Bounds on Distances between Eigenvalues*

Moshe Haviv
Department of Statistics
The Hebrew University
Jerusalem 91905, Israel

and

Uriel G. Rothblum
Faculty of Industrial Engineering and Management
Technion-Israel Institute of Technology
Haifa 32000, Israel

Submitted by Hans Schneider

ABSTRACT

Explicit (computable) lower and upper bounds on the distances between a given real eigenvalue of a real square matrix and the remaining (not necessarily real) eigenvalues of the matrix are developed.

1. INTRODUCTION

The purpose of this paper is to develop lower and upper bounds on the distances between a given real eigenvalue of a real square matrix and the remaining (not necessarily real) eigenvalues of the matrix. Specifically, let A be a real square matrix, and let λ be a real eigenvalue of A. We identify a ring in the complex plane having the form $\{\xi \in C: s \leqslant |\xi - \lambda| \leqslant S\}$ where $0 \leqslant s \leqslant S$, such that all the eigenvalues of A which are different from λ are included in it. We next motivate the interest in such bounds. The bounds s and S we obtain depend only on the eigenvalue λ and the corresponding eigenvector.

Let A and λ be as above. Resolvent expansions of the matrix A at λ are power-series expansions of the inverse of $(A - \xi I)^{-1}$ in the variable $\xi - \lambda$, where $\xi \neq \lambda$ is a complex number. Evidently, if $|\xi - \lambda|$ is large, we have from

*This research was supported by NSF Grant ENG-78-25182 to Yale University.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/84/$3.00

the Neuman expansion that

$$\begin{aligned}(A-\xi I)^{-1} &= [(A-\lambda I)-(\xi-\lambda)I]^{-1}\\ &= -(\xi-\lambda)^{-1}\left[I-(\xi-\lambda)^{-1}(A-\lambda I)\right]^{-1}\\ &= -\sum_{j=0}^{\infty}(\xi-\lambda)^{-j-1}(A-\lambda I)^{j}. \end{aligned}\tag{1.1}$$

Also, if $|\xi-\lambda|$ is small with $\xi\neq\lambda$, we have that (e.g., [9, Theorem 3.1])

$$\begin{aligned}(A-\xi I)^{-1} &= [(A-\lambda I)-(\xi-\lambda)I]^{-1}\\ &= -\sum_{j=-\nu}^{-1}(\xi-\lambda)^{j}(A-\lambda I)^{-j-1}E\\ &\quad+\sum_{j=0}^{\infty}(\xi-\lambda)^{j}\left[(A-\lambda I)^{D}\right]^{j+1}, \end{aligned}\tag{1.2}$$

where E is the eigenprojection of A at λ, ν is the index of λ for A, and $(A-\lambda I)^{D}$ is the Drazin inverse of $A-\lambda I$ (formal definitions will be given in Section 2). There are many applications of resolvent expansions of matrices (e.g., [3] or [9]). It will be shown in Lemma 2.5, below, that the right-hand side of (1.1) converges, in which case it equals the left-hand side of (1.1), if and only if $|\xi-\lambda|$ is greater than the maximal distance between λ and an eigenvalue of A which is different from λ. It is also shown in Lemma 2.5 that the right-hand side of (1.2) converges, in which case it equals the left-hand side of (1.2), if and only if $|\xi-\lambda|$ is smaller than the minimal distance between λ and an eigenvalue of A which is different from λ.

An additional motivation for interest in the minimal distance between the eigenvalues of a real square matrix A and a distinguished real eigenvalue λ occurs when the matrix A is stochastic. It is well known that if A is a stochastic matrix, one is an eigenvalue of A and the multiplicity of this eigenvalue equals the number of ergodic classes of A (e.g., [5]). It now follows from continuity arguments that if $\{A_n\}$ is a sequence of stochastic matrices, each having a single ergodic class, and if $\lim_{n\to\infty}A_n$ exists and the limiting matrix is a (stochastic) matrix having more than a single ergodic class, then there is a sequence $\{\lambda_n\}$, where for each $n=1,2,\ldots,$ $\lambda_n\neq 1$ and λ_n is an eigenvalue of A_n, with $\lim_{n\to\infty}\lambda_n=1$. In particular, $\min\{|\mu-1|:\mu$ is an eigenvalue of A_n and $\mu\neq 1\}$ approaches zero as $n\to\infty$. Call a stochastic matrix *completely decomposable* if it has more than a single ergodic class. We conclude from the above that the smaller the minimal distance between the eigenvalue one of a stochastic matrix A and the other eigenvalues of A is, the "more completely decomposable" A is. Example 6.2 below illustrates the use of the above measure for "nearly complete decomposability." For further discussion on "nearly completely decomposable stochastic matrices" and their eigenvalues see [12, Chapter 4].

The organization of this paper is as follows. Preliminary results and notational conventions are listed in Section 2. General bounds on the distances between a given eigenvalue of a matrix and its remaining eigenvalues are given in Section 3. Section 4 contains some explicit bounds which are easy to compute, and Section 5 contains variants of the general bounds (obtained in Section 3) for the cases where the given eigenvalue is simple. Finally, we list a few numerical examples in Section 6.

2. NOTATIONAL CONVENTIONS AND PRELIMINARY RESULTS

The *real* and *complex* field will be denoted by R and C, respectively. The *coordinates* of a (complex) matrix or vector will be denoted by subscripts, e.g., B_{ij} or b_i. The *null space* and *range space* of a complex $n \times n$ matrix B will be denoted by null B and range B, respectively, i.e., null $B = \{x \in C^n : Bx = 0\}$ and range $B = \{Bx : x \in C^n\}$.

The vector $(1,1,\dots,1)^T \in R^n$ will be denoted by e, and the vector $(0,0,\dots,0)^T \in R^n$ will be denoted by 0. Also, for $i = 1,\dots,n$, the ith *unit vector*, denoted e^i, is the vector in R^n whose coordinates are all zero except for the ith coordinate, which is one. (As usual, these conventions do not have explicit reference to n.)

Let $B \in C^{n \times n}$. The *spectrum* of B, denoted $\sigma(B)$, is the set of all eigenvalues of B. The *spectral radius* of B, denoted $\rho(B)$, is the maximal modulus of an eigenvalue of B, i.e., $\rho(B) = \max\{|\lambda| : \lambda \in \sigma(B)\}$. Let $\mu \in C$, and let $Q_\mu \equiv B - \mu I$. The *index of* μ *for* B, denoted $\nu_\mu(B)$, is the smallest integer $m \geqslant 0$ for which the null spaces of $(Q_\mu)^m$ and $(Q_\mu)^{m+1}$ coincide. Of course, $\mu \in \sigma(B)$ if and only if $\nu_\mu(B) > 0$. Evidently, $\operatorname{null}(Q_\mu)^m = \operatorname{null}(Q_\mu)^\nu$ for $\nu \equiv \nu_\mu(B)$ if and only if $m \geqslant \nu$. Also, the null set$(Q_\mu)^\nu$ for $\nu \equiv \nu_\mu(B)$ is called the *algebraic eigenspace* of B at μ. An eigenvalue μ of B is called *simple* if the dimension of the algebraic eigenspace of B at μ is one.

The following lemma is well known (e.g., [9, Lemma 2.1]).

LEMMA 2.1. *Let Q be a complex $n \times n$ matrix, let ν be the index of zero for Q, and let $\mathscr{R}$ and $\mathscr{N}$ be, respectively, the range and null spaces of Q^ν. Then*:

(1) $C^n = \mathscr{R} \oplus \mathscr{N}$.

(2) *There is a projection E on $\mathscr{N}$ along $\mathscr{R}$.*

(3) $Q^{\nu}E = EQ^{\nu} = 0$, $Q^{\nu-1}E = EQ^{\nu-1} \neq 0$.
(4) *E and Q commute.*
(5) *$Q - E$ is nonsingular.*
(6) *$(Q - E)^{-1}$ and E commute.*
(7) *$DE = ED = 0$, where $D \equiv (Q - E)^{-1}(I - E) = (I - E)(Q - E)^{-1}$.*
(8) *$DQ = QD = I - E$.*
(9) *$D^{n+1}Q = QD^{n+1} = D^n$ for $n = 1, 2, \ldots$.*

For a given complex square matrix B and a complex number μ, the matrix E constructed in Lemma 2.1 for $Q \equiv B - \mu I$ is called the *eigenprojection of B at μ*. Also, it was shown in [6] that the matrix D constructed in Lemma 2.1 is the well-known Drazin inverse of Q, which will be denoted by Q^D (see, for example, [1] for the original definition of the Drazin inverse). One can easily verify that if B and μ are real, so are E and Q^D. Also, $(Q^T)^D = [Q^D]^T$.

The following lemma gives a representation of the Drazin inverse for complex square matrices for which zero is an eigenvalue having index 1.

LEMMA 2.2. *Let Q be a complex $n \times n$ matrix where zero is an eigenvalue of Q, and let ν and E be, respectively, the index of zero for Q and the eigenprojection of Q at zero. Then the following are equivalent:*

(1) $\nu = 1$.
(2) $Q^D = (Q - E)^{-1} + E$.

Proof. As $Q - E$ is invertible, and as $Q^D = (Q - E)^{-1}(I - E)$, we have that (2) holds if and only if $I - E = (Q - E)Q^D = (Q - E)[(Q - E)^{-1} + E] = I + (Q - E)E = I - E + QE$, or equivalently, if and only if $QE = 0$. As zero is an eigenvalue of Q, we have that $E \neq 0$. Hence, part (3) of Lemma 2.1 assures that $QE = 0$ if and only if $\nu = 1$, completing the proof. ■

The next lemma states the relationship between eigenvalues and corresponding eigenvectors of a complex square matrix and those of its Drazin inverse. It is key for our development.

LEMMA 2.3. *Let Q be a complex $n \times n$ matrix, and let zero be an eigenvalue of Q. Then*

$$\sigma(Q^D) = \{\mu^{-1} : 0 \neq \mu \in \sigma(Q)\} \cup \{0\}. \tag{2.1}$$

Moreover, the eigenvectors of Q and Q^D corresponding to the (common) eigenvalue zero coincide.

Proof. The result is given by Campbell and Meyer [1, p. 129]. ■

COROLLARY 2.4. *Let A be a complex square matrix, and let λ be an eigenvalue of A. Then $\rho[(A-\lambda I)^D] = \max\{|\mu-\lambda|^{-1} : \lambda \neq \mu \in \sigma(A)\}$, where the maximum over the empty set is defined to be zero.*

We end this section with the promised characterization of the complex numbers ξ for which (1.1) and (1.2) hold, respectively.

LEMMA 2.5. *Let A be a complex $n \times n$ matrix, let λ be an eigenvalue of A, and let ξ be a complex number. Then:*

(1) *The right-hand side of* (1.1) *converges if and only if* $|\xi-\lambda| > \max\{|\mu-\lambda| : \mu \in \sigma(A)\}$, *and in this case* (1.1) *holds.*

(2) *The right-hand side of* (1.2) *converges if and only if* $|\xi-\lambda| < \min\{|\mu-\lambda| : \lambda \neq \mu \in \sigma(A)\}$, *and in this case* (1.2) *holds.*

Proof. It is well known (e.g., [8 Corollary 4.4]) that for a square matrix B, $\sum_{i=0}^{\infty} B^i$ converges if and only if $\rho(B) < 1$, and in this case $I-B$ is invertible and $(I-B)^{-1} = \sum_{i=0}^{\infty} B^i$.

We conclude that the right-hand side of (1.1) converges if and only if $|\xi-\lambda|^{-1}\rho(A-\lambda I) = \rho[(\xi-\lambda)^{-1}(A-\lambda I)] < 1$. In this case $I-(\xi-\lambda)^{-1}(A-\lambda I) = (\xi-\lambda)^{-1}[(\xi-\lambda)I-(A-\lambda I)] = -(\xi-\lambda)^{-1}(A-\xi I)$ is invertible, and its inverse equals $\sum_{j=0}^{\infty}(\xi-\lambda)^{-j}(A-\lambda I)^j$; in particular, in this case (1.1) holds. Observing that $\rho(A-\lambda I) = \max\{|\mu-\lambda| : \mu \in \sigma(A)\}$, we have that $|\xi-\lambda|^{-1}\rho(A-\lambda I) < 1$ if and only if $|\xi-\lambda| > \max\{|\mu-\lambda| : \mu \in \sigma(A)\}$, establishing (1).

As above, we next conclude that the right-hand side of (1.2) converges if and only if $|\xi-\lambda|\rho[(A-\lambda)^D] = \rho[(\xi-\lambda)(A-\lambda I)^D] < 1$. In this case, arguments of [9, Theorem 3.1] show that (1.2) holds. Corollary 2.4 shows that $[\rho(A-\lambda I)^D]^{-1} = \min\{|\mu-\lambda| : \lambda \neq \mu \in \sigma(A)\}$, where the minimum over the empty set is defined to be $+\infty$. We therefore conclude that $|\xi-\lambda|\rho[(A-\lambda I)^D] < 1$ if and only if $|\xi-\lambda| < \rho[(A-\lambda I)^D]^{-1} = \min\{|\mu-\lambda| : \lambda \neq \mu \in \sigma(A)\}$, establishing (2). ■

3. THE MAIN RESULTS ON BOUNDS ON DISTANCES OF EIGENVALUES

In this section we present the promised bounds on the distances of the eigenvalues of a square matrix from a given eigenvalue.

Let $\|\ \|$ be a norm on R^n, and let $u \in R^n$. We define the coefficient $\tau^u_{\|\ \|}(B)$ for an $n \times n$ matrix B by

$$\tau^u_{\|\ \|}(B) \equiv \max_{\substack{\|x\| \leqslant 1 \\ x^T u = 0 \\ x \in R^n}} \|x^T B\|. \tag{3.1}$$

Of course, the above maximum is not taken over the empty set and is attained.

We are now ready for the main result of this paper. Part of the result [the last inequality in (3.2) below] follows directly from [11]. We include a full proof here for completeness.

THEOREM 3.1. *Let A be a real $n \times n$ matrix, let λ be a real eigenvalue of A, and let w be a corresponding real, right eigenvector. Also, let $\|\ \|$ be a norm on R^n. Then*

$$\left\{\tau^w_{\|\ \|}\left[(A - \lambda I)^D\right]\right\}^{-1} \leqslant \min_{\substack{\mu \in \sigma(A) \\ \mu \neq \lambda}} |\mu - \lambda| \leqslant \max_{\substack{\mu \in \sigma(A) \\ \mu \neq \lambda}} |\mu - \lambda| \leqslant \tau^w_{\|\ \|}(A - \lambda I). \tag{3.2}$$

Proof. We will first establish the last inequality in (3.2), namely, we will show that for every $\lambda \neq \mu \in \sigma(A)$,

$$|\mu - \lambda| \leqslant \tau^w_{\|\ \|}(A - \lambda I). \tag{3.3}$$

Consider the real-valued functional $\phi(\cdot)$ defined over C^n by

$$\phi(z) = \max_{\substack{\|x\| \leqslant 1 \\ x^T w = 0 \\ x \in R^n}} |x^T z|. \tag{3.4}$$

Of course, the above maximum is not taken over the empty set and is attained. Also, it is easy to verify that for every $z \in C^n$ and $\xi \in C$,

$$\phi(\xi z) = |\xi| \phi(z). \tag{3.5}$$

We will next obtain a bound on $\phi[(A - \lambda I)z]$. First observe that for every $x \in R^n$, $[x^T(A - \lambda I)]w = x^T[(A - \lambda I)w] = 0$ and therefore

$$|x^T[(A - \lambda I)z]| = |[x^T(A - \lambda I)]z| \leqslant \|x^T(A - \lambda I)\| \phi(z). \tag{3.6}$$

Taking the maximum of both sides of this inequality over $x \in R^n$ for which $\|x\| \leqslant 1$ and $x^T w = 0$ implies that

$$\phi[(A-\lambda I)z] \leqslant \left[\tau^{w}_{\|\ \|}(A-\lambda I)\right][\phi(z)]. \tag{3.7}$$

Next let $\mu \neq \lambda$ be an eigenvalue of A, and let $0 \neq z \in C^n$ be a corresponding right eigenvector. Then $Az = \mu z$, and (3.5) and (3.7) imply that

$$|\mu-\lambda|\phi(z) = \phi[(\mu-\lambda)z] = \phi[(A-\lambda I)z] \leqslant \left[\tau_{\|\ \|}(A-\lambda I)\right]\phi(z). \tag{3.8}$$

We will next show that $\phi(z) > 0$. This assertion coupled with (3.8) will immediately imply (3.3). As $\phi(z) \geqslant 0$, it suffices to show that $\phi(z) \neq 0$. Now, if $\phi(z) = 0$, then we have that

$$[x^T w = 0, \quad x \in R^n] \quad \Rightarrow \quad [x^T z = 0]. \tag{3.9}$$

Let $u = \mathrm{Re}(z)$ and $v = \mathrm{Im}(z)$ be the vectors in R^n whose components are the real and imaginary parts of z, respectively. We conclude from (3.9) that

$$[x^T w = 0, \quad x \in R^n] \quad \Rightarrow \quad [x^T u = 0, \quad x^T v = 0]. \tag{3.10}$$

Arguments from elementary linear algebra show that (3.10) implies that for some $\alpha, \beta \in R$, $u = \alpha w$ and $v = \beta w$. It follows that $z = u + iv = (\alpha + i\beta)w$ and therefore $(\mu - \lambda)z = (A - \lambda I)z = (A - \lambda I)(\alpha + i\beta)w = 0$. As $\lambda \neq \mu$ and $z \neq 0$, we get a contradiction to the assertion that $\phi(z) = 0$, completing the proof of (3.3).

Since the second inequality of (3.2) is trivial, it remains to prove the first inequality of (3.2). It follows from Lemma 2.3 that $\sigma[(A-\lambda I)^D] = \{(\mu - \lambda)^{-1} : \lambda \neq \mu \in \sigma(A)\} \cup \{0\}$ and that w is a (real) right eigenvector of $(A - \lambda I)^D$ corresponding to the eigenvalue zero. Hence, by applying the last inequality of (3.2) to $(A-\lambda I)^D$ and its eigenvalue zero with its real right eigenvector w, we conclude that for $\lambda \neq \mu \in \sigma(A)$

$$|\mu-\lambda|^{-1} \leqslant \tau^{w}_{\|\ \|}\left[(A-\lambda I)^D\right]; \tag{3.11}$$

and the first inequality of (3.2) follows directly. ■

The following two corollaries state modifications of Theorem 3.1, by applying that theorem to matrices having the same spectrum as the given matrix (cf., [11, Sections 4, 5]).

COROLLARY 3.2. *Let A be a real $n \times n$ matrix, let λ be a real eigenvalue of A, and let v be a corresponding real left eigenvector. Also, let $\|\ \|$ be a norm on R^n. Then*

$$\left\{\tau^{v}_{\|\ \|}\left[(A^T-\lambda I)^D\right]\right\}^{-1} \leqslant \min_{\substack{\mu\in\sigma(A)\\ \mu\neq\lambda}}|\mu-\lambda| \leqslant \max_{\substack{\mu\in\sigma(A)\\ \mu\neq\lambda}}|\mu-\lambda| \leqslant \tau^{v}_{\|\ \|}(A^T-\lambda I). \tag{3.12}$$

Proof. The corollary follows immediately by applying Theorem 3.1 to A^T and recalling the facts that $\sigma(A)=\sigma(A^T)$ and that v is a (real) right eigenvector of A^T corresponding to eigenvalue λ. ∎

COROLLARY 3.3. *Let A be a real $n \times n$ matrix, let λ be a real eigenvalue of A, and let w be a corresponding real right eigenvector. Also, let F be a real nonsingular $n \times n$ matrix, and let $\|\ \|$ be a norm on R^n. Then*

$$\left\{\tau^{F^{-1}w}_{\|\ \|}\left[F^{-1}(A-\lambda I)^D F\right]\right\}^{-1} \leqslant \min_{\substack{\mu\in\sigma(A)\\ \mu\neq\lambda}}|\mu-\lambda| \leqslant \max_{\substack{\mu\in\sigma(A)\\ \mu\neq\lambda}}|\mu-\lambda|$$

$$\leqslant \tau^{F^{-1}w}_{\|\ \|}\left[F^{-1}(A-\lambda I)F\right]. \tag{3.13}$$

Proof. The corollary follows immediately by applying Theorem 3.1 to $F^{-1}AF$ and recalling that $\sigma(A)=\sigma(F^{-1}AF)$, that $F^{-1}w$ is a (real) right eigenvector of $F^{-1}AF$ corresponding to eigenvalue λ, and that $(F^{-1}AF-\lambda I)^D = F^{-1}(A-\lambda I)^D F$. ∎

Examples of Rothblum and Tan [11, Section 8] demonstrate that the bounds obtained from Theorem 3.1 depend on the selection of the norm $\|\ \|$; that the bounds obtained from Corollary 3.2 do not necessarily coincide with those obtained from Theorem 3.1; and that the bounds obtained from Corollary 3.3 depend on the selection of the (nonsingular) matrix F. We remark that a useful selection of the (nonsingular) matrices in Corollary 3.3 are diagonal matrices (e.g., [11, Section 6]). In particular, one can select the nonsingular matrix F to be diagonal with $F_{ii}=w_i$ for $i=1,\dots,n$ with $w_i\neq 0$ and with $F_{ii}=1$ otherwise, assuring that the coordinates of $F^{-1}w$ consist only of zeros and ones. The usefulness of such selections will be demonstrated in Section 4. We also observe that additional corresponding bounds can be obtained through a combination of Corollaries 3.2 and 3.3 (applying Theorem 3.1 to $F^{-1}A^TF$, where F is a nonsingular matrix).

The bounds obtained in Theorem 3.1 and Corollaries 3.2 and 3.3 are all expressed in terms of the functionals $\tau^{\cdot}_{\cdot}(\cdot)$ which are defined, by (3.1), as the maximum of a real-valued convex function over a convex set. Thus, techniques of mathematical programming can be used to compute these bounds. For example, the corresponding maxima are always obtained at extreme points of the convex sets. In particular, explicit forms can be obtained when the number of corresponding extreme points is finite (e.g., Section 4).

We remark that the results of this section can be extended to complex matrices and corresponding complex eigenvalues by changing the definition of the τ coefficients in (3.1) and replacing the condition $x \in R^n$ by $x \in C^n$. The corresponding results are easier to prove, but we do not find them interesting, since the corresponding coefficients are not easy to compute (being optimal values of complex optimization problems). We remark that Zarling [12] obtained such a variant of Theorem 3.1 for a stochastic matrix, the eigenvalue $\lambda = 1$, and the l_1 norm (though he claims that his arguments apply with our definition of τ as a maximum of the corresponding functional over R^n and not C^n).

We finally note that Rothblum and Tan [11] discuss the selection of the norm for corresponding bounds. In particular, it follows that appropriate selections of the norms can get the bounds to be arbitrarily tight.

4. SOME EXPLICIT BOUNDS

Recall that for a vector $x \in R^n$ and $1 \leqslant p \leqslant \infty$, the l_p norm of x, denoted $\|x\|_p$, is given by $(\sum_{i=1}^n |x_i|^p)^{1/p}$ when $1 \leqslant p < \infty$ and by $\max_{1 \leqslant i \leqslant n}|x_i|$ when $p = \infty$. For notational convenience, we denote the coefficients $\tau^u_{\|\ \|}(B)$ by $\tau^u_p(B)$ when the norm $\|\ \|$ is the l_p norm. In this section we obtain explicit expressions for $\tau^u_p(B)$ when $p = 1$ and when $p = \infty$ rather than implicit optimal values of optimization problems. Our development follows [11, Sections 6, 9) and relies on [10].

We first state an explicit form for $\tau^u_1(B)$.

THEOREM 4.1. *Let B be an $n \times n$ real matrix, and let $u \in R^n$ where $n > 1$. Then*

$$\tau^u_1(B) = \max_{\substack{i,j=1,\dots,n \\ i \neq j \\ |u_i|+|u_j| \neq 0}} \left(|u_i| + |u_j|\right)^{-1}\left[\sum_{k=1}^n |u_j B_{ik} - u_i B_{jk}|\right]. \tag{4.1}$$

Proof. As the maximization problem defining $\tau_1^u(B)$ [in (3.1)] is of a convex function over the convex set $\mathscr{C} \equiv \{x \in R^n : \|x\|_1 \leqslant 1,\ x^T u = 0\}$, its maximum is attained at an extreme point of the set $\mathscr{C}$. The set of extreme points of the set $\mathscr{C}$ were identified in [10] to be the set $F \equiv \{(|u_i| + |u_j|)^{-1}(u_j e^i - u_i e^j) : i, j = 1, \dots, n,\ i \neq j,\ |u_i| + |u_j| \neq 0\}$. It is easily seen that (4.1) is the maximum of the corresponding convex function over F. ■

The computational effort required to compute $\tau_1^u(B)$ with the expression given in Theorem 4.1 is of the order of $O(n^3)$. We remark that the explicit expression for $\tau_1^u(B)$, obtained in Theorem 4.1, simplifies when the coordinates of the vector u consist only of zeros and ones.

Before presenting the explicit form of $\tau_\infty^u(B)$, we need an additional definition. For a vector $u \in R^n$, define the real-valued functional M^u on R^n by

$$M^u(a) \equiv \max_{\substack{\|x\|_\infty \leqslant 1 \\ x^T u = 0 \\ x \in R^n}} x^T a \qquad \text{for} \quad a \in R^n. \tag{4.2}$$

The following explicit expression for the above functional was obtained in [10]. First observe that by making the change of variable $x_i \to x_i$ if $u_i \geqslant 0$ and $x_i \to -x_i$ if $u_i < 0$, one can assume, without loss of generality, that $u \geqslant 0$. Let $i(1), \dots, i(n)$ be an enumeration of the indices $1, \dots, n$ such that

$$\frac{a_{i(1)}}{u_{i(1)}} \geqslant \frac{a_{i(2)}}{u_{i(2)}} \geqslant \cdots \geqslant \frac{a_{i(n)}}{u_{i(n)}}. \tag{4.3}$$

where $\alpha/0$ is defined to be $+\infty$ if $\alpha > 0$ and $-\infty$ if $\alpha \leqslant 0$. Then Theorem 4.1 of [10] shows that

$$M^u(a) = \sum_{t=1}^{m-1} a_{i(t)} + \gamma a_{i(m)} - \sum_{t=m+1}^{n} a_{i(t)}, \tag{4.4}$$

where $m \in \{1, \dots, n\}$ is the smallest integer for which $2\sum_{t=1}^m u_{i(t)} > \sum_{i=1}^n u_i$ and where

$$\gamma \equiv 1 + \left(\sum_{i=1}^n u_i - 2 \sum_{t=1}^m u_{i(t)} \right) u_{i(m)}^{-1}. \tag{4.5}$$

One can compute $M^u(a)$ by sorting the n numbers $\{a_i/u_i : i = 1, \dots, n\}$ and then computing $i(m)$, γ, and the right-hand side of (4.5). The computational complexity of this procedure is $O(n \log n)$ (e.g., [4, Vol. 3]). An alternative method for computing $M^u(a)$, whose computational complexity is $O(n)$, is given in [10].

When u is a vector consisting of zeros and ones, the explicit expression for $M^u(a)$ given in (4.4) simplifies, as $a_{i(m)}$ is a median of $\{a_i : i = 1, \dots, n, u_i \neq 0\}$ (e.g., [10, Section 4]). In particular, if k and p are such that $\{t = 1, \dots, n : u_{i(t)} \neq 0\} = \{k, k+1, \dots, p\}$, then $\gamma = 0$ if $p - k$ is even and $\gamma = -1$ if $p - k$ is odd. Also,

$$M^u(a) = \sum_{t=1}^{k-1} |a_{i(t)}| + \sum_{t=k}^{p} |a_{i(t)} - a_{i(m)}| + \sum_{t=p+1}^{n} |a_{i(t)}|. \tag{4.6}$$

We will next state the promised explicit form of $\tau_\infty^u(B)$.

THEOREM 4.2. *Let B be an $n \times n$ matrix, and let $u \in R^n$. For $j = 1, \dots, n$, let B^j be the jth column of B. Then*

$$\tau_\infty^u(B) = \max_{j=1,\dots,n} M^u(B^j). \tag{4.7}$$

Proof. Let $\mathscr{C} \equiv \{x \in R^n : \|x\|_\infty \leqslant 1,\ x^T u = 0\}$. Evidently,

$$\begin{aligned}\tau_\infty^u(B) &= \max_{x \in \mathscr{C}} \|x^T B\|_\infty = \max_{x \in \mathscr{C}} \max_{j=1,\dots,n} |x^T B^j| \\ &= \max_{j=1,\dots,n} \max_{x \in \mathscr{C}} |x^T B^j| = \max_{j=1,\dots,n} \max_{x \in \mathscr{C}} x^T B^j \\ &= \max_{j=1,\dots,n} M^u(B^j).\end{aligned}$$ ■

The computation of $\tau_\infty^u(B)$ requires n evaluations of the functional $M^u(\cdot)$. As noted, each of these evaluations can be accomplished by a method whose complexity is $O(n)$. Thus, $\tau_\infty^u(B)$ can be computed by a method whose complexity is $O(n^2)$. We remark that the computation of $\tau_\infty^u(B)$ simplifies when the coordinates of u consist only of zeros and ones, by using (4.6).

5. SIMPLE EIGENVALUES

In this section we present some variants of the results of Section 3 in the case where the eigenvalue λ is simple. We start with a summary of properties of simple eigenvalues.

We first need some additional notation. For a complex $n \times n$ matrix B, a vector $b \in C^n$ and an index $i \in \{1, \dots, n\}$, we use the notation $B^{(i,b)}$ ($B_{(i,b^T)}$) to denote the matrix all of whose columns (rows) coincide with those of B except for the ith one, which is b (b^T).

LEMMA 5.1. *Let Q be a complex $n \times n$ matrix where zero is a simple eigenvalue of Q, and let w and v^T be corresponding right and left eigenvectors, respectively. Also, let E be the eigenprojection of Q at zero, and let $i \in \{1, \dots, n\}$ satisfy $w_i \neq 0$. Then:*

(1) $\nu_0(Q) = 1$.
(2) $v^T w \neq 0$.
(3) $E = (v^T w)^{-1}(w v^T)$.
(4) $Q^{(i,w)}$ *is invertible.*
(5) *The ith row of* $[Q^{(i,w)}]^{-1}$ *is* $(v^T w)^{-1} v^T$.
(6) $Q^D = (I - E)^{(i,0)}[Q^{(i,w)}]^{-1}$.
(7) *If $x^T w = 0$, then for some $y \in C^n$, $y^T Q^D = x^T$. Moreover, if Q and x are real, y can be selected to be real.*

Proof. Let $\nu = \nu_0(Q)$. As zero is a simple eigenvalue of Q, we have that $\nu \geqslant 1$ and that $1 \leqslant \dim(\text{null}\, Q) \leqslant \dim(\text{null}\, Q^\nu) = 1$. We conclude that $\dim(\text{null}\, Q) = \dim(\text{null}\, Q^\nu)$, assuring that $\nu \leqslant 1$. It follows that $\nu = 1$, establishing (1).

Parts (2) and (3) follow directly from [7, Theorem 3.1].

Next observe that as $Qw = 0$ and $w_i \neq 0$, the ith column of Q can be expressed as a linear combination of the remaining columns. In particular, the subspaces spanned by the columns of Q and by the columns of $Q^{(i,0)}$ coincide. As zero is a simple eigenvalue of Q and, by part (1), $\nu_0(Q) = 1$, we have that null Q is spanned by a single vector. In particular, null $Q = \{\alpha w : \alpha \in C\}$. It follows that the columns of $Q^{(i,w)}$ span $(\text{range}\, Q) \oplus (\text{null}\, Q)$. As $\nu_0(Q) = 1$, we have from Lemma 2.1, part (1), that $(\text{range}\, Q) \oplus (\text{null}\, Q) = C^n$. Hence, the n columns of $Q^{(i,w)}$ span C^n, and standard results from linear algebra assure that $Q^{(i,w)}$ is invertible, establishing (4).

Next recall that $\nu^T Q = 0$. It follows that $\nu^T Q^{(i,w)} = (\nu^T w)(e^i)^T$, and therefore the invertibility of $Q^{(i,w)}$ implies that $(\nu^T w)^{-1} \nu^T = (e^i)^T [Q^{(i,w)}]^{-1}$, i.e., $(v^T w)^{-1} v^T$ is the ith row of $[Q^{(i,w)}]^{-1}$, establishing (5).

Next observe that $Q^D Q = I - E$ [see Lemma 2.1, part (8)] and that $(I - E)w = w - (v^T w)^{-1}(w v^T)w = 0$. It follows that $Q^D Q^{(i,w)} = (I - E)^{(i,0)}$ and therefore the invertibility of $Q^{(i,w)}$ implies that $Q^D = (I - E)^{(i,0)}[Q^{(i,w)}]^{-1}$, establishing (6).

Finally assume that $x^T w = 0$. Then $x^T E = x^T (v^T w)^{-1}(w v^T) = 0$, and therefore for $y^T \equiv x^T(Q - E)$, $y^T Q^D = [x^T(Q - E)][(Q - E)^{-1}(I - E)] =$

$x^T(I - E) = x^T$. Of course, if Q and x are real, so is y. This completes the proof of (7). ■

Parts (4)–(6) of Lemma 5.1 were first established by Denardo [2] for the special case where the matrix Q has the form $P - I$ where P is stochastic. In particular, in this case $w = e = (1,\dots,1)^T \in R^n$ has $w_i \neq 0$ for each $i \in \{1,\dots,n\}$.

Let A, λ, and w be as in Theorem 3.1. We next show that when λ is simple, the lower bound in (3.2), namely $[\tau^w_{\|\ \|}(A - \lambda I)^D]^{-1}$, can be computed without the prior computation of $(A - \lambda I)^D$. Our key tool is the following observation, which provides a (trivial) sufficient condition for two matrices to have identical τ coefficients (e.g., [11, Section 7]).

LEMMA 5.2. *Let B be a $n \times n$ matrix, let a and u be vectors in R^n, and let $\|\ \|$ be a norm on R^n. Then*

$$\tau^u_{\|\ \|}(B) = \tau^u_{\|\ \|}(B - ua^T). \tag{5.1}$$

Proof. The conclusion is immediate from the observation that if $x^Tu = 0$ for $x \in R^n$, then $x^TB = x^T(B - ua^T)$. ■

THEOREM 5.3. *Let A be a real $n \times n$ matrix, let λ be a real simple eigenvalue of A, and let w be a corresponding real, right eigenvector with $w_i \neq 0$ where $i \in \{1,\dots,n\}$. Also, let $\|\ \|$ be a norm on R^n. Then*

$$\tau^w_{\|\ \|}\left[(A - \lambda I)^D\right] = \tau^w_{\|\ \|}\left\{\left[(A - \lambda I)^{(i,w)}\right]^{-1}_{(i,0^T)}\right\} \tag{5.2}$$

and

$$\tau^w_{\|\ \|}\left[(A - \lambda I)^D\right] = \tau^w_{\|\ \|}\left[(A - \lambda I - E)^{-1}\right], \tag{5.3}$$

where E is the eigenprojection of A at λ.

Proof. Let J be the $n \times n$ diagonal matrix all of whose diagonal elements are one except of the ith one, which is zero. Evidently, for a real $n \times n$ matrix B,

$$B^{(i,0)} = BJ \quad \text{and} \quad B_{(i,0^T)} = JB.$$

Let v^T be a left eigenvector of A corresponding to eigenvalue λ. Lemma 5.1 assures that $v^T w \neq 0$, that $E = (v^T w)^{-1}(wv^T)$, and that $(A - \lambda I)^D = (I - E)^{(i,0)}[(A - \lambda I)^{(i,w)}]^{-1} = (I - E)J[(A - \lambda I)^{(i,w)}]^{-1}$. It follows that

$$
\begin{aligned}
(A - \lambda I)^D - \left[(A - \lambda I)^{(i,w)}\right]^{-1}_{(i,0^T)} &= (I - E)J\left[(A - \lambda I)^{(i,w)}\right]^{-1} \\
&\quad - J\left[(A - \lambda I)^{(i,w)}\right]^{-1} \\
&= -EJ\left[(A - \lambda I)^{(i,w)}\right]^{-1} \\
&= -w\left\{v^T(v^T w)^{-1}J\left[(A - \lambda I)^{(i,w)}\right]^{-1}\right\}.
\end{aligned}
$$

Hence, (5.2) follows directly from Lemma 5.2.

Next observe that Lemma 2.2 assures that $(A - \lambda I)^D = (A - \lambda I - E)^{-1} + E = (A - \lambda I - E)^{-1} + w[v^T(wv^T)^{-1}]$. Hence, (5.3) follows directly from Lemma 5.2. ∎

Let A, λ, and w be as in Theorem 5.3. We note that (5.2) and (5.3) show that $\tau^w_{\| \|}[(A - \lambda I)^D]$ can be determined by computing the corresponding τ coefficient of a matrix obtained from $A - \lambda I$ through a single matrix inversion; of course, the use of (5.2) requires the prior computation of the corresponding right eigenvector w, and the use of (5.3) requires the prior computation of both w and the corresponding eigenprojection E. The simplest method for computing E, given in part (3) of Lemma 5.1, requires the identification of a right and a left eigenvector of A corresponding to λ. It follows that (5.2) yields a faster method for computing $\tau^w_{\| \|}[(A - \lambda I)^D]$ than (5.3) does. We also note that Lemma 5.1 shows that the inversion of $(A - \lambda I)^{(i,w)}$ produces a left eigenvector v^T of A with respect to λ for which $E = wv^T$.

We end this section with the observation that when the eigenvalue λ of the matrix A is simple, the lower bound in Theorem 3.1 can be stated in terms of the matrix $A - \lambda I$ itself rather than in terms of its Drazin inverse. The disadvantage of this formulation is that it is not easily computable, as the corresponding expression for the bound is as a minimum of a (convex) function over a set which is *not* convex. Our proof follows Zarling [12, p. 37], who derives the results for stochastic matrices and the eigenvalue one.

We will need the following additional definition. Let $\| \ \|$ be a norm on R^n, and let $u \in R^n$. We define the coefficient $\eta^u_{\| \|}(B)$ for an $n \times n$ matrix B by

$$\eta^u_{\| \|}(B) \equiv \min_{\substack{\|x\| = 1 \\ x^T u = 0 \\ x \in R^n}} \|x^T B\|. \tag{5.4}$$

We next show that the bound obtained in Theorem 3.1 can be expressed in terms of the functionals $\eta^{\cdot}_{\cdot}(\cdot)$.

THEOREM 5.4. *Let A be a real $n \times n$ matrix, let λ be a real, simple eigenvalue of A, and let w be a corresponding real right eigenvector. Also, let $\|\ \|$ be a norm on R^n. Then*

$$\left[\tau^{w}_{\|\ \|}(A-\lambda I)^{D}\right]^{-1} = \eta^{w}_{\|\ \|}(A-\lambda I) \tag{5.5}$$

and

$$\tau^{w}_{\|\ \|}(A-\lambda I) = \left\{\eta^{w}_{\|\ \|}\left[(A-\lambda I)^{D}\right]\right\}^{-1}. \tag{5.6}$$

Proof. We first note that the homogeneity of the norm $\|\ \|$ assures that in the maximization problem defining $\tau^{w}_{\|\ \|}(\cdot)$ in (3.1), one can require that $\|x\| = 1$ rather than $\|x\| \leqslant 1$. For notational convenience we let $B \equiv A - \lambda I$.

Assume that $x^T w = 0$. As λ is a simple eigenvalue of A, zero is a simple eigenvalue of B and therefore, by Lemma 5.1, part (7), $x^T = y^T B^D$ for some $y \in R^n$. It therefore follows from part (9) of Lemma 2.1 that

$$x^T B^D B = y^T B^D B^D B = y^T B^D = x^T. \tag{5.7}$$

We also observe that $(x^T B)w = x^T(Bw) = 0$, and, by Lemma 2.3, we have that $(x^T B^D)w = x^T(B^D w) = 0$. It now follows from (5.7) that

$$\|x^T\| = \|x^T B B^D\| \leqslant \|x^T B\|\left[\tau^{w}_{\|\ \|}(B^D)\right] \tag{5.8}$$

and that

$$\|x^T\| = \|x^T B^D B\| \geqslant \|x^T B^D\|\left[\eta^{w}_{\|\ \|}(B)\right]. \tag{5.9}$$

Taking the maximum of (5.8) and minimum of (5.9) over $\{x \in R^n : \|x\| = 1,\ x^T w = 0\}$, implies, respectively, that

$$1 \leqslant \left[\eta^{w}_{\|\ \|}(B)\right]\left[\tau^{w}_{\|\ \|}(B^D)\right] \tag{5.10}$$

and that

$$1 \geqslant \left[\tau^{w}_{\|\ \|}(B^D)\right]\left[\eta^{w}_{\|\ \|}(B)\right]. \tag{5.11}$$

Of course, combining (5.10) and (5.11) establishes (5.5). Finally, (5.6) follows by interchanging the roles of B and B^D in the above arguments. ■

Theorems 5.3 and 5.4 can be extended to situations where λ is not necessarily a simple eigenvalue of A but $v_\lambda(A)=1$ (see [11, Section 9]).

6. EXAMPLES

Our first example demonstrates that the bounds in Theorem 3.1 are tight, namely, they can be attained.

EXAMPLE 6.1. Let

$$A=\begin{bmatrix} .2 & .5 & .3 \\ .4 & .3 & .3 \\ .4 & .2 & .4 \end{bmatrix}.$$

Then $\sigma(A)=\{-0.2, 0.1, 1\}$, and $e=(1,1,1)^T$ is a right eigenvector of A corresponding to the eigenvalue one. Also, $\max\{|\mu-1|: 1\neq\mu\in\sigma(A)\}=1.2$ and $\min\{|\mu-1|: 1\neq\mu\in\sigma(A)\}=0.9$. Next observe that

$$(A-I)^D=\begin{bmatrix} -.56 & .19 & .37 \\ .28 & -.65 & .37 \\ .28 & .46 & -.74 \end{bmatrix}.$$

It follows from the results of Section 4 that

$$\tau_1^e(A-I)=\tau_\infty^e(A-I)=1.2,$$

and that

$$\tau_1^e\left[(A-I)^D\right]=\tau_\infty\left[(A-I)^D\right]=1.11=(0.9)^{-1}.$$

The following example demonstrates that our lower bound on $\min\{|\mu-1|: 1\neq\mu\in\sigma(A)\}$ for a stochastic matrix A provides a measure of the complete indecomposability of A (see Section 1).

EXAMPLE 6.2. For $0<\varepsilon\leqslant 2^{-1}$, let

$$A_\varepsilon=\begin{bmatrix} 2^{-1}-\varepsilon & 2^{-1}-\varepsilon & \varepsilon & \varepsilon \\ 2^{-1}-\varepsilon & 2^{-1}-\varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 2^{-1}-\varepsilon & 2^{-1}-\varepsilon \\ \varepsilon & \varepsilon & 2^{-1}-\varepsilon & 2^{-1}-\varepsilon \end{bmatrix}.$$

Then $e=(1,1,1,1)^T$ is a right eigenvector of A_ε corresponding to the eigenvalue one and

$$[A_\varepsilon - I]^D = (16\varepsilon)^{-1}\begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}$$

$$+2^{-1}\begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

It follows from the results of Section 4 that $\tau_1^e[(A_\varepsilon - I)^D] = \tau_\infty^e[(A_\varepsilon - I)^D] = (4\varepsilon)^{-1}$. Of course, the smaller ε is, the more decomposable A_ε is, and indeed, that is accompanied by 4ε being smaller.

We are indebted to Ludo Van der Heyden for helpful comments on an earlier version of this paper.

REFERENCES

1 S. L. Campbell and C. D. Meyer, Jr., *Generalized Inverses of Linear Transformations*, Pitman, London, 1979.
2 E. V. Denardo, A Markov decision problem, in *Mathematical Programming* (T. C. Hu and S. M. Robinson, Eds.), Academic, New York, 1973.
3 T. Kato, *Perturbation Theory for Linear Operators*, Springer, New York, 1966.
4 D. E. Knuth, *The Art of Computer Programming*, Addison-Wesley, Reading, Mass., 1973.
5 U. G. Rothblum, Algebraic eigenspaces of nonnegative matrices, *Linear Algebra Appl.* 12:281–292 (1975).
6 U. G. Rothblum, A representation of the Drazin inverse and characterization of the index, *SIAM. J. Appl. Math.* 31:646–648 (1976).
7 U. G. Rothblum, Computation of the eigenprojection of a nonnegative matrix at its spectral radius, in *Mathematical Programming Studies 6, Stochastic Systems: Modeling, Identification and Optimization II* (Roger J.-B. Wets, Ed.), 1976, pp. 188–201.
8 U. G. Rothblum, Expansions of sums of matrix powers, *SIAM Rev.* 23:143–164 (1981).
9 U. G. Rothblum, Resolvent expansions of matrices and applications, *Linear Algebra Appl.* 38:33–49 (1981).
10 U. G. Rothblum, Explicit solutions to optimization problems on the intersections of the unit ball of the l_1 and l_∞ norms with a hyperplane, unpublished.

11 U. G. Rothblum and C. P. Tan, Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices, unpublished.
12 R. L. Zarling, Numerical solution of nearly completely decomposable queuing systems, Ph.D. Dissertation, Univ. of North Carolina, Chapel Hill, N.C., 1976.

Received 18 July 1983; revised 3 October 1983