigma_{h+1}, \qquad h = 1 : n-2. \tag{4.4}$$

If T is irreducible, then (4.2) and (4.4) lead to either

$$\sigma_h = \tau_h, \qquad h = 1 : n-1,$$

or

$$\sigma_h = -\tau_h, \qquad h = 1 : n-1. \tag{4.5}$$

When (4.5) holds, equation (4.3) yields $\delta_1 = \delta_2 = \ldots = \delta_n$.

If T is reducible, then T may have diagonal blocks, which are diagonal and therefore normal. The above discussion on irreducible matrices applies to the remaining tridiagonal blocks on the diagonal (if any). This concludes the proof. ∎

COROLLARY 4.2. *A real normal tridiagonal matrix can be partitioned into diagonal blocks that are either diagonal or tridiagonal and normal.*

5. The distances to $\mathcal{S}_\mathcal{T}$ and $\mathcal{A}_\mathcal{T}^+$. It is easy to see that, given a real tridiagonal matrix $T = (n; \sigma, \delta, \tau)$, the closest matrix in $\mathcal{S}$ in the Frobenius norm is the matrix $T^{(s)} = (n; \sigma^{(s)}, \delta^{(s)}, \tau^{(s)})$ with entries

$$\begin{array}{rcll} \sigma_h^{(s)} & = & \tau_h^{(s)} = \dfrac{\sigma_h + \tau_h}{2}, & h = 1 : n-1, \\ \delta_h^{(s)} & = & \delta_h, & h = 1 : n. \end{array}$$

ETNA
Kent State University
etna@mcs.kent.edu

It follows that the distance in the Frobenius norm between T and $\mathcal{S}$ is given by

$$d_F(T,\mathcal{S}) = \sqrt{\frac{1}{2}\sum_{i=1}^{n-1}(\sigma_i-\tau_i)^2}.$$

Similarly, the closest matrix to T in $\mathcal{A}^+$ in the Frobenius norm is the matrix $T^{(a)} = (n;\sigma^{(a)},\delta^{(a)},\tau^{(a)})$ with entries

$$\begin{aligned} \sigma_h^{(a)} &= -\tau_h^{(a)} = \frac{\sigma_h-\tau_h}{2}, && h=1:n-1,\\ \delta_h^{(a)} &= \frac{1}{n}\sum_{i=1}^{n}\delta_i, && h=1:n. \end{aligned}$$

It follows that the distance in the Frobenius norm between $T^{(a)}$ and $\mathcal{A}^+$ is given by

$$d_F(T,\mathcal{A}^+) = \sqrt{\frac{1}{2}\sum_{i=1}^{n-1}(\sigma_i+\tau_i)^2+\sum_{i=1}^{n}\left(\delta_i-\frac{\sum_{j=1}^{n}\delta_j}{n}\right)^2}.$$

Let $\mathcal{M}=\mathcal{S}_{\mathcal{T}}\cup\mathcal{A}^+_{\mathcal{T}}$. Then

$$d_F(T,\mathcal{M}) = \min\left\{d_F(T,\mathcal{S}_{\mathcal{T}}),d_F(T,\mathcal{A}^+_{\mathcal{T}})\right\}. \tag{5.1}$$

The following theorem provides upper and lower bounds for (5.1).

THEOREM 5.1. *For every real tridiagonal matrix $T=(n;\sigma,\delta,\tau)$, one has*

$$\begin{aligned} d_F(T,\mathcal{N}) &\leq d_F(T,\mathcal{N}_{\mathcal{T}}) \leq d_F(T,\mathcal{M}) = d_F(T_0,\mathcal{M})\\ &= \min\{d_F(T_0,\mathcal{S}_{\mathcal{T}}),d_F(T_0,\mathcal{A}_{\mathcal{T}})\} \leq \tfrac{1}{\sqrt{2}}\|T_0\|_F, \end{aligned} \tag{5.2}$$

where T_0 is defined by (3.4). The upper bound for $d_F(T,\mathcal{M})$ is attained when (3.5) holds. Then

$$\frac{d_F(T,\mathcal{M})}{\|T_0\|_F} = \frac{1}{\sqrt{2}}. \tag{5.3}$$

Proof. The first two inequalities of (5.2) follow from the inclusions $\mathcal{N}\supset\mathcal{N}_{\mathcal{T}}\supset\mathcal{M}$. The distance to $\mathcal{M}$ is invariant under addition of a multiple of the identity matrix. This gives the first equality. It follows from $\mathrm{trace}\,(T_0)=0$ that the matrix $T_0^{(a)}$ has a vanishing diagonal and therefore lives in $\mathcal{A}_{\mathcal{T}}$. The second equality follows from the definition of $\mathcal{M}$. The last inequality of (5.2) follows from (3.3). This inequality is achieved when (3.5) holds; cf. Remark 3.3. This shows (5.3). ∎

6. The distance $d_F(T,\mathcal{I})$. We are in a position to discuss the computation of the distance of a matrix $T\in\mathcal{T}$ to the set $\mathcal{I}$ of real normal irreducible tridiagonal matrices. The following theorem reduces this problem to the determination of the distance $d_F(T,\mathcal{M})$, which already has been discussed.

THEOREM 6.1.

$$d_F(T,\mathcal{I}) = d_F(T,\mathcal{M}), \qquad \forall T\in\mathcal{T}.$$

Proof. We first note that the set $\mathcal{M}$ is closed, because it is the union of the closed sets $\mathcal{S}_{\mathcal{T}}$ and $\mathcal{A}_{\mathcal{T}}^{+}$. By Theorem 4.1, a real irreducible normal tridiagonal matrix belongs to $\mathcal{M}$. Thus, $\mathcal{I} \subset \mathcal{M}$. Moreover, in every neighborhood of any matrix T in $\mathcal{S}_{\mathcal{T}}$ [in $\mathcal{A}_{\mathcal{T}}^{+}$] there is a real normal irreducible tridiagonal matrix $T^* \neq T$ in $\mathcal{S}_{\mathcal{T}}$ [in $\mathcal{A}_{\mathcal{T}}^{+}$]. Thus, $\mathcal{M}$ is the closure of $\mathcal{I}$, and (6.1) follows. □

REMARK 6.1. Let $T = (n; \sigma, \delta, \tau)$. When $\sigma_i^2 \neq \tau_i^2$, $i = 1 : n-1$, the closest matrix to T in $\mathcal{M}$, $T^{(s)}$ or $T^{(a)}$, is irreducible. Otherwise, at least one of the matrices $T^{(s)}$ and $T^{(a)}$ is reducible and, hence, the closest matrix to T in $\mathcal{M}$ may be reducible.

REMARK 6.2. Theorem 6.1 is applicable to any matrix $T \in \mathcal{T}$, also reducible ones. However, note that $d_F(T, \mathcal{I}) = 0$ does not imply that T is irreducible; it just indicates that there is a real irreducible normal tridiagonal matrix in every open neighborhood of T. For instance, $d_F(T, \mathcal{I}) = 0$ when T is the zero matrix.

REMARK 6.3. There are matrices $T \in \mathcal{T}$, such that $d_F(T, \mathcal{I}) \gg d_F(T, \mathcal{N})$. Indeed, there are matrices $T \in \mathcal{T}$, such that $d_F(T, \mathcal{I}) \gg d_F(T, \mathcal{N}_{\mathcal{T}})$; see Example 9.2 in section 9 below.

7. Tridiagonal Toeplitz matrices. This section is concerned with real tridiagonal Toeplitz matrices $T = (n; \sigma, \delta, \tau)$, i.e.,

$$T = \begin{bmatrix} \delta & \tau & & & & & \mathbf{0} \\ \sigma & \delta & \tau & & & & \\ & \sigma & \delta & & & & \\ & & & \ddots & & & \\ & & \ddots & \ddots & \tau & & \\ & & & \sigma & \delta & \tau \\ \mathbf{0} & & & & \sigma & \delta \end{bmatrix} \in \mathbf{R}^{n \times n}. \tag{7.1}$$

THEOREM 7.1. *The real tridiagonal Toeplitz matrix (7.1) is normal if and only if its entries satisfy $\sigma = \tau$ or $\sigma = -\tau$.*

Proof. The result follows from Theorem 4.1. □

REMARK 7.1. Notice that Theorem 7.1 implies that a real normal tridiagonal Toeplitz matrix is reducible if and only if it is diagonal.

The following results are consequences of (5.1) and the discussion leading up to that result.

THEOREM 7.2. *Let T be a real tridiagonal Toeplitz matrix (7.1). The closest real tridiagonal matrix $T^{(s)}$ to T in the set $\mathcal{S}_{\mathcal{T}}$ is a Toeplitz matrix with diagonal entries δ, and sub- and super-diagonal entries $\frac{1}{2}(\sigma + \tau)$. The closest real tridiagonal matrix $T^{(a)}$ to T in the set $\mathcal{A}_{\mathcal{T}}^{+}$ is a Toeplitz matrix with sub-diagonal entries $\frac{1}{2}(\sigma - \tau)$, diagonal entries δ, and super-diagonal entries $-\frac{1}{2}(\sigma - \tau)$. Moreover,*

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

COROLLARY 7.3. *Let T be a real irreducible tridiagonal Toeplitz matrix. Then the closest matrix in $\mathcal{M}$ is irreducible. Moreover,*

$$d_F(T, \mathcal{M}) = d_F(T, \mathcal{N}_{\mathcal{T}}). \tag{7.2}$$

ETNA
Kent State University
etna@mcs.kent.edu

Proof. Let $T^* = (n; \sigma^*, \delta^*, \tau^*)$ denote the closest matrix in $\mathcal{M}$ to T. Since T is assumed to be irreducible, it follows from Theorem 7.2 that T^* is an irreducible normal Toeplitz matrix. Thus, it suffices to show that there is no normal reducible tridiagonal matrix $T' = (n; \sigma', \delta', \tau')$ closer to T than T^*. Since T is a Toeplitz matrix, T' must belong to $\mathcal{M}$. Any pair of vanishing entries $\sigma'_h = \tau'_h = 0$ gives a contribution to the distance $\|T - T'\|_F$ that is larger than the contribution of the pair of entries $\{\sigma^*, \tau^*\}$ of T^* to the distance $\|T - T^*\|_F$. This completes the proof. $\square$

We finally consider the situation when the given real tridiagonal matrix is not Toeplitz, and we wish to determine the closest real normal Toeplitz matrix.

THEOREM 7.4. *Let $T = (n; \sigma, \delta, \tau)$ be a real tridiagonal matrix (2.1). The closest tridiagonal Toeplitz matrix $T^{(s)}$ to T in the set $\mathcal{S}_T$ has diagonal entries δ, and sub- and super-diagonal entries σ and τ, respectively, given by*

$$\delta = \frac{1}{n}\sum_{i=1}^{n} \delta_i, \qquad \sigma = \tau = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(\sigma_i + \tau_i). \tag{7.3}$$

The closest tridiagonal Toeplitz matrix $T^{(a)}$ to T in the set $\mathcal{A}_T^+$ has diagonal entries δ given by (7.3), and sub- and super-diagonal entries σ and τ, respectively, given by

$$\sigma = -\tau = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(\sigma_i - \tau_i).$$

Proof. It follows from Theorem 4.1 that the desired matrix belongs to $\mathcal{M}$ and is of Toeplitz form. This yields the equations for the entries δ, σ, and τ. $\square$

8. Eigenvalue condition numbers. We define the eigenvalue condition number for nondefective irreducible tridiagonal matrices. Let $T \in \mathbf{R}^{n\times n}$ be such a matrix, and let x_j and y_j denote right and left eigenvectors of unit length, respectively, associated with the eigenvalue λ_j. Following [10, 21], we define the condition number for the eigenvalue λ_j by

$$\kappa(\lambda_j) = |y_j^* x_j|^{-1},$$

where the superscript $*$ denotes transposition and complex conjugation. The eigenvalue condition number for T is defined by

$$\kappa(T) = \max_{1\le j\le n} \kappa(\lambda_j). \tag{8.1}$$

9. Examples. This section presents a few examples that illustrate some properties of the structured distance to normality.

Example 9.1. Consider the quasi-Jordan block

$$J = \begin{bmatrix} \lambda & \mu & 0 & \dots & \dots & 0 & 0 \\ 0 & \lambda & \mu & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & \mu & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & \mu & 0 \\ 0 & 0 & 0 & \dots & \dots & \lambda & \mu \\ 0 & 0 & 0 & \dots & \dots & \dots & \lambda \end{bmatrix} \in \mathbf{R}^{n\times n}, \qquad \text{with } \mu \neq 0.$$

ETNA
Kent State University
etna@mcs.kent.edu

Thus, J is an upper bidiagonal Toeplitz matrix. The circulant matrix

$$N = \begin{bmatrix} \lambda & \frac{n-1}{n}\mu & 0 & \dots & \dots & 0 & 0 \\ 0 & \lambda & \frac{n-1}{n}\mu & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & \frac{n-1}{n}\mu & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & \frac{n-1}{n}\mu & 0 \\ 0 & 0 & 0 & \dots & \dots & \lambda & \frac{n-1}{n}\mu \\ \frac{n-1}{n}\mu & 0 & 0 & \dots & \dots & \dots & \lambda \end{bmatrix} \in \mathbf{R}^{n\times n}$$

is normal. An easy computation yields

$$\frac{d_F(J,\mathcal{N})}{\|J_0\|_F} \leq \frac{\|J-N\|_F}{\|J_0\|_F} = \frac{1}{\sqrt{n}}, \tag{9.1}$$

where $J_0 = J - \lambda I$.

The following irreducible $n \times n$ matrices,

$$J^{(s)} = \begin{bmatrix} \lambda & \mu/2 & 0 & \dots & \dots & 0 & 0 \\ \mu/2 & \lambda & \mu/2 & 0 & \dots & 0 & 0 \\ 0 & \mu/2 & \lambda & \mu/2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \mu/2 & \lambda & \mu/2 & 0 \\ 0 & 0 & 0 & \dots & \mu/2 & \lambda & \mu/2 \\ 0 & 0 & 0 & \dots & \dots & \mu/2 & \lambda \end{bmatrix},$$

$$J^{(a)} = \begin{bmatrix} \lambda & \mu/2 & 0 & \dots & \dots & 0 & 0 \\ -\mu/2 & \lambda & \mu/2 & 0 & \dots & 0 & 0 \\ 0 & -\mu/2 & \lambda & \mu/2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & -\mu/2 & \lambda & \mu/2 & 0 \\ 0 & 0 & 0 & \dots & -\mu/2 & \lambda & \mu/2 \\ 0 & 0 & 0 & \dots & \dots & -\mu/2 & \lambda \end{bmatrix},$$

are the closest matrices to J in $\mathcal{S}_\mathcal{T}$ and $\mathcal{A}^+_\mathcal{T}$, respectively. They are equidistant to J, and we obtain from (5.1) and Theorem 6.1 that

$$\frac{d_F(J,\mathcal{I})}{\|J_0\|_F} = \frac{d_F(J,\mathcal{M})}{\|J_0\|_F} = \frac{1}{\sqrt{2}}. \tag{9.2}$$

The eigenvalues of a Jordan block are sensitive to perturbations of the matrix, while eigenvalues of normal matrices are not. We would like our measure of the distance to normality to reflect this fact, i.e., Jordan blocks should be distant from the set of normal matrices considered. Indeed, the normalized structured distance (9.2) is maximal; cf. Theorem 3.2. We also note that the normalized unstructured distance (9.1) is not; the latter distance decreases to zero as the size n of the Jordan block increases. The normalized unstructured distance therefore is a poor indicator of the conditioning of the eigenvalue problem.

We conclude this example by noting that

$$d_F(J,\mathcal{N}_\mathcal{T}) = d_F(J,\mathcal{I}), \tag{9.3}$$

ETNA
Kent State University
etna@mcs.kent.edu

from which it follows that

$$\frac{d_F(J,\mathcal{N}_{\mathcal{T}})}{\|J_0\|_F} = \frac{1}{\sqrt{2}}.$$

The equality (9.3) can be shown similarly as (7.2). Thus, we need to show that there is no real normal reducible tridiagonal matrix $T = (n;\sigma,\delta,\tau)$ closer to J than $J^{(s)}$ and $J^{(a)}$. If there were such a matrix, then, according to Theorem 4.1, the condition $\sigma_h = \tau_h = 0$ would necessarily hold for at least one value of h. However, such a pair of entries would give a larger contribution to the distance of J to $\mathcal{N}_{\mathcal{T}}$ than the pair $\{\sigma_h, \tau_h\} = \{\pm\mu/2, \mu/2\}$.

Example 9.2. The above example illustrates that $d_F(T,\mathcal{N}_{\mathcal{T}}) = d_F(T,\mathcal{M})$ for certain $T \in \mathcal{T}$. This example shows that when a tridiagonal matrix $T = (n;\sigma,\delta,\tau)$ has a pair of "tiny" off-diagonal entries σ_h and τ_h, then $d_F(T,\mathcal{N}_{\mathcal{T}})$ and $d_F(T,\mathcal{M})$ may differ significantly. Thus, let

$$T = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & \varepsilon & 0 \\ 0 & \varepsilon & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

with $\varepsilon > 0$ a tiny parameter. Then

$$\frac{d_F(T,\mathcal{M})}{||T_0||_F} = \frac{1}{\sqrt{2+\varepsilon^2}}, \qquad \frac{d_F(T,\mathcal{N}_{\mathcal{T}})}{||T_0||_F} \le \frac{\varepsilon}{\sqrt{2+\varepsilon^2}},$$

where we obtain the inequality by setting $\varepsilon = 0$ in T. Thus, the normalized distance to the set $\mathcal{M}$ is close to maximal for $\varepsilon > 0$ small, while the normalized distance to $\mathcal{N}_{\mathcal{T}}$ is small in this situation. This example indicates that the computed normalized distances to the set $\mathcal{M}$ may be most useful for tridiagonal matrices with no tiny off-diagonal pairs.

The following numerical examples have been carried out in MATLAB with about 16 significant decimal digits.

Example 9.3. We consider tridiagonal matrices T, whose eigenvalues are zeros of generalized Bessel polynomials. These polynomials depend on two parameters a and $b \neq 0$. The entries of T are given by

$$\delta_1 = -\frac{a}{b}, \qquad \tau_1 = -\delta_1, \qquad \sigma_1 = \frac{\delta_1}{a+1},$$

and, for $j \geq 2$,

$$\begin{aligned} \delta_j &= -b\,\frac{a-2}{(2j+a-2)(2j+a-4)}, \\ \tau_j &= b\,\frac{j+a-2}{(2j+a-2)(2j+a-3)}, \\ \sigma_j &= -b\,\frac{j}{(2j+a-1)(2j+a-2)}; \end{aligned}$$

see [18] for a recent discussion on generalized Bessel polynomials, their applications, and the computation of their zeros.

For the tridiagonal matrix T of order 30 with $a = -17/2$ and $b = 2$, we obtain

$$\frac{d_F(T,\mathcal{N})}{\|T_0\|} \le 0.6835, \qquad \frac{d_F(T,\mathcal{I})}{\|T_0\|} = 0.7071,$$

ETNA
Kent State University
etna@mcs.kent.edu

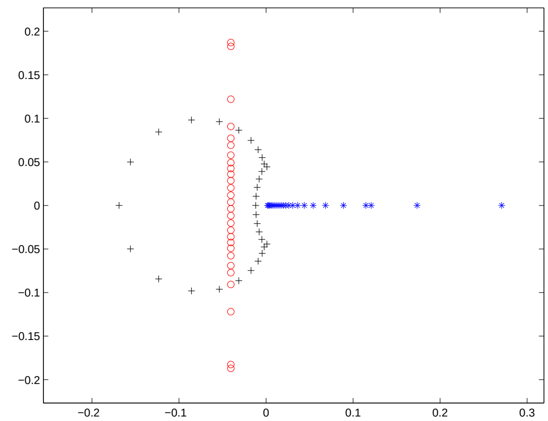

FIG. 9.1. *Example 9.3: The computed zeros of a generalized Bessel polynomial of degree* 30 *close to the origin (black* $+$*), eigenvalues of* $T^{(s)}$ *close to the origin (blue* $*$*), and eigenvalues of* $T^{(a)}$ *close to the origin (red* o*).*

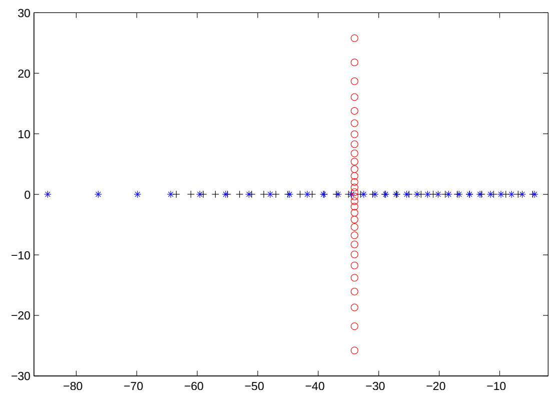

FIG. 9.2. *Example 9.4: The eigenvalues of the tridiagonal matrix* $T \in \mathbf{R}^{30\times 30}$ *defined by (9.4) (black* $+$*), eigenvalues of* $T^{(s)}$ *(blue* $*$*), and eigenvalues of* $T^{(a)}$ *(red* o*).*

where T_0 is given by (3.4). Thus, the normalized structured distance to normality is very close to its maximal value $1/\sqrt{2}$. An upper bounds for the unstructured distance $d_F(T, \mathcal{N})$ in this and the following examples is computed by the method described by Ruhe [20]; for many matrices this method yields $d_F(T, \mathcal{N})$. In the present example, the computed bound for the normalized unstructured distance is nearly as large as the normalized structured distance.

The eigenvalue condition number (8.1) is very large; we have $\kappa(T) = 1.4 \cdot 10^{13}$. Thus, in this example the unstructured and structured distances to normality are large, and so is the eigenvalue condition number. Figure 9.1 shows the computed eigenvalues of T, $T^{(s)}$, and $T^{(a)}$ closest to the origin.

Example 9.4. Consider the tridiagonal matrix $T = (30; \sigma, \delta, \tau)$ defined by

$$\delta_j = -3 - 2j, \qquad \tau_j = j + 1, \qquad \sigma_j = \frac{1}{j+1}, \qquad j \geq 1. \tag{9.4}$$

This matrix is discussed in [17]. We obtain

$$\frac{d_F(T, \mathcal{N})}{\|T_0\|} \leq 0.23, \qquad \frac{d_F(T, \mathcal{I})}{\|T_0\|} = 0.50,$$

ETNA
Kent State University
etna@mcs.kent.edu

and $\kappa(T) = 6.7 \cdot 10^7$, where T_0 is given by (3.4). The unstructured distance to normality is fairly small, but the eigenvalue condition number is large. This example illustrates, similarly as Example 9.1, that the unstructured distance to normality can be a poor measure of the conditioning of the eigenvalue problem. The structured distance to normality is large enough to indicate that the eigenvalues may be sensitive. Figure 9.2 shows the eigenvalues of T, $T^{(s)}$, and $T^{(a)}$.

Example 9.5. We compute the structured and unstructured distances to normality for three Toeplitz matrices. The tridiagonal Toeplitz matrix $T = (30; \sigma, \delta, \tau)$, defined by

$$\delta = 0, \qquad \tau = \frac{1}{4}, \qquad \sigma = 1, \tag{9.5}$$

yields

$$\frac{d_F(T, \mathcal{N})}{\|T_0\|} \leq 0.17, \qquad \frac{d_F(T, \mathcal{I})}{\|T_0\|} = 0.51.$$

The eigenvalue condition number is given by $\kappa(T) = 3.7 \cdot 10^7$. Figure 9.3 shows the eigenvalues of T, $T^{(s)}$, and $T^{(a)}$. The spectrum of T is seen to be real.

The tridiagonal Toeplitz matrix $T = (30; \sigma, \delta, \tau)$ determined by

$$\delta = 1, \qquad \tau = \frac{9}{10}, \qquad \sigma = -\frac{1}{10}, \tag{9.6}$$

satisfies

$$\frac{d_F(T, \mathcal{N})}{\|T_0\|} \leq 0.18, \qquad \frac{d_F(T, \mathcal{I})}{\|T_0\|} = 0.62,$$

and has the eigenvalue condition number $\kappa(T) = 4.5 \cdot 10^{12}$. Figure 9.4 shows the eigenvalues of T, $T^{(s)}$, and $T^{(a)}$. All eigenvalues of T and $T^{(a)}$ have real part one.

The unstructured distances to normality for the matrices (9.5) and (9.6) are small, but the structured distances to normality are fairly large, and so are the eigenvalue condition numbers. Moreover, the structured distance to normality is larger for the matrix with the largest eigenvalue condition number.

For the tridiagonal Toeplitz matrix $T = (30; \sigma, \delta, \tau)$ defined by

$$\delta = 1, \qquad \tau = \frac{9}{10}, \qquad \sigma = -\frac{11}{10}, \tag{9.7}$$

we obtain the distances

$$\frac{d_F(T, \mathcal{N})}{\|T_0\|} \leq 8.3 \cdot 10^{-2}, \qquad \frac{d_F(T, \mathcal{I})}{\|T_0\|} = 1.0 \cdot 10^{-1},$$

and the eigenvalue condition number $\kappa(T) = 3.6$. Figure 9.5 shows the eigenvalues of T, $T^{(s)}$, and $T^{(a)}$. The eigenvalues of T are real and difficult to distinguish from the eigenvalues of $T^{(s)}$. In some applications, it therefore may suffice to compute the eigenvalues of the symmetric tridiagonal matrix $T^{(s)}$ instead of the eigenvalues of the nonnormal matrix T. This can be attractive since there are fast accurate algorithms available for the computation of the eigenvalues of $T^{(s)}$.

10. Conclusion. The structured distance to normality for real irreducible tridiagonal matrices is easy to compute. Numerous computed examples suggest that for many matrices a small structured distance to normality implies a small to moderate eigenvalue condition number. If the matrix does not have pairs of tiny off-diagonal entries, such as in Example 9.2, then a large structured distance to normality generally indicates that the eigenvalue condition number is large. Further analysis that sheds light on these observations is required.

ETNA
Kent State University
etna@mcs.kent.edu

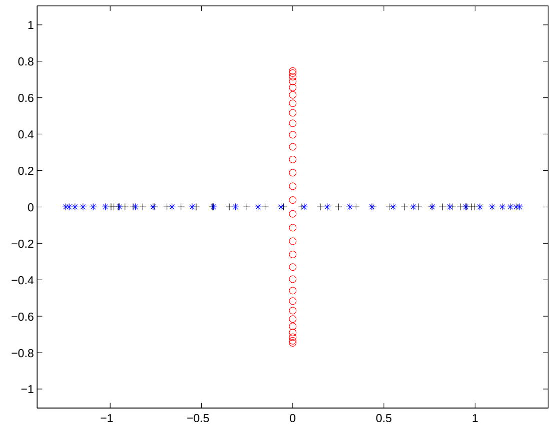

FIG. 9.3. *Example 9.5: The eigenvalues of the tridiagonal Toeplitz matrix $T \in \mathbf{R}^{30\times 30}$ defined by (9.5) (black $+$), eigenvalues of $T^{(s)}$ (blue $*$), and eigenvalues of $T^{(a)}$ (red o).*

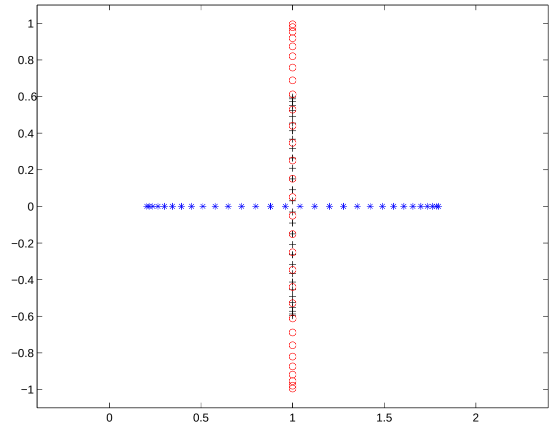

FIG. 9.4. *Example 9.5: The eigenvalues of the tridiagonal Toeplitz matrix $T \in \mathbf{R}^{30\times 30}$ defined by (9.6) (black $+$), eigenvalues of $T^{(s)}$ (blue $*$), and eigenvalues of $T^{(a)}$ by (red o).*

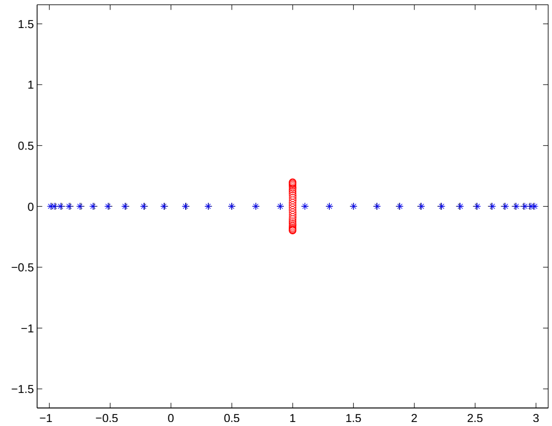

FIG. 9.5. *Example 9.5: The eigenvalues of the tridiagonal Toeplitz matrix $T \in \mathbf{R}^{30\times 30}$ defined by (9.7) (black $+$), eigenvalues of $T^{(s)}$ (blue $*$), and eigenvalues of $T^{(a)}$ by (red o). The eigenvalues of T and $T^{(s)}$ are too close to distinguish.*

Acknowledgement. LR would like to thank Volker Mehrmann for an enjoyable visit to TU Berlin, during which work on the present paper was carried out.

REFERENCES

[1] P. ARBENZ AND G. H. GOLUB, *On the spectral decomposition of Hermitian matrices modified by low rank perturbations with applications*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 40–58.
[2] ———, *Matrix shapes invariant under the symmetric QR algorithm*, Numer. Linear Algebra Appl., 2 (1995), pp. 87–93.
[3] J. W. DEMMEL, *On condition numbers and the distance to the nearest ill–posed problem*, Numer. Math., 51 (1987), pp. 251–289.
[4] ———, *Nearest defective matrices and the geometry of ill–conditioning*, in Reliable Numerical Computation, M. G. Cox and S. Hammarling, eds., Clarendon Press, Oxford, England, 1990, pp. 35–55.
[5] L. ELSNER AND K. D. IKRAMOV, *Normal matrices: an update*, Linear Algebra Appl., 285 (1998), pp. 291–303.
[6] L. ELSNER AND M. H. C. PAARDEKOOPER, *On measures of nonnormality of matrices*, Linear Algebra Appl., 92 (1987), pp. 107–124.
[7] R. GABRIEL, *The normal ΔH-matrices with connection to some Jacobi-like methods*, Linear Algebra Appl., 91 (1987), pp. 181–194.
[8] G. H. GOLUB AND W. KAHAN, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM J. Numer. Anal., 2, Ser. B, (1965), pp. 205–224.
[9] G. H. GOLUB AND J. H. WELSCH, *Calculation of Gauss quadrature rules*, Math. Comp., 23 (1969), pp. 221–230.
[10] 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.
[11] R. GRONE, C. R. JOHNSON, E. M. SA, AND H. WOLKOWICZ, *Normal matrices*, Linear Algebra Appl., 87 (1987), pp. 213–225.
[12] P. HENRICI, *Bounds for iterates, inverses, spectral variation and field of values of non–normal matrices*, Numer. Math., 4 (1962), pp. 24–40.
[13] 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, England, 1989, pp. 1–27.
[14] L. LÁSZLÓ, *An attainable lower bound for the best normal approximation*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 1035–1043.
[15] S. L. LEE, *A practical upper bound for departure from normality*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 462–468.
[16] ———, *Best available bounds for departure from normality*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 984–991.
[17] H. W. J. LENFERINK AND M. N. SPIJKER, *On the use of stability regions in the numerical analysis of initial value problems*, Math. Comp., 57 (1991), pp. 221–237.
[18] L. PASQUINI, *Accurate computation of the zeros of the generalized Bessel polynomials*, Numer. Math., 86 (2000), pp. 507–538.
[19] A. RUHE, *On the closeness of eigenvalues and singular values for almost normal matrices*, Linear Algebra Appl., 11 (1975), pp. 87–94.
[20] ———, *Closest normal matrix finally found!*, BIT, 27 (1987), pp. 585–598.
[21] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, England, 1965.
[22] T. ZHANG, G. H. GOLUB, AND K. H. LAW, *Eigenvalue perturbation and generalized Krylov subspace method*, Appl. Numer. Math., 27 (1998), pp. 185–202.