Electronic Transactions on Numerical Analysis.
Volume 31, pp. 1-11, 2008.
Copyright © 2008, Kent State University.
ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

MAJORIZATION BOUNDS FOR RITZ VALUES OF HERMITIAN MATRICES*

CHRISTOPHER C. PAIGE† AND IVO PANAYOTOV‡

Abstract. Given an approximate invariant subspace we discuss the effectiveness of majorization bounds for assessing the accuracy of the resulting Rayleigh-Ritz approximations to eigenvalues of Hermitian matrices. We derive a slightly stronger result than previously for the approximation of k extreme eigenvalues, and examine some advantages of these majorization bounds compared with classical bounds. From our results we conclude that the majorization approach appears to be advantageous, and that there is probably much more work to be carried out in this direction.

Key words. Hermitian matrices, angles between subspaces, majorization, Lidskii's eigenvalue theorem, perturbation bounds, Ritz values, Rayleigh-Ritz method, invariant subspace.

AMS subject classifications. 15A18, 15A42, 15A57.

1. Introduction.

The Rayleigh-Ritz method for approximating eigenvalues of a Hermitian matrix A finds the eigenvalues of Y^HAY, where the columns of the matrix Y form an orthonormal basis for a subspace $\mathcal{Y}$ which is an approximation to some invariant subspace $\mathcal{X}$ of A, and Y^H denotes the complex conjugate transpose of Y. Here $\mathcal{Y}$ is called a trial subspace. The eigenvalues of Y^HAY do not depend on the particular choice of basis and are called Ritz values of A with respect to $\mathcal{Y}$. See Parlett [11, Chapters 10–13] for a nice treatment. If $\mathcal{Y}$ is one-dimensional and spanned by the unit vector y there is only one Ritz value—namely the Rayleigh quotient y^HAy. The Rayleigh-Ritz method is a classical approximation method. With the notation $\|x\| \equiv \sqrt{x^Hx}$ write

$$\mathrm{spr}(A) \equiv \lambda_{\max}(A) - \lambda_{\min}(A), \qquad A = A^H,$$
$$\theta(x,y) \equiv \arccos|x^Hy| \in [0,\pi/2], \qquad \|x\| = \|y\| = 1,$$

$\theta(x,y)$ being the acute angle between x and y. The classical result that motivates our research is the following: the Rayleigh quotient approximates an eigenvalue of a Hermitian matrix with accuracy proportional to the *square* of the eigenvector approximation error, see [12] and for example [1]: when $Ax = x \cdot x^HAx$, $\|x\| = \|y\| = 1$,

$$|x^HAx - y^HAy| \le \mathrm{spr}(A)\sin^2\theta(x,y). \tag{1.1}$$

Let $Ax = x\lambda$, then $x^HAx = \lambda$ so $|x^HAx - y^HAy| = |y^H(A-\lambda I)y|$. We now plug in the orthogonal decomposition $y = u + v$ where $u \in \mathrm{span}\{x\}$ and $v \in (\mathrm{span}\{x\})^\perp$. Thus $(A-\lambda I)u = 0$ and $\|v\| = \sin\theta(x,y)$, which results in

$$|y^H(A-\lambda I)y| = |v^H(A-\lambda I)v| \le \|A-\lambda I\| \cdot \|v\|^2 = \|A-\lambda I\|\sin^2\theta(x,y),$$

where $\|\cdot\|$ denotes the matrix norm subordinate to the vector norm $\|\cdot\|$. But $\|A-\lambda I\| \le \mathrm{spr}(A)$, proving the result.

It is important to realize that this bound depends on the unknown quantity $\theta(x,y)$, and thus is an *a priori* result. Such results help our understanding rather than produce computationally useful *a posteriori* results. As Wilkinson [14, p. 166] pointed out, *a priori* bounds are

* Received November 29, 2007. Accepted May 4, 2008. Published online on September 9, 2008. Recommended by Anne Greenbaum.

†School of Computer Science, McGill University, Montreal, Quebec, Canada, H3A 2A7 (paige@cs.mcgill.ca). Research supported by NSERC of Canada Grant OGP0009236.

‡Department of Mathematics and Statistics, McGill University, Montreal, Quebec, Canada, H3A 2K6 (ipanay@math.mcgill.ca). Research supported by FQRNT of Quebec Scholarship 100936.

of great value in assessing the relative performance of algorithms. Thus while (1.1) is very interesting in its own right—depending on $\sin^2\theta(x,y)$ rather than $\sin\theta(x,y)$—it could also be useful for assessing the performance of algorithms that iterate vectors y approximating x, in order to also approximate $x^H Ax$.

Now suppose an algorithm produced a succession of k-dimensional subspaces $\mathcal{Y}^{(j)}$ approximating an invariant subspace $\mathcal{X}$ of A. For example the block Lanczos algorithm of Golub and Underwood [4] is a Krylov subspace method which does this. In what ways can we generalize (1.1) to subspaces $\mathcal{X}$ and $\mathcal{Y}$ with $\dim\mathcal{X} = \dim\mathcal{Y} = k > 1$? In [7] Knyazev and Argentati stated the following conjecture generalizing (1.1) to the multidimensional setting. (See Section 2.1 for the definitions of $\lambda(\cdot)$ and $\theta(\cdot,\cdot)$).

CONJECTURE 1.1. *Let $\mathcal{X}$, $\mathcal{Y}$ be subspaces of $\mathbb{C}^n$ having the same dimension k, with orthonormal bases given by the columns of the matrices X and Y respectively. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix, and let $\mathcal{X}$ be A-invariant. Then*

$$|\lambda(X^H AX) - \lambda(Y^H AY)| \prec_w \operatorname{spr}(A)\sin^2\theta(\mathcal{X},\mathcal{Y}). \tag{1.2}$$

Here '$\prec_w$' denotes the weak submajorization relation, a concept which is explained in Section 2.2. Argentati, Knyazev, Paige and Panayotov [1] provided the following answer to the conjecture.

THEOREM 1.2. *Let $\mathcal{X}$, $\mathcal{Y}$ be subspaces of $\mathbb{C}^n$ having the same dimension k, with orthonormal bases given by the columns of the matrices X and Y respectively. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix, and let $\mathcal{X}$ be A-invariant. Then*

$$|\lambda(X^H AX) - \lambda(Y^H AY)| \prec_w \operatorname{spr}(A)\left(\sin^2\theta(\mathcal{X},\mathcal{Y}) + \frac{\sin^4\theta(\mathcal{X},\mathcal{Y})}{2}\right). \tag{1.3}$$

Moreover, if the A-invariant subspace $\mathcal{X}$ corresponds to the set of k largest or smallest eigenvalues of A then

$$|\lambda(X^H AX) - \lambda(Y^H AY)| \prec_w \operatorname{spr}(A)\sin^2\theta(\mathcal{X},\mathcal{Y}). \tag{1.4}$$

REMARK 1.3. This is slightly weaker than Conjecture 1.1—we were unable to prove the full conjecture, although all numerical tests we have done suggest that it is true. In numerical analysis we are mainly interested in these results as the angles become small, and then there is minimal difference between the right hand sides of (1.3) and (1.2), so proving the full Conjecture 1.1 is largely of mathematical interest.

Having thus motivated and reviewed Conjecture 1.1, in Section 2 we give the necessary notation and basic theory, then in Section 3 prove a slightly stronger result than (1.4), since in practice we are usually interested in the extreme eigenvalues. In Section 4 we derive results to show some benefits of these majorization bounds in comparison with the classical *a priori* eigenvalue error bounds (1.1), and add comments in Section 5. This is ongoing research, and there is probably much more to be found on this topic.

2. Definitions and Prerequisites.

2.1. Notation. For $x = [\xi_1,\ldots,\xi_n]^T$, $y = [\eta_1,\ldots,\eta_n]^T$, $u = [\mu_1,\ldots,\mu_n]^T \in \mathbb{R}^n$, we use $x^\downarrow \equiv [\xi_1^\downarrow,\ldots,\xi_n^\downarrow]^T$ to denote x with its elements rearranged in descending order, while $x^\uparrow \equiv [\xi_1^\uparrow,\ldots,\xi_n^\uparrow]^T$ denotes x with its elements rearranged in ascending order. We use $|x|$ to denote the vector x with the absolute value of its components and use '$\leq$' to compare real vectors componentwise. Notice that $x \leq y \Rightarrow x^\downarrow \leq y^\downarrow$, otherwise there would exist a first i such that $x_1^\downarrow \geq \cdots \geq x_i^\downarrow > y_i^\downarrow \geq \cdots \geq y_n^\downarrow$, leaving only $i-1$ elements $y_1^\downarrow,\ldots,y_{i-1}^\downarrow$ to dominate the i elements $x_1^\downarrow,\ldots,x_i^\downarrow$, a contradiction.

ETNA
Kent State University
etna@mcs.kent.edu

For real vectors x and y the expression $x \prec y$ means that x is majorized by y, while $x \prec_w y$ means that x is weakly submajorized by y. These concepts are explained in Section 2.2.

In our discussion $A \in \mathbb{C}^{n\times n}$ is a Hermitian matrix, $\mathcal{X}, \mathcal{Y}$ are subspaces of $\mathbb{C}^n$, and $\mathcal{X}$ is A-invariant. We write $\mathcal{X} = \mathcal{R}(X) \subset \mathbb{C}^n$ whenever the subspace $\mathcal{X}$ is equal to the range of the matrix X with n rows. The unit matrix is I and $e \equiv [1, \ldots, 1]^T$. We use $\mathrm{offdiag}(B)$ to denote B with its diagonal elements set to zero, while $\mathrm{diag\_of}(B) \equiv B - \mathrm{offdiag}(B)$.

We write $\lambda(A) \equiv \lambda^{\downarrow}(A)$ for the *vector* of eigenvalues of $A = A^H$ arranged in descending order, and $\sigma(B) \equiv \sigma^{\downarrow}(B)$ for the vector of singular values of B arranged in descending order. Individual eigenvalues and singular values are denoted by $\lambda_i(A)$ and $\sigma_i(B)$, respectively. The distance between the largest and smallest eigenvalues of A is denoted by $\mathrm{spr}(A) = \lambda_1(A) - \lambda_n(A)$, and the 2-norm of B is $\sigma_1(B) = \|B\|$.

The acute angle between two unit vectors x and y is denoted by $\theta(x, y)$ and is defined by $\cos\theta(x, y) = |x^H y| = \sigma(x^H y)$. Let $\mathcal{X}$ and $\mathcal{Y} \subset \mathbb{C}^n$ be subspaces of the same dimension k, each with orthonormal bases given by the columns of the matrices X and Y respectively. We denote the vector of principal angles between $\mathcal{X}$ and $\mathcal{Y}$ by $\theta(\mathcal{X}, \mathcal{Y}) \equiv \theta^{\downarrow}(\mathcal{X}, \mathcal{Y})$, and define it using $\cos\theta(\mathcal{X}, \mathcal{Y}) = \sigma^{\uparrow}(X^H Y)$; e.g., [3], [5, §12.4.3].

2.2. Majorization. Majorization compares two real n-vectors. Majorization inequalities appear naturally, e.g., when describing the spectrum or singular values of sums and products of matrices. Majorization is a well developed tool applied extensively in theoretical matrix analysis (see, e.g., [2, 6, 10]), but recently it has also been applied in the analysis of matrix algorithms; e.g., [8]. We briefly introduce the subject and state a few theorems which we use, followed by two nice theorems we do not use.

We say that $x \in \mathbb{R}^n$ is weakly submajorized by $y \in \mathbb{R}^n$, written $x \prec_w y$, if

$$\sum_{i=1}^{k} \xi_i^{\downarrow} \le \sum_{i=1}^{k} \eta_i^{\downarrow}, \qquad 1 \le k \le n, \tag{2.1}$$

while x is majorized by y, written $x \prec y$, if (2.1) holds together with

$$\sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n} \eta_i. \tag{2.2}$$

The linear inequalities of these two majorization relations define convex sets in $\mathbb{R}^n$. Geometrically $x \prec y$ if and only if the vector x is in the convex hull of all vectors obtained by permuting the coordinates of y; see, e.g., [2, Theorem II.1.10]. If $x \prec_w y$ one can also infer that x is in a certain convex set depending on y, but in this case the description is more complicated. In particular this convex set need not be bounded. However if $x, y \ge 0$ then the corresponding convex set is indeed bounded, see for example the pentagon in Figure 3.2.

From (2.1) $x \le y \Rightarrow x^{\downarrow} \le y^{\downarrow} \Rightarrow x \prec_w y$, but $x \prec_w y \not\Rightarrow x^{\downarrow} \le y^{\downarrow}$. The majorization relations '$\prec$' and '$\prec_w$' share some properties with the usual inequality relation '$\le$', but not all, so one should deal with them carefully. Here are basic results we use. It follows from (2.1) and (2.2) that $x + u \prec x^{\downarrow} + u^{\downarrow}$ (see, e.g., [2, Corollary II.4.3]), so with the logical '&'

$$\{x \prec_w y\} \;\&\; \{u \prec_w v\} \;\&\; \cdots \;\Rightarrow\; x + u + \cdots \prec x^{\downarrow} + u^{\downarrow} + \cdots \prec_w y^{\downarrow} + v^{\downarrow} + \cdots. \tag{2.3}$$

Summing the elements shows this also holds with '$\prec_w$' replaced by '$\prec$'.

THEOREM 2.1. *Let $x, y \in \mathbb{R}^n$. Then*

$$x \prec_w y \Leftrightarrow \exists u \in \mathbb{R}^n \ \text{ such that } \ x \le u \ \& \ u \prec y; \tag{2.4}$$

$$x \prec_w y \Leftrightarrow \exists u \in \mathbb{R}^n \ \text{ such that } \ x \prec u \ \& \ u \le y. \tag{2.5}$$

ETNA
Kent State University
etna@mcs.kent.edu

Proof. See, e.g., [2, p. 39] for (2.4). If $x \prec u$ & $u \leq y$, then $u^\downarrow \leq y^\downarrow$ and from (2.1) $x \prec_w y$. Suppose $x = x^\downarrow \prec_w y = y^\downarrow$. Define $\tau \equiv e^T y - e^T x$, then $\tau \geq 0$. Define $u \equiv y - e_n\tau$, then $u \leq y$ and $u = u^\downarrow$. But $e^T u = e^T y - \tau = e^T x$ with $\sum_{i=1}^j \xi_i \leq \sum_{i=1}^j \eta_i = \sum_{i=1}^j \mu_i$ for $1 \leq j \leq n-1$, so $x \prec u$, proving (2.5). $\square$

THEOREM 2.2. (Lidskii [9], see also, e.g., [2, p. 69]).
Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian. Then $\lambda(A) - \lambda(B) \prec \lambda(A - B)$.

THEOREM 2.3. (See, e.g., [6, Theorem 3.3.16], [2, p. 75]). *$\sigma(AB) \leq \|A\|\sigma(B)$ and $\sigma(AB) \leq \|B\|\sigma(A)$ for arbitrary matrices A and B such that AB exists.*

THEOREM 2.4. (“Schur's Theorem”, see, e.g., [2, p. 35]).
Let $A \in \mathbb{C}^{n\times n}$ be Hermitian, then $\mathrm{diag\_of}(A)e \prec \lambda(A)$.

For interest, here are two results involving ‘$\prec_w$’ that we do not use later.

THEOREM 2.5. (Weyl, see, e.g., [6, Theorem 3.3.13 (a), pp. 175–6]).
For any $A \in \mathbb{C}^{n\times n}$, $|\lambda(A)| \prec_w \sigma(A)$.

Majorization inequalities are intimately connected with norm inequalities:

THEOREM 2.6. (Fan 1951, see, e.g., [6, Corollary 3.5.9], [13, § II.3]). *Let $A, B \in \mathbb{C}^{m\times n}$. Then $\sigma(A) \prec_w \sigma(B) \Leftrightarrow |||A||| \leq |||B|||$ for every unitarily invariant norm $|||\cdot|||$.*

3. The special case of extreme eigenvalues. In (1.1) we saw that if x is an eigenvector of a Hermitian matrix A and y is an approximation to x, $x^H x = y^H y = 1$, then the Rayleigh quotient $y^H Ay$ is a superior approximation to the eigenvalue $x^H Ax$ of A. A similar situation occurs in the multi-dimensional case. Suppose $X, Y \in \mathbb{C}^{n\times k}$, $X^H X = Y^H Y = I_k$, $\mathcal{X} = \mathcal{R}(X)$, $\mathcal{Y} = \mathcal{R}(Y)$, where $\mathcal{X}$ is A-invariant, i.e. $AX = X(X^H AX)$. Then $\lambda(X^H AX)$ is a vector containing the k eigenvalues of the matrix A corresponding to the invariant $\mathcal{X}$. Suppose that $\mathcal{Y}$ is some approximation to $\mathcal{X}$, then $\lambda(Y^H AY)$, called the vector of Ritz values of A relative to $\mathcal{Y}$, approximates $\lambda(X^H AX)$. Theorem 1.2 extends (1.1) by providing an upper bound for $d \equiv |\lambda(Y^H AY) - \lambda(X^H AX)|$. The componentwise inequality $d^\downarrow \leq \mathrm{spr}(A)\sin^2\theta(\mathcal{X}, \mathcal{Y})$ is false, but it can be relaxed to weak submajorization to give Theorem 1.2. For the proof of the general statement (1.3) of Theorem 1.2 and for some other special cases not treated here we refer the reader to [1]. That paper also shows that the conjectured bound cannot be made any tighter, and discusses the issues which make the proof of the full Conjecture 1.1 difficult.

Instead of (1.4) in Theorem 1.2, in Theorem 3.3 we prove a stronger result involving ‘$\prec$’ (rather than ‘$\prec_w$’) for this special case of extreme eigenvalues. We will prove the result for $\mathcal{X} = \mathcal{R}(X)$ being the invariant space for the k largest eigenvalues of A. We would replace A by $-A$ to prove the result for the k smallest eigenvalues. The eigenvalues and Ritz values depend on the subspaces $\mathcal{X}, \mathcal{Y}$ and not on the choice of orthonormal bases. If we choose Y such that $\mathcal{Y} = \mathcal{R}(Y)$, $Y^H Y = I$, and $Y^H AY$ is the diagonal matrix of Ritz values, then the columns of Y are called Ritz vectors. In this section we choose bases which usually are not eigenvectors or Ritz vectors, so we use the notation $\widetilde{X}, \widetilde{Y}$ to indicate this. We first provide a general result for $A = A^H$.

THEOREM 3.1. (See [1]). *Let $\mathcal{X}$, $\mathcal{Y}$ be subspaces of $\mathbb{C}^n$ having the same dimension k, with orthonormal bases given by the columns of the matrices $\widetilde{X}$ and $\widetilde{Y}$ respectively. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix, $\mathcal{X}$ be A-invariant, $[\widetilde{X}, \widetilde{X}_\perp] \in \mathbb{C}^{n\times n}$ unitary, and write $C \equiv \widetilde{X}^H\widetilde{Y}$, $S \equiv \widetilde{X}_\perp^H\widetilde{Y}$, $A_{11} \equiv \widetilde{X}^H A\widetilde{X}$, $A_{22} \equiv \widetilde{X}_\perp^H A\widetilde{X}_\perp$. Then $\widetilde{X}$ and $\widetilde{Y}$ may be chosen to give real diagonal $C \geq 0$ with $C^2 + S^H S = I_k$, and*

$$d \equiv \lambda(\widetilde{X}^H A\widetilde{X}) - \lambda(\widetilde{Y}^H A\widetilde{Y}) = \lambda(A_{11}) - \lambda(CA_{11}C + S^H A_{22}S). \tag{3.1}$$

ETNA
Kent State University
etna@mcs.kent.edu

Proof. By using the singular value decomposition we can choose $\widetilde{Y}$ and unitary $[\widetilde{X}, \widetilde{X}_\perp]$ to give $k \times k$ diagonal $\widetilde{X}^H \widetilde{Y}^H = C \geq 0$, and $(n-k) \times k$ S, in

$$\mathcal{X} = \mathcal{R}(\widetilde{X}), \quad \mathcal{Y} = \mathcal{R}(\widetilde{Y}), \quad [\widetilde{X}, \widetilde{X}_\perp]^H \widetilde{Y} = \begin{bmatrix} \widetilde{X}^H \widetilde{Y} \\ \widetilde{X}_\perp^H \widetilde{Y} \end{bmatrix} = \begin{bmatrix} C \\ S \end{bmatrix}, \quad C^2 + S^H S = I_k, \tag{3.2}$$

where with the definition of angles between subspaces, and appropriate ordering,

$$\begin{aligned} &\cos\theta(\mathcal{X}, \mathcal{Y}) = \sigma^\uparrow(\widetilde{X}^H \widetilde{Y}) = Ce, \\ &\sin^2\theta(\mathcal{X}, \mathcal{Y}) = e - \cos^2\theta(\mathcal{X}, \mathcal{Y}) = \lambda(I_k - C^2) = \lambda(S^H S) = \sigma(S^H S). \end{aligned} \tag{3.3}$$

Since $\mathcal{X}$ is A-invariant and $[\widetilde{X}, \widetilde{X}_\perp]$ is unitary:

$$[\widetilde{X}, \widetilde{X}_\perp]^H A\, [\widetilde{X}, \widetilde{X}_\perp] = \operatorname{diag}(A_{11}, A_{22}), \text{ and } A = [\widetilde{X}, \widetilde{X}_\perp] \operatorname{diag}(A_{11}, A_{22}) [\widetilde{X}, \widetilde{X}_\perp]^H,$$

where $\widetilde{X}^H A \widetilde{X} = A_{11} \in \mathbb{C}^{k \times k}$ and $(\widetilde{X}_\perp)^H A \widetilde{X}_\perp = A_{22} \in \mathbb{C}^{(n-k)\times(n-k)}$.
We can now use $\widetilde{Y}^H [\widetilde{X}, \widetilde{X}_\perp] = [C^H, S^H] = [C, S^H]$ to show that

$$\begin{aligned} \widetilde{Y}^H A \widetilde{Y} &= \widetilde{Y}^H \left([\widetilde{X}, \widetilde{X}_\perp] \operatorname{diag}(A_{11}, A_{22}) [\widetilde{X}, \widetilde{X}_\perp]^H \right) \widetilde{Y} \\ &= \begin{bmatrix} C & S^H \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} C \\ S \end{bmatrix} = C A_{11} C + S^H A_{22} S. \end{aligned}$$

The expression we will later bound thus takes the form in (3.1). ∎

Now assume A_{11} in Theorem 3.1 has the k largest eigenvalues of A. We see that (3.1) is shift independent, so we assume we have shifted $A := A - \lambda_{\min}(A_{11}) I$ to make both the new A_{11} and $-A_{22}$ nonnegative definite, see Figure 3.1, and we now have nonnegative definite square roots $\sqrt{A_{11}}$ and $\sqrt{-A_{22}}$, giving

$$\|A_{11}\| + \|A_{22}\| = \operatorname{spr}(A). \tag{3.4}$$

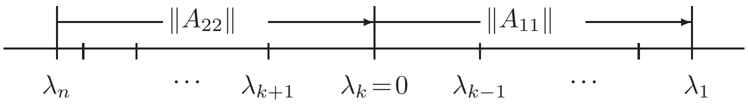

FIGURE 3.1. *Eigenvalues of the shifted matrix*

We give a lemma to use in our theorem for the improved version of (1.4).

LEMMA 3.2. *If $-A_{22} \in \mathbb{C}^{(n-k)\times(n-k)}$ is Hermitian nonnegative definite and $S \in \mathbb{C}^{(n-k)\times k}$, then $0 \leq \lambda(-S^H A_{22} S) \leq \|A_{22}\| \sigma(S^H S)$.*

Proof. With the Cholesky factorization $-A_{22} = L_2 L_2^H$ we have from Theorem 2.3

$$0 \leq \lambda(-S^H A_{22} S) = \sigma(-S^H A_{22} S) = \sigma^2(L_2^H S) \leq \|L_2\|^2 \sigma^2(S) = \|A_{22}\| \sigma(S^H S). \quad ∎$$

THEOREM 3.3. *Assume the notation and conditions of Theorem 3.1, but now also assume that the A-invariant $\mathcal{X}$ corresponds to the k largest eigenvalues of A. If we shift so that $A := A - \lambda_{\min}(A_{11}) I$, then the new A_{11} and $-A_{22}$ are nonnegative definite and*

$$\begin{aligned} 0 \leq d \equiv \lambda(\widetilde{X}^H A \widetilde{X}) - \lambda(\widetilde{Y}^H A \widetilde{Y}) \prec u \equiv \lambda\left(\sqrt{A_{11}} S^H S \sqrt{A_{11}}\right) + \lambda(-S^H A_{22} S) \\ \leq \operatorname{spr}(A) \sin^2\theta(\mathcal{X}, \mathcal{Y}). \end{aligned} \tag{3.5}$$

Proof. Cauchy's Interlacing Theorem shows $d \geq 0$; see, e.g., [2, p. 59]. From (3.1)

$$d = \{\lambda(A_{11}) - \lambda(CA_{11}C)\} \;+\; \left\{\lambda(CA_{11}C) - \lambda(CA_{11}C + S^H A_{22} S)\right\}. \tag{3.6}$$

Using Lidskii's Theorem (Theorem 2.2 here) we have in (3.6)

$$\lambda(CA_{11}C) - \lambda(CA_{11}C + S^H A_{22}S) \;\prec\; \lambda(-S^H A_{22} S), \tag{3.7}$$

where since $-A_{22}$ is nonnegative definite we see from Lemma 3.2 and (3.3) that

$$0 \leq \lambda(-S^H A_{22}S) \leq \|A_{22}\|\sigma(S^H S) = \|A_{22}\| \sin^2\theta(\mathcal{X},\mathcal{Y}). \tag{3.8}$$

Next since AB and BA have the same nonzero eigenvalues, by using Lidskii's Theorem, $C^2 + S^H S = I$, and Theorem 2.3, we see in (3.6) that (this was proven in [1]):

$$\begin{aligned} 0 \leq \lambda(A_{11}) - \lambda(CA_{11}C) &= \lambda(\sqrt{A_{11}}\sqrt{A_{11}}) - \lambda\left(\sqrt{A_{11}}C^2\sqrt{A_{11}}\right) \\ &\prec \lambda\left(\sqrt{A_{11}}\sqrt{A_{11}} - \sqrt{A_{11}}C^2\sqrt{A_{11}}\right) \\ &= \lambda\left(\sqrt{A_{11}}\,(I - C^2)\,\sqrt{A_{11}}\right) = \lambda\left(\sqrt{A_{11}}S^H S\sqrt{A_{11}}\right) \qquad (3.9) \\ &\leq \|A_{11}\|\sigma(S^H S) = \operatorname{spr}(A_{11})\sin^2\theta(\mathcal{X},\mathcal{Y}). \qquad (3.10) \end{aligned}$$

Combining (3.6), (3.7) and (3.9) via the '$\prec$' version of (2.3) gives

$$d \equiv \lambda(\widetilde{X}^H A\widetilde{X}) - \lambda(\widetilde{Y}^H A\widetilde{Y}) \prec u \equiv \lambda\left(\sqrt{A_{11}}S^H S\sqrt{A_{11}}\right) + \lambda(-S^H A_{22}S), \tag{3.11}$$

and using the bounds (3.8) and (3.10) with (3.4) proves (3.5). ◻

REMARK 3.4. Since $0 \leq d$, we see from (2.5) that (3.5) implies (1.4), and (1.4) implies (3.5) *for some* u. The improvement in Theorem 3.3 is that it provides a useful such u in (3.11). This is a small advance, but any insight might help in this area.

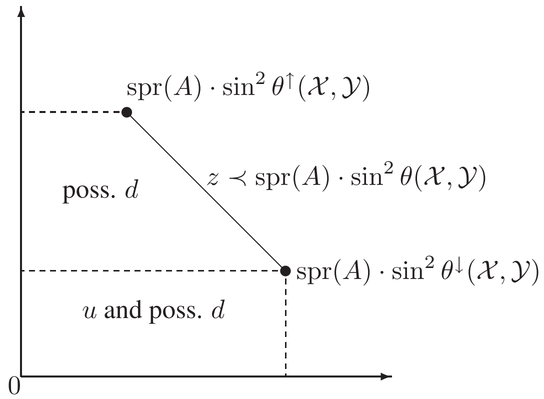

FIGURE 3.2. $0 \leq d \prec u \leq b \equiv \operatorname{spr}(A)\cdot\sin^2\theta^{\downarrow}(\mathcal{X},\mathcal{Y})$, *so* $d \prec_w b$, *and* d *must lie in the pentagon.*

Figure 3.2 illustrates the $\mathbb{R}^2$ case of possible d and u if we know only the vector $b \equiv \operatorname{spr}(A)\cdot\sin^2\theta(\mathcal{X},\mathcal{Y})$. Note that this illustrates possible $d^{\uparrow}$ as well as $d^{\downarrow}$. Later we show we can do better by using more information about $u \equiv \lambda\left(\sqrt{A_{11}}S^H S\sqrt{A_{11}}\right) + \lambda(-S^H A_{22}S)$.

ETNA
Kent State University
etna@mcs.kent.edu

4. Comparison with the classical bounds. Here we compare the present majorization approach with the classical approach in the case where the conjectured bound (1.2) holds. For comments on this see Remark 1.3. We can add:

COROLLARY 4.1. *If Conjecture 1.1 is true, then from [2, Example II.3.5 (iii)]*

$$|\lambda(X^HAX)-\lambda(Y^HAY)|^p \prec_w \operatorname{spr}(A)^p \sin^{2p}\theta(\mathcal{X},\mathcal{Y}) \quad \textit{and}$$
$$\|\lambda(X^HAX)-\lambda(Y^HAY)\|_p \le \|\operatorname{spr}(A)\sin^2\theta(\mathcal{X},\mathcal{Y})\|_p, \quad \textit{for } p\ge 1, \tag{4.1}$$

where the exponent is applied to each element of a vector, (4.1) comes from the last inequality in (2.1), and these are the standard vector p-norms; see, e.g., [2, p. 84].

Given an A-invariant subspace $\mathcal{X}$ and an approximation $\mathcal{Y}$, both of dimension k, the choice of respective orthonormal bases X and Y does not change the Ritz values. Take $X = [x_1,\ldots,x_k]$, $Y=[y_1,\ldots,y_k]$, each with orthonormal columns so that X^HAX, Y^HAY are diagonal matrices with elements decreasing along the main diagonal. Thus the x_i are (some choice of) eigenvectors of A corresponding to the subspace $\mathcal{X}$, while the y_i are Ritz vectors of A corresponding to $\mathcal{Y}$. Then the classical result (1.1) shows that

$$|x_i^HAx_i - y_i^HAy_i| \le \operatorname{spr}(A)\sin^2\theta(x_i,y_i), \qquad i=1,\ldots,k. \tag{4.2}$$

Because it uses angles between vectors rather than angles between subspaces, this bound can be unnecessarily weak. As an extreme example, if x_1 and x_2 correspond to a double eigenvalue, then it is possible to have $y_1 = x_2$ and $y_2 = x_1$, giving the extremely poor bound in (4.2) of $0 = |x_i^HAx_i - y_i^HAy_i| \le \operatorname{spr}(A)$ for both $i=1$ and $i=2$.

Setting $c \equiv [\cos\theta(x_1,y_1),\ldots,\cos\theta(x_k,y_k)]^T$, $s\equiv[\sin\theta(x_1,y_1),\ldots,\sin\theta(x_k,y_k)]^T$, and c^2, s^2 to be the respective vectors of squares of the elements of c and s, here we can rewrite these k classical eigenvalue bounds as

$$d \equiv |X^HAX - Y^HAY|e = |\lambda(X^HAX)-\lambda(Y^HAY)| \le \operatorname{spr}(A)s^2. \tag{4.3}$$

We will compare this to the conjectured (1.2) and the known (1.4):

$$d \equiv |\lambda(X^HAX)-\lambda(Y^HAY)| \prec_w \operatorname{spr}(A)\sin^2\theta(\mathcal{X},\mathcal{Y}). \tag{4.4}$$

This does not have the weakness mentioned regarding (4.2), which gives it a distinct advantage. The expressions (4.3) and (4.4) have similar forms, but differ in the angles and relations that are used. Notice that $s^2 = e - |\text{diag\_of}(X^HY)|^2e$, whereas $\sin^2\theta(\mathcal{X},\mathcal{Y}) = e - [\sigma^2(X^HY)]^\uparrow$. Here X^HY contains information about the relative positions of $\mathcal{X}$ and $\mathcal{Y}$. In the classical case we use only the diagonal of X^HY to estimate the eigenvalue approximation error, whereas in the majorization approach we use the singular values of this product. Note in comparing the two bounds that in the inequality relation the order of the elements must be respected, whereas in the majorization relation the order in which the errors are given does not play a role.

Before dealing with more theory we present an illustrative example. Let

$$A = \begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}, \quad X = \begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}, \quad Y = \begin{bmatrix}\frac{1}{\sqrt3}&0\\ \frac{1}{\sqrt3}&\frac{1}{\sqrt2}\\ \frac{1}{\sqrt3}&\frac{-1}{\sqrt2}\end{bmatrix},$$

where the columns of $X=[x_1,x_2]$ are eigenvectors of A corresponding to the eigenvalues 1 and 0 respectively. Since

$$Y^HAY = \begin{bmatrix}\frac13&0\\0&0\end{bmatrix}$$

ETNA
Kent State University
etna@mcs.kent.edu

is diagonal, the ordered Ritz values for the subspace $\mathcal{Y} = \mathcal{R}(Y)$ are just 1/3 and 0, with corresponding Ritz vectors given by the columns y_1 and y_2 of Y. Hence $\mathrm{spr}(A) = 1$ and

$$\begin{bmatrix}\cos\theta(x_1,y_1)\\ \cos\theta(x_2,y_2)\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} \quad\Rightarrow\quad s^2 \equiv \begin{bmatrix}\sin^2\theta(x_1,y_1)\\ \sin^2\theta(x_2,y_2)\end{bmatrix} = \begin{bmatrix}\frac{2}{3}\\ \frac{1}{2}\end{bmatrix}.$$

On the other hand we have for $\sin^2\theta^{\downarrow}(\mathcal{X},\mathcal{Y})$ and the application of (4.3) and (4.4):

$$\cos\theta^{\uparrow}(\mathcal{X},\mathcal{Y}) = \sigma(X^HY) = \sigma\left(\begin{bmatrix}\frac{1}{\sqrt{3}} & 0\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}}\end{bmatrix}\right) = \begin{bmatrix}1\\ \frac{1}{\sqrt{6}}\end{bmatrix} \Rightarrow \sin^2\theta^{\downarrow}(\mathcal{X},\mathcal{Y}) = \begin{bmatrix}\frac{5}{6}\\ 0\end{bmatrix},$$

$$d \equiv |\lambda(X^HAX) - \lambda(Y^HAY)| = \begin{bmatrix}\frac{2}{3}\\ 0\end{bmatrix} \leq s^2 = \begin{bmatrix}\frac{2}{3}\\ \frac{1}{2}\end{bmatrix}, \quad d = \begin{bmatrix}\frac{2}{3}\\ 0\end{bmatrix} \prec_w \sin^2\theta^{\downarrow}(\mathcal{X},\mathcal{Y}) = \begin{bmatrix}\frac{5}{6}\\ 0\end{bmatrix},$$

showing how (4.3) holds, so this nonnegative d lies in the dashed-line bounded area in Figure 4.1, and how (4.4) holds, so that d lies below the thin outer diagonal line. We also see how the later theoretical relationships (4.5), (4.7), and (4.8) are satisfied.

In this example we are approximating the two largest eigenvalues 0 and 1, so (3.5) must hold. In this example $A_{22} = 0$, so $u = \lambda(\sqrt{A_{11}}S^HS\sqrt{A_{11}})$. To satisfy (3.2) we need the SVD of X^HY (here $\widetilde{X} = XU$, $\widetilde{Y} = YV$):

$$U^HX^HYV = \frac{1}{\sqrt{5}}\begin{bmatrix}-2 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}1/\sqrt{3} & 0\\ 1/\sqrt{3} & 1/\sqrt{2}\end{bmatrix}\begin{bmatrix}-\sqrt{2} & \sqrt{3}\\ \sqrt{3} & \sqrt{2}\end{bmatrix}\frac{1}{\sqrt{5}} = C = \begin{bmatrix}1/\sqrt{6} & 0\\ 0 & 1\end{bmatrix}.$$

The form of C shows $S = [\sqrt{5/6}, 0]$ in (3.2), giving $u^T = [2/3, 0]$ since

$$A_{11} = U^HX^HAXU = U^He_1e_1^TU = \frac{1}{5}\begin{bmatrix}4 & -2\\ -2 & 1\end{bmatrix}, \quad A_{11}^2 = A_{11}, \quad u = \lambda\left(\frac{2}{3}A_{11}\right).$$

Since $d \prec u$ in (3.5), d must lie *on* the thick inner diagonal line in Figure 4.1. In fact it is at the bottom right corner of this convex set. It can be seen that $d \prec u$ is very much stronger than $d \leq \mathrm{spr}(A)\sin^2\theta(\mathcal{X},\mathcal{Y})$ in (3.5), and that $d \prec u$ describes by far the smallest of the three sets containing d.

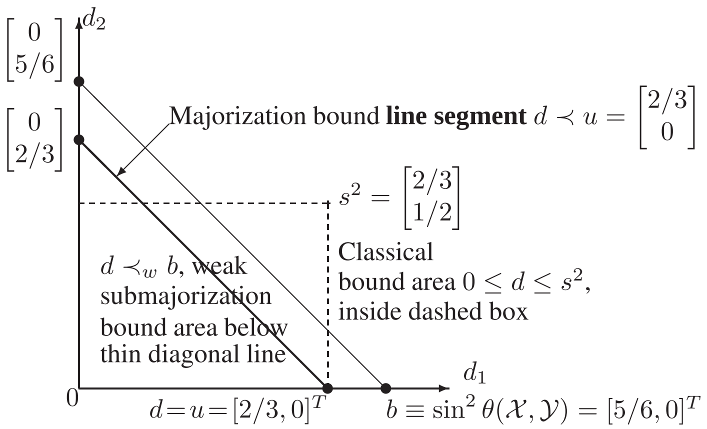

FIGURE 4.1. *Majorization and classical bounds on the eigenvalue error vector d,* $\mathrm{spr}(A) = 1$.

The next theorem and its corollaries illustrate situations where the general majorization bound is superior to the classical bound.

ETNA
Kent State University
etna@mcs.kent.edu

THEOREM 4.2. *Let $X = [x_1, \ldots, x_k]$ be a matrix of k orthonormal eigenvectors of A, and $Y = [y_1, \ldots, y_k]$ a matrix of k orthonormal Ritz vectors of A. Then with $\mathcal{X} \equiv \mathcal{R}(X)$ and $\mathcal{Y} \equiv \mathcal{R}(Y)$ and the notation presented earlier we have*

$$\sum_{i=1}^{j} \sin^2 \theta_i^{\uparrow}(\mathcal{X}, \mathcal{Y}) \le \sum_{i=1}^{j} \sin^2 \theta^{\uparrow}(x_i, y_i), \qquad 1 \le j \le k. \tag{4.5}$$

Proof. Notice that $c^2 = |\mathrm{diag\_of}(X^H Y)|^2 e \le \mathrm{diag\_of}(Y^H X X^H Y)e$. Using Schur's Theorem (Theorem 2.4 here) we have

$$\begin{aligned} c^2 = |\mathrm{diag\_of}(X^H Y)|^2 e &\le \mathrm{diag\_of}(Y^H X X^H Y)e \\ &\prec \lambda(Y^H X X^H Y) = \sigma^2(X^H Y) = \cos^2 \theta(\mathcal{X}, \mathcal{Y}). \end{aligned}$$

We apply (2.4) to this, then use (2.1), to give

$$c^2 \prec_w \cos^2 \theta(\mathcal{X}, \mathcal{Y}), \qquad \sum_{i=1}^{j} \cos^2 \theta^{\uparrow}(x_i, y_i) \le \sum_{i=1}^{j} \cos^2 \theta_i^{\uparrow}(\mathcal{X}, \mathcal{Y}), \quad 1 \le j \le k,$$

from which (4.5) follows. ☐

Notice that the angles in (4.5) are given in *increasing* order. Thus (4.5) does *not* show that $\sin^2 \theta(\mathcal{X}, \mathcal{Y}) \prec_w s^2$ for comparing (4.3) and (4.4). It is not important here, but the majorization literature has a special notation for denoting relations of the form (4.5), whereby (4.5) can be rewritten as

$$s^2 \prec^w \sin^2 \theta(\mathcal{X}, \mathcal{Y}).$$

Here '$\prec^w$' means 'is weakly supermajorized by'. In general $x \prec_w y \Leftrightarrow -x \prec^w -y$, see for example [2, pp. 29–30].

Theorem 4.2 has the following important consequence.

COROLLARY 4.3. *The majorization bound (4.4) provides a better estimate for the total error defined as the sum of all the absolute errors (or equivalently k times the average error) of eigenvalue approximation than the classical bounds (4.3). That is*

$$\|\lambda(X^H A X) - \lambda(Y^H A Y)\|_1 \le \mathrm{spr}(A) \|\sin^2 \theta(\mathcal{X}, \mathcal{Y})\|_1 \le \mathrm{spr}(A) \|s^2\|_1. \tag{4.6}$$

It follows that if we are interested in the overall (average) quality of approximation of the eigenvalue error, rather than a specific component, the majorization bound provides a better estimate than the classical one. The improvement Δ^2 in this total error bound satisfies

$$\begin{aligned} \Delta^2/\mathrm{spr}(A) &\equiv e^T s^2 - e^T \sin^2 \theta(\mathcal{X}, \mathcal{Y}) = e^T \cos^2 \theta(\mathcal{X}, \mathcal{Y}) - e^T c^2 \\ &= e^T \sigma^2(X^H Y) - e^T c^2 \\ &= \|X^H Y\|_F^2 - \|\mathrm{diag\_of}(X^H Y)\|_F^2 = \|\mathrm{offdiag}(X^H Y)\|_F^2. \end{aligned} \tag{4.7}$$

Note that $\Delta^2 \to 0$ as $Y \to X$, but that Δ^2 can stay positive even as $\mathcal{Y} \to \mathcal{X}$. This is a weakness of the classical bound similar to that mentioned following (4.2). Thus since $\sin^2 \theta(\mathcal{X}, \mathcal{Y}) = 0 \Leftrightarrow \mathcal{Y} = \mathcal{X}$, the majorization bound is tight as $\mathcal{Y} \to \mathcal{X}$, while the classical bound might not be.

Equation (4.7) also leads to a nice geometrical result. See Figure 4.1 for insight.

COROLLARY 4.4. *The point $\mathrm{spr}(A)s^2$ of the classical bound (4.3) is never contained within the majorization bound (4.4) unless $s^2 = 0$, in which case $|X^H Y| = I$ and both* *bounds are zero. So unless $|X^HY| = I$ the majorization bound always adds information to the classical bound. Mathematically*

$$|X^HY| \neq I \Leftrightarrow s \neq 0, \qquad s \neq 0 \Rightarrow s^2 \not\prec_w \sin^2\theta(\mathcal{X},\mathcal{Y}). \tag{4.8}$$

Proof. The first part of (4.8) follows from the definition $s \equiv (\sin\theta(x_i, y_i))$, and then the second follows from (4.7) since $e^Ts^2 > e^T\sin^2\theta(\mathcal{X},\mathcal{Y})$ shows that $s^2 \prec_w \sin^2\theta(\mathcal{X},\mathcal{Y})$ does not hold, see the definition of '$\prec_w$' in (2.1). □

From the previous corollary we see that $\mathrm{spr}(A)s^2$ must lie outside the pentagon in Figure 3.2. It might be thought that we could still have $\mathrm{spr}(A)s^2$ lying between zero and the extended diagonal line in Figure 3.2, but this is not the case. In general:

COROLLARY 4.5. *Let $\mathcal{H}$ denote the convex hull of the set*

$$\{P\sin^2\theta(\mathcal{X},\mathcal{Y}) : \ P \textit{ is a } k\times k \textit{ permutation matrix}\}.$$

Then $\mathcal{H}$ lies on the boundary of the half-space $\mathcal{S} \equiv \{z \,:\, e^Tz \leq e^T\sin^2\theta(\mathcal{X},\mathcal{Y})\}$, and s^2 is not in the strict interior of this half-space.

Proof. For any $z \in \mathcal{H}$ we have $z = (\sum_i \alpha_i P_i)\sin^2\theta(\mathcal{X},\mathcal{Y})$ for a finite number of permutation matrices P_i with $\sum_i \alpha_i = 1$, all $\alpha_i \geq 0$. Thus $e^Tz = e^T\sin^2\theta(\mathcal{X},\mathcal{Y})$ and z lies on the boundary of $\mathcal{S}$. Therefore $\mathcal{H}$ lies on the boundary of $\mathcal{S}$. From (4.7) $e^Ts^2 = e^T\sin^2\theta(\mathcal{X},\mathcal{Y}) + \|\mathrm{offdiag}(X^HY)\|_F^2 \geq e^T\sin^2\theta(\mathcal{X},\mathcal{Y})$, so s^2 cannot lie in the interior of $\mathcal{S}$. □

5. Comments and conclusions. For a given approximate invariant subspace we discussed majorization bounds for the resulting Rayleigh-Ritz approximations to eigenvalues of Hermitian matrices. We showed some advantages of this approach compared with the classical bound approach. We gave a proof in Theorem 3.3 of (3.5), a slightly stronger result than (1.4), proven in [1]. This suggests the possibility that knowing

$$u \equiv \lambda\left(\sqrt{A_{11}}S^HS\sqrt{A_{11}}\right) + \lambda(-S^HA_{22}S)$$

in (3.5) could be more useful than just knowing its bound $u \leq \mathrm{spr}(A)\sin^2\theta(\mathcal{X},\mathcal{Y})$, and this is supported by Figure 4.1.

In Section 4 the majorization result (4.4) with bound $\mathrm{spr}(A)\sin^2\theta(\mathcal{X},\mathcal{Y})$ was compared with the classical bound $\mathrm{spr}(A)s^2$ in (4.3). It was shown in (4.8) that $s \neq 0 \Rightarrow s^2 \not\prec_w \sin^2\theta(\mathcal{X},\mathcal{Y})$, so that this majorization result always gives added information. It was also seen to give a stronger 1-norm bound in (4.6). From these, the other results here, and [1, 7, 8], we conclude that this majorization approach is worth further study.

Care was taken to illustrate some of the ideas with simple diagrams in $\mathbb{R}^2$.

Acknowledgments. Comments from an erudite and very observant referee helped us to give a nicer and shorter proof for Lemma 3.2, and to improve the presentation in general.

REFERENCES

[1] M. E. ARGENTATI, A. V. KNYAZEV, C. C. PAIGE, AND I. PANAYOTOV, *Bounds on changes in Ritz values for a perturbed invariant subspace of a Hermitian matrix*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 548–559.

[2] R. BHATIA, *Matrix Analysis*, Springer, New York, 1997.

[3] Å. BJÖRCK AND G. H. GOLUB, *Numerical methods for computing angles between linear subspaces*, Math. Comp., 27 (1973), pp. 579–594.

[4] G. H. GOLUB AND R. UNDERWOOD, *The Block Lanczos Method for Computing Eigenvalues*, in Mathematical Software III, J. Rice Ed., Academic Press, New York, 1977, pp. 364–377.

[5] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1996.

[6] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, New York, 1991.

[7] A. V. KNYAZEV AND M. E. ARGENTATI, *Majorization for changes in angles between subspaces, Ritz values, and graph Laplacian spectra*, SIAM J. Matrix Anal. Appl., 29 (2006), pp. 15–32.

[8] A. V. KNYAZEV AND M. E. ARGENTATI, *Rayleigh-Ritz majorization error bounds with applications to FEM and subspace iterations*, submitted for publication January 2007 (http://arxiv.org/abs/math/0701784).

[9] V. B. LIDSKII, *On the proper values of a sum and product of symmetric matrices*, Dokl. Akad. Nauk SSSR, 75 (1950), pp. 769–772.

[10] A. W. MARSHALL AND I. OLKIN, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, 1979.

[11] B. N. PARLETT, *The Symmetric Eigenvalue Problem*, Society for Industrial and Applied Mathematics, Philadelphia, 1998.

[12] A. RUHE, *Numerical methods for the solution of large sparse eigenvalue problems*, in Sparse Matrix Techniques, Lect. Notes Math. 572, V.A. Barker Ed., Springer, Berlin-Heidelberg-New York, 1976, pp. 130–184.

[13] G. W. STEWART AND J.-G. SUN, *Matrix Perturbation Theory*, Academic Press, San Diego, 1990.

[14] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, U.K., 1965.