SIAM J. MATRIX ANAL. APPL.
Vol. 19, No. 4, pp. 956–982, October 1998

009

RELATIVE PERTURBATION THEORY: I. EIGENVALUE AND SINGULAR VALUE VARIATIONS*

REN-CANG LI†

Abstract. The classical perturbation theory for Hermitian matrix eigenvalue and singular value problems provides bounds on the absolute differences between approximate eigenvalues (singular values) and the true eigenvalues (singular values) of a matrix. These bounds may be bad news for small eigenvalues (singular values), which thereby suffer worse relative uncertainty than large ones. However, there are situations where even small eigenvalues are determined to high relative accuracy by the data much more accurately than the classical perturbation theory would indicate. In this paper, we study how eigenvalues of a Hermitian matrix A change when it is perturbed to $\widetilde{A} = D^*AD$, where D is close to a unitary matrix, and how singular values of a (nonsquare) matrix B change when it is perturbed to $\widetilde{B} = D_1^*BD_2$, where D_1 and D_2 are nearly unitary. It is proved that under these kinds of perturbations small eigenvalues (singular values) suffer relative changes no worse than large eigenvalues (singular values). Many well-known perturbation theorems, including the Hoffman–Wielandt and Weyl–Lidskii theorems, are extended.

Key words. multiplicative perturbation, relative perturbation theory, relative distance, eigenvalue, singular value, graded matrix

AMS subject classifications. 15A18, 15A42, 65F15, 65F35, 65G99

PII. S089547989629849X

1. Introduction.

The classical perturbation theory for Hermitian matrix eigenvalue problems provides bounds on the absolute differences $|\lambda - \widetilde{\lambda}|$ between approximate eigenvalues $\widetilde{\lambda}$ and the true eigenvalues λ of a Hermitian matrix A. When $\widetilde{\lambda}$ is computed using standard numerical software, the bounds on $|\lambda - \widetilde{\lambda}|$ are typically only moderately bigger than $\epsilon\|A\|$ [15, 33, 40], where ϵ is the rounding error threshold characteristic of the computer's arithmetic. These bounds are bad news for small eigenvalues, which thereby suffer worse relative uncertainty than large ones.

Generally, the classical error bounds are best possible if perturbations are arbitrary. However, there are situations where perturbations have special structures and, under these special perturbations, even small eigenvalues (singular values) are determined to high relative accuracy by the data much more accurately than the classical perturbation theory would indicate. A relative perturbation theory is then called for to exploit the situations for better bounds on the *relative* differences between $\widetilde{\lambda}$ and λ.

*Received by the editors February 2, 1996; accepted for publication (in revised form) by R. Bhatia June 23, 1997; published electronically July 7, 1998. A preliminary version of this paper appeared as Technical Report UCB//CSD-94-855, Computer Science Division, Department of EECS, University of California at Berkeley, 1994, and also appeared as LAPACK working note 85 (revised January 1996) available online at http://www.netlib.org/lapack/lawns/lawns84.ps. This research was supported in part by Argonne National Laboratory under grant 20552402, by the University of Tennessee through the Advanced Research Projects Agency under contract DAAL03-91-C-0047, by the National Science Foundation under grant ASC-9005933, by National Science Infrastructure grants CDA-8722788 and CDA-9401156, and by a Householder Fellowship in Scientific Computing at Oak Ridge National Laboratory, supported by the Applied Mathematical Sciences Research Program, Office of Energy Research, United States Department of Energy contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp.

http://www.siam.org/journals/simax/19-4/29849.html

†Mathematical Science Section, Oak Ridge National Laboratory, P.O. Box 2008, Bldg. 6012, Oak Ridge, TN 37831-6367. Present address: Department of Mathematics, University of Kentucky, Lexington, KY 40506 (rcli@cs.uky.edu).

The development of such a theory goes back to Kahan [20] and is becoming a very active area of research [1, 6, 7, 8, 9, 11, 12, 14, 16, 10, 28, 34]. In this paper, we develop a theory by a unifying treatment that sharpens some existing bounds and covers many previously studied cases. We shall deal with perturbations that have multiplicative structures; namely, perturbations to unperturbed matrices are realized by multiplying the unperturbed ones with matrices that are nearly unitary. (To be exact, our theorems only require those multiplying matrices to be nonsingular, but our bounds are interesting only when they are close to some unitary matrices.) For Hermitian eigenvalue problems, we shall assume that A is perturbed to $\widetilde{A} = D^* A D$, where D is nonsingular; and for singular value problems we shall consider that B is perturbed to $\widetilde{B} = D_1^* B D_2$, where D_1 and D_2 are nonsingular. It is proved that these kinds of perturbations introduce no bigger uncertainty to small eigenvalues (in magnitude) and small singular values than they would to large ones. Although special, these perturbations cover *componentwise relative perturbations* of entries of symmetric tridiagonal matrices with zero diagonal [8, 20] and *componentwise relative perturbations* of entries of bidiagonal and biacyclic matrices [1, 7, 8]. More realistically, perturbations of graded nonnegative Hermitian matrices [9, 28] and perturbations of graded matrices of singular value problems [9, 28] can be transformed to take forms of multiplicative perturbations as will be seen from later proofs.

Additive perturbations are the most general in the sense that if A is perturbed to $\widetilde{A}$, the only possible known information is on some norm of $\Delta A \stackrel{\text{def}}{=} \widetilde{A} - A$. Such perturbations, no matter how small, may not guarantee relative accuracy in eigenvalues (singular values) of the matrix under consideration. For example, when A is singular, $\widetilde{A}$ can be made nonsingular no matter how small a norm of ΔA is; thus some zero eigenvalues are perturbed to nonzero ones and therefore lose their relative accuracy completely. (Retaining any relative accuracy of a zero at all ends up not changing it.)

The rest of this paper is organized as follows. Section 2 defines two kinds of relative distances ϱ_p $(1 \le p \le \infty)$ and χ, and Appendices A and B present proofs of some crucial properties of ϱ_p and χ needed in this paper. We devote two sections to present and discuss our main theorems—section 3 for relative perturbation theorems for Hermitian matrix eigenvalue problems and section 4 for relative perturbation theorems for singular value problems. Long proofs of our main theorems are postponed to sections 5 and 6. Section 7 briefly discusses how our relative perturbation theorems can be applied to generalized eigenvalue problems and generalized singular value problems.

Notation. We shall adopt the following convention: capital letters denote unperturbed matrices and capital letters with *tildes* denote their perturbed matrices. For example, *X is perturbed to $\widetilde{X}$.* Throughout the paper, capital letters are for matrices, lowercase Latin letters for column vectors or scalars, and lowercase Greek letters for scalars. Also,

$\mathbb{C}^{m\times n}$:	the set of $m \times n$ complex matrices, and $\mathbb{C}^m = \mathbb{C}^{m\times 1}$;
$\mathbb{R}^{m\times n}$:	the set of $m \times n$ real matrices, and $\mathbb{R}^m = \mathbb{R}^{m\times 1}$;
$\mathbb{U}_n$:	the set of $n \times n$ unitary matrices;
$0_{m,n}$:	the $m \times n$ zero matrix (we may simply write 0 instead);
I_n:	the $n \times n$ identity matrix (we may simply write I instead);

X^*:	the conjugate transpose of a matrix X;
$\lambda(X)$:	the set of the eigenvalues of X, counted according to their algebraic multiplicities;
$\sigma(X)$:	the set of the singular values of X, counted according to their algebraic multiplicities;
$\sigma_{\min}(X)$:	the smallest singular value of $X \in \mathbb{C}^{m\times n}$;
$\sigma_{\max}(X)$:	the largest singular value of $X \in \mathbb{C}^{m\times n}$;
$\\|X\\|_2$:	the spectral norm of X, i.e., $\sigma_{\max}(X)$;
$\\|X\\|_{\mathrm{F}}$:	the Frobenius norm of X, i.e., $\sqrt{\sum_{i,j} \lvert x_{ij}\rvert^2}$, where $X = (x_{ij})$.

2. Relative distances. Classically, the relative error in $\widetilde{\alpha} = \alpha(1+\delta)$ as an approximation to α is measured by

$$\delta = \text{relative error in } \widetilde{\alpha} = \frac{\widetilde{\alpha}-\alpha}{\alpha}. \tag{2.1}$$

When $|\delta| \le \epsilon$, we say that the relative perturbation to α is at most ϵ (see, e.g., [8]). Such a measurement lacks mathematical properties upon which a nice relative perturbation theory can be built; for example, it lacks symmetry between α and $\widetilde{\alpha}$ and thus it cannot be a metric. Nonetheless, it is good enough and is convenient to use for measuring correct digits in numerical approximations.

Our new *relative distances* have better mathematical properties, such as symmetry in the arguments. Topologically they are all equivalent to the classical δ-measurement defined by (2.1). The *p-relative distance* between $\alpha, \widetilde{\alpha} \in \mathbb{C}$ is defined as

$$\varrho_p(\alpha,\widetilde{\alpha}) \stackrel{\text{def}}{=} \frac{|\alpha-\widetilde{\alpha}|}{\sqrt[p]{|\alpha|^p + |\widetilde{\alpha}|^p}} \quad \text{for } 1 \le p \le \infty. \tag{2.2}$$

We define, for convenience, $0/0 \stackrel{\text{def}}{=} 0$. ϱ_∞ has been used by Deift et al. [6] to define relative gaps. Another *relative distance* that is of interest to us is

$$\chi(\alpha,\widetilde{\alpha}) \stackrel{\text{def}}{=} \frac{|\alpha-\widetilde{\alpha}|}{\sqrt{|\alpha\widetilde{\alpha}|}}. \tag{2.3}$$

This χ-distance has been used by Barlow and Demmel [1] and Demmel and Veselić [9] to define relative gaps between the spectra of two matrices.

Appendix B will show that ϱ_p $(1 \le p \le \infty)$ *is indeed a metric on* $\mathbb{R}$; see also Li [24]. (We suspect that ϱ_p is a metric on $\mathbb{C}$ also, but we cannot give a proof at this point.) Unfortunately χ violates the triangle inequality and thus cannot be a metric. In fact, one can prove that $\chi(\alpha,\gamma) > \chi(\alpha,\beta) + \chi(\beta,\gamma)$ for $\alpha < \beta < \gamma$; see Lemma 6.1.

We refer the reader to Li [24] for a detailed study of the two relative distances. Here, only properties that are most relevant to our relative perturbation theory will be presented, and those proofs that require little work and seem to be straightforward are omitted. Complicated proofs will be given in Appendix A.

PROPOSITION 2.1 (see [24]). *Let* $\alpha, \widetilde{\alpha} \in \mathbb{R}$.

1. *For* $0 \le \epsilon < 1$,

$$\left|\frac{\widetilde{\alpha}}{\alpha} - 1\right| \le \epsilon \Rightarrow \varrho_p(\alpha,\widetilde{\alpha}) \le \frac{\epsilon}{\sqrt[p]{1+(1-\epsilon)^p}}, \tag{2.4}$$

$$\left|\frac{\widetilde{\alpha}}{\alpha} - 1\right| \le \epsilon \Rightarrow \chi(\alpha,\widetilde{\alpha}) \le \frac{\epsilon}{\sqrt{1-\epsilon}}. \tag{2.5}$$

2. *For* $0 \le \epsilon < 1$,

$$\varrho_p(\alpha, \widetilde{\alpha}) \le \epsilon \Rightarrow \max\left\{\left|\frac{\widetilde{\alpha}}{\alpha} - 1\right|, \left|\frac{\alpha}{\widetilde{\alpha}} - 1\right|\right\} \le \frac{2^{1/p}\,\epsilon}{1-\epsilon}. \tag{2.6}$$

For $0 \le \epsilon < 2$,

$$\chi(\alpha, \widetilde{\alpha}) \le \epsilon \Rightarrow \max\left\{\left|\frac{\widetilde{\alpha}}{\alpha} - 1\right|, \left|\frac{\alpha}{\widetilde{\alpha}} - 1\right|\right\} \le \left(\frac{\epsilon}{2} + \sqrt{1 + \frac{\epsilon^2}{4}}\right)\epsilon. \tag{2.7}$$

3. *Asymptotically,*

$$\lim_{\widetilde{\alpha}\to\alpha} \frac{\varrho_p(\alpha, \widetilde{\alpha})}{\left|\frac{\widetilde{\alpha}}{\alpha} - 1\right|} = 2^{1/p} \quad \textit{and} \quad \lim_{\widetilde{\alpha}\to\alpha} \frac{\chi(\alpha, \widetilde{\alpha})}{\left|\frac{\widetilde{\alpha}}{\alpha} - 1\right|} = 1.$$

Thus (2.4), (2.6), (2.5), *and* (2.7) *are at least asymptotically sharp.*

The following proposition establishes a relation between ϱ_p and χ.

PROPOSITION 2.2 (see [24]). *For* α, $\widetilde{\alpha} \in \mathbb{C}$,

$$\varrho_p(\alpha, \widetilde{\alpha}) \le 2^{-1/p}\,\chi(\alpha, \widetilde{\alpha}),$$

and the equality holds if and only if $|\alpha| = |\widetilde{\alpha}|$.

Next we ask *what are the best one-one pairings between two sets of* n *real numbers*? Such a question will become important later in this paper when we try to pair the eigenvalues or the singular values of one matrix to those of another.

PROPOSITION 2.3 (see [24]). *Let* $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ *and* $\{\widetilde{\alpha}_1, \widetilde{\alpha}_2, \dots, \widetilde{\alpha}_n\}$ *be two sets of* n *real numbers ordered in descending order, i.e.,*

$$\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_n, \qquad \widetilde{\alpha}_1 \ge \widetilde{\alpha}_2 \ge \cdots \ge \widetilde{\alpha}_n. \tag{2.8}$$

We have for $p = 1$,

$$\max_{1\le i\le n} \varrho_1(\alpha_i, \widetilde{\alpha}_i) = \min_{\tau} \max_{1\le i\le n} \varrho_1(\alpha_i, \widetilde{\alpha}_{\tau(i)}).$$

For $p > 1$, *if in addition all* α_i*'s and* $\widetilde{\alpha}_j$*'s are nonnegative,*

$$\max_{1\le i\le n} \varrho_p(\alpha_i, \widetilde{\alpha}_i) = \min_{\tau} \max_{1\le i\le n} \varrho_p(\alpha_i, \widetilde{\alpha}_{\tau(i)}). \tag{2.9}$$

Both minimizations are taken over all permutations τ *of* $\{1, 2, \dots, n\}$.

Proofs of this proposition and Proposition 2.4 below are given in Appendix A.

Remark 2.1. Equation (2.9) of Proposition 2.3 may fail if not all the α_i's and $\widetilde{\alpha}_j$'s are of the same sign. A *counterexample* is as follows: $n = 2$ and

$$\alpha_1 = 1 > \alpha_2 = -2 \quad \text{and} \quad \widetilde{\alpha}_1 = 4 > \widetilde{\alpha}_2 = 2.$$

Then for $p > 1$,

$$\begin{aligned} &\max\{\varrho_p(\alpha_1, \widetilde{\alpha}_1),\ \varrho_p(\alpha_2, \widetilde{\alpha}_2)\} = \varrho_p(\alpha_2, \widetilde{\alpha}_2) = 2^{1-1/p} \\ &> \frac{6}{\sqrt[p]{2^p + 4^p}} = \varrho_p(\alpha_2, \widetilde{\alpha}_1) = \max\{\varrho_p(\alpha_1, \widetilde{\alpha}_2),\ \varrho_p(\alpha_2, \widetilde{\alpha}_1)\}. \end{aligned}$$

Remark 2.2. Given two sets of α_i's and $\widetilde{\alpha}_j$'s ordered as in (2.8), generally,

$$\sum_{i=1}^{n} [\varrho_p(\alpha_i, \widetilde{\alpha}_i)]^2 \neq \min_{\tau} \sum_{i=1}^{n} \left[\varrho_p(\alpha_i, \widetilde{\alpha}_{\tau(i)})\right]^2, \tag{2.10}$$

even if all α_i, $\widetilde{\alpha}_j > 0$. Here is a *counterexample*: $n = 2$,

$$\widetilde{\alpha}_1 > \alpha_1 = \widetilde{\alpha}_1/2 > \widetilde{\alpha}_2 > \alpha_2 > 0,$$

where α_2 is sufficiently close to 0, and $\widetilde{\alpha}_2$ is sufficiently close to α_1 which is fixed. Since, as $\alpha_2 \to 0^+$ and $\widetilde{\alpha}_2 \to \alpha_1^-$,

$$\begin{aligned} &[\varrho_p(\alpha_1, \widetilde{\alpha}_2)]^2 + [\varrho_p(\alpha_2, \widetilde{\alpha}_1)]^2 \to 1, \\ &[\varrho_p(\alpha_1, \widetilde{\alpha}_1)]^2 + [\varrho_p(\alpha_2, \widetilde{\alpha}_2)]^2 \to \frac{1}{\sqrt[p]{2^p + 1}} + 1, \end{aligned}$$

(2.10) must fail for some $\widetilde{\alpha}_1 > \alpha_1 = \widetilde{\alpha}_1/2 > \widetilde{\alpha}_2 > \alpha_2 > 0$.

PROPOSITION 2.4 (see [24]). *Let* $\{\alpha_1, \ldots, \alpha_n\}$ *and* $\{\widetilde{\alpha}_1, \ldots, \widetilde{\alpha}_n\}$ *be two sets of* n *positive numbers ordered as in* (2.8). *Then*

$$\max_{1 \le i \le n} \chi(\alpha_i, \widetilde{\alpha}_i) = \min_{\tau} \max_{1 \le i \le n} \chi(\alpha_i, \widetilde{\alpha}_{\tau(i)}), \tag{2.11}$$

$$\sum_{i=1}^{n} [\chi(\alpha_i, \widetilde{\alpha}_i)]^2 = \min_{\tau} \sum_{i=1}^{n} \left[\chi(\alpha_i, \widetilde{\alpha}_{\tau(i)})\right]^2, \tag{2.12}$$

where the minimization is taken over all permutations τ *of* $\{1, 2, \ldots, n\}$.

Remark 2.3. Both (2.11) and (2.12) of Proposition 2.4 may fail if the α_i's and $\widetilde{\alpha}_j$'s are not all of the same sign. A *counterexample* for (2.11) is that $n = 2$ and

$$\alpha_1 = 1 > \alpha_2 = -1 \quad \text{and} \quad \widetilde{\alpha}_1 = 2 > \widetilde{\alpha}_2 = \frac{1}{4},$$

for which

$$\begin{aligned} &\max\{\chi(\alpha_1, \widetilde{\alpha}_1),\, \chi(\alpha_2, \widetilde{\alpha}_2)\} = \max\left\{1/\sqrt{2}, 5/2\right\} = 5/2 \\ &> 3/\sqrt{2} = \max\left\{3/2, 3/\sqrt{2}\right\} = \max\{\chi(\alpha_1, \widetilde{\alpha}_2),\, \chi(\alpha_2, \widetilde{\alpha}_1)\}. \end{aligned}$$

A *counterexample* for (2.12) is that $n = 2$ and

$$\alpha_1 = 1 > \alpha_2 = -2 \quad \text{and} \quad \widetilde{\alpha}_1 = 2 > \widetilde{\alpha}_2 = 1,$$

for which

$$\begin{aligned} [\chi(\alpha_1, \widetilde{\alpha}_1)]^2 + [\chi(\alpha_2, \widetilde{\alpha}_2)]^2 &= \left(1/\sqrt{2}\right)^2 + \left(3/\sqrt{2}\right)^2 = 5 \\ > 4 = 0^2 + \left(4/\sqrt{4}\right)^2 &= [\chi(\alpha_1, \widetilde{\alpha}_2)]^2 + [\chi(\alpha_2, \widetilde{\alpha}_1)]^2. \end{aligned}$$

3. Relative perturbation theorems for Hermitian matrix eigenvalue problems. Throughout the section, A, $\widetilde{A} \in \mathbb{C}^{n\times n}$ are Hermitian and one is a perturbation of the other. Denote their eigenvalues by

$$\lambda(A) = \{\lambda_1, \dots, \lambda_n\} \quad \text{and} \quad \lambda(\widetilde{A}) = \{\widetilde{\lambda}_1, \dots, \widetilde{\lambda}_n\} \tag{3.1}$$

ordered so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n, \quad \widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq \cdots \geq \widetilde{\lambda}_n. \tag{3.2}$$

THEOREM 3.1. *Let A and $\widetilde{A} = D^*AD$ be two $n \times n$ Hermitian matrices with eigenvalues* (3.1) *ordered as in* (3.2)*, where D is nonsingular. Then*

1. *there is a permutation τ of $\{1, 2, \dots, n\}$ such that*

$$\sqrt{\sum_{i=1}^{n} \left[\varrho_2(\lambda_i, \widetilde{\lambda}_{\tau(i)})\right]^2} \leq \sqrt{\|I - \Sigma_{\rm d}\|_{\rm F}^2 + \|I - \Sigma_{\rm d}^{-1}\|_{\rm F}^2}, \tag{3.3}$$

where $\Sigma_{\rm d}$ is diagonal and its diagonal entries are D's singular values.

2. *if, in addition, A is nonnegative definite,*[1] *then*

$$\max_{1\leq i\leq n} \chi(\lambda_i, \widetilde{\lambda}_i) \leq \|D^* - D^{-1}\|_2, \tag{3.4}$$

$$\sqrt{\sum_{i=1}^{n} \left[\chi(\lambda_i, \widetilde{\lambda}_i)\right]^2} \leq \|D^* - D^{-1}\|_{\rm F}. \tag{3.5}$$

A proof of Theorem 3.1 will be given in section 5.

A corollary of (3.3) is

$$\sqrt{\sum_{i=1}^{n} \left[\varrho_2(\lambda_i, \widetilde{\lambda}_{\tau(i)})\right]^2} \leq \sqrt{\|I - D\|_{\rm F}^2 + \|I - D^{-1}\|_{\rm F}^2} \tag{3.3a}$$

by a well-known (absolute) perturbation theorem for singular values; see (4.7). On the other hand, (3.3a) leads to (3.3) as well by considering $U_{\rm d}^*AU_{\rm d}$ and $V_{\rm d}^*\widetilde{A}V_{\rm d} = \Sigma_{\rm d}(U_{\rm d}^*AU_{\rm d})\Sigma_{\rm d}$ instead, where

$$D = U_{\rm d}\Sigma_{\rm d}V_{\rm d}^* \tag{3.6}$$

is D's singular value decomposition (SVD) [15, p. 71]. It is also possible to relate the right-hand sides of (3.4) and (3.5) to the singular values of D, since for every unitarily invariant norm[2] $|||\cdot|||$,

$$|||D^* - D^{-1}||| = |||V_{\rm d}(\Sigma_{\rm d} - \Sigma_{\rm d}^{-1})U_{\rm d}^*||| = |||\Sigma_{\rm d} - \Sigma_{\rm d}^{-1}|||.$$

---

[1] Then $\widetilde{A}$ must be nonnegative definite as well.

[2] In this we follow Mirsky [30], Stewart and Sun [35], and Bhatia [3]. That a norm $|||\cdot|||$ is *unitarily invariant* on $\mathbb{C}^{m\times n}$ means that it also satisfies, besides the usual properties of any norm,

1. $|||UYV||| = |||Y|||$, for any $U \in \mathbb{U}_m$, and $V \in \mathbb{U}_n$;
2. $|||Y||| = \|Y\|_2$, for any $Y \in \mathbb{C}^{m\times n}$ with $\mathrm{rank}(Y) = 1$.

Two unitarily invariant norms most frequently used are the *spectral norm* $\|\cdot\|_2$ and the *Frobenius norm* $\|\cdot\|_{\rm F}$. Let $|||\cdot|||$ be a unitarily invariant norm on some matrix space. The following inequalities [35, p. 80] will be employed later in this paper:

$$|||WY||| \leq \|W\|_2|||Y||| \quad \text{and} \quad |||YZ||| \leq |||Y|||\|Z\|_2.$$

The earliest relative perturbation result for eigenvalue problems goes back to a theorem due to Ostrowski [32] (see also [18, pp. 224–225]), though he did not interpret his theorem in the way we do now. Ostrowski proved that

for two $n \times n$ Hermitian matrices A and $\widetilde{A} = D^* A D$ with eigenvalues (3.1) *ordered as in* (3.2), *where D is nonsingular, we have*

$$\sigma_{\min}(D)^2 \cdot \lambda_i \le \widetilde{\lambda}_i \le \sigma_{\max}(D)^2 \cdot \lambda_i \quad \textit{for} \ \le i \le n. \tag{3.7}$$

Inequalities (3.7) immediately imply a relative perturbation bound

$$\max_{1 \le i \le n} \frac{|\widetilde{\lambda}_i - \lambda_i|}{|\lambda_i|} \le \|I - D^* D\|_2.$$

This result of Ostrowski's is independent of (3.4). Both may be attainable for the scalar case ($n = 1$) or for the case when A and D are diagonal. Our bounds (3.3) and (3.5) are the first of their kind.

Roughly speaking, the classical perturbation theory for Hermitian matrix eigenvalue problems establishes one uniform bound for all differences $|\lambda_i - \widetilde{\lambda}_i|$ regardless of magnitudes of λ_i's. In this regard, we have the following.

Let both A and $\widetilde{A}$ be Hermitian. (*No special form of $\widetilde{A}$ is assumed.*) *Then for any unitarily invariant norm $|||\cdot|||$*,

$$|||\mathrm{diag}(\lambda_1 - \widetilde{\lambda}_1, \dots, \lambda_n - \widetilde{\lambda}_n)||| \le |||A - \widetilde{A}|||. \tag{3.8}$$

There is a long history associated with this inequality; see Bhatia [3] for details. Theorem 3.1 extends (3.8) to the relative perturbation theory for $|||\cdot||| = \|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$. Two main differences between Theorem 3.1 and (3.8) are as follows.

1. Inequality (3.8) bounds the absolute differences $|\lambda_i - \widetilde{\lambda}_i|$. It is in fact the best possible as far as arbitrary perturbations are concerned. However, it may overestimate the differences $|\lambda_j - \widetilde{\lambda}_j|$ too much for eigenvalues λ_j of much smaller magnitudes than $\|A\|_2$ when perturbations have special structures such as multiplicative perturbations, for which it is possible that $|||A - \widetilde{A}|||$ is larger than $|\lambda_j - \widetilde{\lambda}_j|$ by many orders of magnitudes while, on the other hand, $D^* D \approx I$.
2. Theorem 3.1 exploits fully multiplicative perturbation structures by bounding directly the relative differences $\chi(\lambda_i, \widetilde{\lambda}_i)$ or $\varrho_2(\lambda_i, \widetilde{\lambda}_i)$ in terms of D's departures from unitary matrices $|||D^* - D^{-1}|||$ and $\sqrt{\|I - \Sigma_d\|_{\mathrm{F}}^2 + \|I - \Sigma_d^{-1}\|_{\mathrm{F}}^2}$. Thus, all eigenvalues of the same as or much smaller magnitudes than $\|A\|_2$ alike provably suffer small uncertainty as long as D's departures from unitary matrices are small.

Such arguments more or less apply to our other relative perturbation theorems in this paper in comparison to their counterparts in the classical absolute perturbation theory.

In Theorem 3.1, the perturbation to A is rather restrictive but *is* applicable to a more realistic situation when scaled A is much better conditioned. In Theorem 3.2, S is a scaling matrix, often highly graded and diagonal in practice, though the theorem does not assume this.

THEOREM 3.2. *Let $A = S^* H S$ and $\widetilde{A} = S^* \widetilde{H} S$ be two $n \times n$ nonnegative definite Hermitian matrices with eigenvalues* (3.1) *ordered as in* (3.2), *and let $\Delta H = \widetilde{H} - H$.*

If $\|H^{-1}\|_2\|\Delta H\|_2 < 1$, *then*

$$\max_{1\le i\le n} \chi(\lambda_i, \widetilde{\lambda}_i) \le \|D - D^{-1}\|_2, \tag{3.9}$$

$$\le \frac{\|H^{-1}\|_2\|\Delta H\|_2}{\sqrt{1 - \|H^{-1}\|_2\|\Delta H\|_2}}, \tag{3.10}$$

$$\sqrt{\sum_{i=1}^{n} \left[\chi(\lambda_i, \widetilde{\lambda}_i)\right]^2} \le \|D - D^{-1}\|_{\mathrm{F}}, \tag{3.11}$$

$$\le \frac{\|H^{-1}\|_2\|\Delta H\|_{\mathrm{F}}}{\sqrt{1 - \|H^{-1}\|_2\|\Delta H\|_2}}, \tag{3.12}$$

where $D = \big(I + H^{-1/2}(\Delta H)H^{-1/2}\big)^{1/2}$.

Proof. Rewrite A and $\widetilde{A}$ as

$$\begin{aligned} A &= S^*HS = (H^{1/2}S)^*\, H^{1/2}S, \\ \widetilde{A} &= S^*H^{1/2}\big(I + H^{-1/2}(\Delta H)H^{-1/2}\big)H^{1/2}S \\ &= \Big(\big(I + H^{-1/2}(\Delta H)H^{-1/2}\big)^{1/2}H^{1/2}S\Big)^*\ \big(I + H^{-1/2}(\Delta H)H^{-1/2}\big)^{1/2}H^{1/2}S. \end{aligned}$$

Set $B \stackrel{\text{def}}{=} H^{1/2}S$ and $\widetilde{B} \stackrel{\text{def}}{=} \big(I + H^{-1/2}(\Delta H)H^{-1/2}\big)^{1/2}H^{1/2}S$, then $A = B^*B$ and $\widetilde{A} = \widetilde{B}^*\widetilde{B}$. We have $\widetilde{B} = DB$, where $D = \big(I + H^{-1/2}(\Delta H)H^{-1/2}\big)^{1/2}$. Notice that

$$\lambda(A) = \lambda(B^*B) = \lambda(BB^*) \quad \text{and} \quad \lambda(\widetilde{A}) = \lambda(\widetilde{B}^*\widetilde{B}) = \lambda(\widetilde{B}\widetilde{B}^*),$$

and $\widetilde{B}\widetilde{B}^* = DBB^*D^*$. Applying Theorem 3.1 to BB^* and $\widetilde{B}\widetilde{B}^*$ yields both (3.9) and (3.11). Inequalities (3.10) and (3.12) follow from the fact that for any Hermitian matrix E with $\|E\|_2 < 1$ and for any unitarily invariant norm $|||\cdot|||$,

$$|||(I+E)^{1/2} - (I+E)^{-1/2}||| \le \|(I+E)^{-1/2}\|_2 |||E||| \le \frac{|||E|||}{\sqrt{1 - \|E\|_2}}. \qquad \square$$

Inequality (3.10) can also be derived from the following bound essentially due to Demmel and Veselić [9] (see also Mathias [28]).

Let the conditions of Theorem 3.2 *hold. Then*

$$\max_{1\le i\le n} \frac{|\widetilde{\lambda}_i - \lambda_i|}{|\lambda_i|} \le \|H^{-1}\|_2\|\Delta H\|_2. \tag{3.13}$$

To see how (3.13) leads to (3.10), we notice that[3]

$$\chi(\lambda_i, \widetilde{\lambda}_i) = \frac{|\widetilde{\lambda}_i - \lambda_i|}{|\lambda_i|} \cdot \sqrt{\frac{\lambda_i}{\widetilde{\lambda}_i}} \le \frac{|\widetilde{\lambda}_i - \lambda_i|}{|\lambda_i|} \cdot \|D^{-1}\|_2$$

by Ostrowski's theorem (3.7) and that $\|D^{-1}\|_2 \le 1/\sqrt{1 - \|H^{-1}\|_2\|\Delta H\|_2}$.

Remark 3.1. Li [24] also considered extending Theorem 3.1 to diagonalizable matrices under multiplicative perturbations. But the bounds obtained in a recent paper [26] are better. Both Li [24] and Eisenstat and Ipsen [13] extended the classical Bauer–Fike theorem [2].

---

[3] $\lambda_i = 0$ if and only if $\widetilde{\lambda}_i = 0$, since A and $\widetilde{A}$ have the same number of zero eigenvalues, if any. So we only need to consider those i such that $\lambda_i \ne 0$.

4. Relative perturbation theorems for singular value problems. Throughout the section, B, $\widetilde{B} \in \mathbb{C}^{m\times n}$ and one is a perturbation of the other. (We shall assume, without loss of generality, that $m \geq n$ in this section.) Denote their singular values by

$$\sigma(B) = \{\sigma_1, \dots, \sigma_n\} \quad \text{and} \quad \sigma(\widetilde{B}) = \{\widetilde{\sigma}_1, \dots, \widetilde{\sigma}_n\} \tag{4.1}$$

ordered so that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0, \qquad \widetilde{\sigma}_1 \geq \widetilde{\sigma}_2 \geq \cdots \geq \widetilde{\sigma}_n \geq 0. \tag{4.2}$$

THEOREM 4.1. *Let B and $\widetilde{B} = D_1^* B D_2$ be two $m \times n$ matrices with singular values* (4.1) *ordered as in* (4.2), *where D_1 and D_2 are square and nonsingular. If $\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2 < 32$, then*

$$\max_{1\leq i\leq n} \chi(\sigma_i, \widetilde{\sigma}_i) \leq \frac{1}{2} \cdot \frac{\|D_1^* - D_1^{-1}\|_2 + \|D_2^* - D_2^{-1}\|_2}{1 - \frac{1}{32}\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2}, \tag{4.3}$$

$$\sqrt{\sum_{i=1}^{n} [\chi(\sigma_i, \widetilde{\sigma}_i)]^2} \leq \frac{1}{2} \cdot \frac{\|D_1^* - D_1^{-1}\|_{\mathrm{F}} + \|D_2^* - D_2^{-1}\|_{\mathrm{F}}}{1 - \frac{1}{32}\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2}. \tag{4.4}$$

A proof of Theorem 4.1 will be given in section 6.

The restriction $\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2 < 32$, though mild, is unpleasant. But we argue that neither this restriction nor the factor $\left(1 - \frac{1}{32}\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2\right)^{-1}$ plays any visible role for any applications where one might expect that perturbing B to $\widetilde{B} = D_1^* B D_2$ retains any significant digits of B's singular values. Our arguments go as follows.

1. For the ease of explanation, consider the case when B and D_j are diagonal. In order for each of B's singular values to have at least one significant decimal digit the same as that of the corresponding $\widetilde{B}$'s, it is necessary that[4]

$$0.9 \leq \sigma_{\min}(D_j) \leq \sigma_{\max}(D_j) \leq 1.05 \tag{4.5}$$

which imply that $\|D_j^* - D_j^{-1}\|_2 \leq 0.2$, and thus the factor

$$\left(1 - \frac{1}{32}\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2\right)^{-1} \leq 1.01.$$

2. In fact, the restriction $\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2 < 32$ is satisfied and the factor is almost 1 even for D_j's singular values being fairly away from 1. It can be seen that

$$\|D_j^* - D_j^{-1}\|_2 \leq 1 \;\text{ if }\; 0.618 \approx \frac{\sqrt{5}-1}{2} \leq \sigma_{\min}(D_j) \leq \sigma_{\max}(D_j) \leq \frac{\sqrt{5}+1}{2} \approx 1.618,$$

under which circumstances the unpleasant factor is

$$\left(1 - \frac{1}{32}\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2\right)^{-1} \leq 32/31 \approx 1.03.$$

[4]This is for the worse case in the sense that if (4.5) is violated, then there are D_j's such that some of the B's singular values retain no significant decimal digits at all under the perturbations.

3. In applications where $\|D_j^* - D_j^{-1}\|_2 \ll 1$, the quantity $\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2$ is of second order. Then the restriction and the factor act as if they were not there. Even more in some applications, as in Corollary 4.2, one of the D_j's is I for which the restriction and the factor disappear completely.
Eisenstat and Ipsen [12] obtained the following result which is essentially a consequence of Ostrowski's theorem (see inequalities (3.7)) and which can also be seen from known inequalities for singular values of a product of two matrices:[5]

Let the conditions of Theorem 4.1, *except* $\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2 < 32$, *hold. We have*

$$\sigma_{\min}(D_1)\sigma_{\min}(D_2) \cdot \sigma_i \le \widetilde{\sigma}_i \le \sigma_{\max}(D_1)\sigma_{\max}(D_2) \cdot \sigma_i \quad \text{for } 1 \le i \le n. \tag{4.6}$$

Inequalities (4.6) imply immediately the following relative perturbation bound:

$$\max_{1 \le i \le n} \frac{|\widetilde{\sigma}_i - \sigma_i|}{\sigma_i} \le \max\{|1 - \sigma_{\min}(D_1)\sigma_{\min}(D_2)|,\, |1 - \sigma_{\max}(D_1)\sigma_{\max}(D_2)|\}.$$

The classical perturbation theory for singular value problems establishes one uniform bound for all differences $\sigma_i - \widetilde{\sigma}_i$, regardless of magnitudes of σ_i's. The following theorem was established by Mirsky [30], based on results from Lidskii [27] and Wielandt [39].

For any unitarily invariant norm $|||\cdot|||$, *we have*

$$|||\mathrm{diag}(\sigma_1 - \widetilde{\sigma}_1, \ldots, \sigma_n - \widetilde{\sigma}_n)||| \le |||B - \widetilde{B}|||. \tag{4.7}$$

(*No special form of* $\widetilde{B}$ *is assumed.*)

A possible application of Theorem 4.1 is related to *deflation* in computing SVD of a bidiagonal matrix. For more details, the reader is referred to [6, 8, 12, 29].

Corollary 4.2. *Assume, in Theorem* 4.1, *that one of* D_1 *and* D_2 *is the identity matrix and the other takes the form*

$$D = \begin{pmatrix} I & X \\ & I \end{pmatrix},$$

where X *is a matrix of suitable dimensions. Then*

$$\max_{1 \le i \le n} \chi(\sigma_i, \widetilde{\sigma}_i) \le \frac{1}{2}\|X\|_2, \tag{4.8}$$

$$\sqrt{\sum_{i=1}^{n} [\chi(\sigma_i, \widetilde{\sigma}_i)]^2} \le \frac{1}{\sqrt{2}}\|X\|_{\mathrm{F}}. \tag{4.9}$$

Proof. Notice that

$$D^* - D^{-1} = \begin{pmatrix} I & \\ X^* & I \end{pmatrix} - \begin{pmatrix} I & -X \\ & I \end{pmatrix} = \begin{pmatrix} & X \\ X^* & \end{pmatrix},$$

[5] Arranging the singular values of a matrix in the decreasing order, we have (see, e.g., [19])

$$(\text{the } i\text{th singular value of } XY) \le (\text{the } i\text{th singular value of } X) \cdot \|Y\|_2.$$

and thus $\|D^* - D^{-1}\|_2 = \|X\|_2$ and $\|D^* - D^{-1}\|_{\rm F} = \sqrt{2}\|X\|_{\rm F}$. $\square$

Eisenstat and Ipsen [12] showed that

$$|\widetilde{\sigma}_i - \sigma_i| \le \|X\|_2 \sigma_i, \quad \text{or equivalently} \quad \left|\frac{\widetilde{\sigma}_i}{\sigma_i} - 1\right| \le \|X\|_2. \tag{4.10}$$

Our inequality (4.8) is sharper by roughly a factor of $1/2$, as long as $\|X\|_2$ is small. As a matter of fact, it follows from (4.8) and Proposition 2.1 that if $\|X\|_2 < 4$, then

$$\left|\frac{\widetilde{\sigma}_i}{\sigma_i} - 1\right| \le \left(\frac{\|X\|_2}{4} + \sqrt{1 + \frac{\|X\|_2^2}{16}}\right)\frac{\|X\|_2}{2} = \frac{\|X\|_2}{2} + O\left(\left(\frac{\|X\|_2}{4}\right)^2\right).$$

Our inequality (4.9) is the first of its kind.

THEOREM 4.3. *Let B and $\widetilde{B} = D_1^* B D_2$ be two $m \times n$ matrices with singular values* (4.1) *ordered as in* (4.2)*, where D_1 and D_2 are square and nonsingular. Then*

$$\max_{1 \le i \le n} \varrho_p(\sigma_i, \widetilde{\sigma}_i) \le \frac{1}{2^{1+1/p}}\left(\|D_1^* - D_1^{-1}\|_2 + \|D_2^* - D_2^{-1}\|_2\right), \tag{4.11}$$

$$\sqrt{\sum_{i=1}^n [\varrho_p(\sigma_i, \widetilde{\sigma}_i)]^2} \le \frac{1}{2^{1+1/p}}\left(\|D_1^* - D_1^{-1}\|_{\rm F} + \|D_2^* - D_2^{-1}\|_{\rm F}\right). \tag{4.12}$$

A straightforward combination of Proposition 2.2 and Theorem 4.1 will lead to bounds that are slightly weaker than those in Theorem 4.3 by a factor of

$$\left(1 - \frac{1}{32}\|D_1^* - D_1^{-1}\|_2 \|D_2^* - D_2^{-1}\|_2\right)^{-1}.$$

A proof of Theorem 4.3 will be given in section 6.

Again we shall now consider a more realistic situation when scaled B is much better conditioned. In Theorem 4.4 below, S is a scaling matrix, often highly graded and diagonal in practice, though the theorem does not assume this.

THEOREM 4.4. *Let $B = GS$ and $\widetilde{B} = \widetilde{G}S$ be two $n \times n$ matrices with singular values* (4.1) *ordered as in* (4.2)*, where G and $\widetilde{G}$ are nonsingular, and let $\Delta G = \widetilde{G} - G$. If $\|\Delta G\|_2 \|G^{-1}\|_2 < 1$, then*

$$\max_{1 \le i \le n} \chi(\sigma_i, \widetilde{\sigma}_i) \le \frac{1}{2}\left\|\left(I + (\Delta G)G^{-1}\right)^* - \left(I + (\Delta G)G^{-1}\right)^{-1}\right\|_2, \tag{4.13}$$

$$\le \left(1 + \frac{1}{1 - \|G^{-1}\|_2\|\Delta G\|_2}\right)\frac{\|G^{-1}\|_2\|\Delta G\|_2}{2}, \tag{4.14}$$

$$\sqrt{\sum_{i=1}^n [\chi(\sigma_i, \widetilde{\sigma}_i)]^2} \le \frac{1}{2}\left\|\left(I + (\Delta G)G^{-1}\right)^* - \left(I + (\Delta G)G^{-1}\right)^{-1}\right\|_{\rm F}, \tag{4.15}$$

$$\le \left(1 + \frac{1}{1 - \|G^{-1}\|_2\|\Delta G\|_2}\right)\frac{\|G^{-1}\|_2\|\Delta G\|_{\rm F}}{2}. \tag{4.16}$$

Proof. Write

$$\widetilde{B} = (G + \Delta G)S = (I + (\Delta G)G^{-1})GS = DB, \tag{4.17}$$

where $D = I + (\Delta G)G^{-1}$. Now, applying Theorem 4.1 to B and $\widetilde{B} = DB$ yields both (4.13) and (4.15). We notice that

$$(I+E)^* - (I+E)^{-1} = I + E^* - \sum_{i=0}^{\infty}(-1)^i E^i = E^* + E + E\sum_{i=2}^{\infty}(-1)^i E^{i-1},$$

where $E = (\Delta G)G^{-1}$ and $\|E\|_2 \le \|G^{-1}\|_2\|\Delta G\|_2 < 1$; therefore, for any unitarily invariant norm $|||\cdot|||$,

$$|||(I+E)^* - (I+E)^{-1}||| \le |||E + E^*||| + |||E|||\sum_{i=1}^{\infty}\|E\|_2^i$$

$$= \left(\frac{|||E+E^*|||}{|||E|||} + \frac{\|E\|_2}{1-\|E\|_2}\right)|||E||| \tag{4.18}$$

$$\le \left(1 + \frac{1}{1-\|E\|_2}\right)|||E|||. \tag{4.19}$$

An application of (4.19) for $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$ completes the proof. □

Equation (4.17) also makes (4.6) applicable and leads to the following.

Let the conditions of Theorem 4.4 *hold. We have*

$$\max_{1\le i\le n}\frac{|\widetilde{\sigma}_i - \sigma_i|}{\sigma_i} \le \|G^{-1}\|_2\|\Delta G\|_2. \tag{4.20}$$

This inequality also follows from [10, Theorem 1.1]. Inequality (4.14) can actually be derived from (4.20) as follows. Notice that

$$\chi(\sigma_i, \widetilde{\sigma}_i) = \frac{|\widetilde{\sigma}_i - \sigma_i|}{|\sigma_i|}\cdot\sqrt{\frac{\sigma_i}{\widetilde{\sigma}_i}} \le \frac{|\widetilde{\sigma}_i - \sigma_i|}{|\sigma_i|}\cdot\|D^{-1}\|_2^{1/2},$$

and that

$$\|D^{-1}\|_2^{1/2} \le \frac{1}{\sqrt{1-\|G^{-1}\|_2\|\Delta G\|_2}} \le \frac{1}{2}\left(1 + \frac{1}{1-\|G^{-1}\|_2\|\Delta G\|_2}\right).$$

Remark 4.1. When $(\Delta G)G^{-1}$ is nearly skew Hermitian, (4.13) and (4.15) lead to bounds that are much better than (4.14) and (4.16). This can be seen from (4.18):

Under the conditions of Theorem 4.4*, we have*

$$\max_{1\le i\le n}\chi(\sigma_i,\widetilde{\sigma}_i) \le \left(\frac{\|(\Delta G)G^{-1} + G^{-*}(\Delta G)^*\|_2}{\|(\Delta G)G^{-1}\|_2} + \frac{\|(\Delta G)G^{-1}\|_2}{1-\|(\Delta G)G^{-1}\|_2}\right)\frac{\|(\Delta G)G^{-1}\|_2}{2},$$

$$\sqrt{\sum_{i=1}^{n}[\chi(\sigma_i,\widetilde{\sigma}_i)]^2} \le \left(\frac{\|(\Delta G)G^{-1} + G^{-*}(\Delta G)^*\|_{\mathrm{F}}}{\|(\Delta G)G^{-1}\|_{\mathrm{F}}} + \frac{\|(\Delta G)G^{-1}\|_2}{1-\|(\Delta G)G^{-1}\|_2}\right)\frac{\|(\Delta G)G^{-1}\|_{\mathrm{F}}}{2}.$$

Now if $(\Delta G)G^{-1}$ is nearly skew Hermitian, then $\chi(\sigma_i,\widetilde{\sigma}_i) = o\left(\|(\Delta G)G^{-1}\|_2\right)$; moreover,

$$\|(\Delta G)G^{-1} + G^{-*}(\Delta G)^*\|_2 = O\left(\|(\Delta G)G^{-1}\|_2^2\right) \Rightarrow \chi(\sigma_i,\widetilde{\sigma}_i) = O\left(\|(\Delta G)G^{-1}\|_2^2\right).$$

Remark 4.2. Theorem 4.4 can be extended to nonsquare matrices. Assume $B = GS$ and $\widetilde{B} = \widetilde{G}S$ are $m \times n$ ($m \geq n$); S is a scaling matrix and both G and $\widetilde{G}$ are $m \times n$; G has full column rank. Let $G^{\dagger} = (G^*G)^{-1}G^*$ be the pseudo-inverse of G. Notice that $G^{\dagger}G = I$. We have

$$\widetilde{B} = \widetilde{G}S = (G + \Delta G)S = (I + (\Delta G)G^{\dagger})GS = (I + (\Delta G)G^{\dagger})B \equiv DB.$$

Now, apply Theorem 4.1 to B and $\widetilde{B} = DB$.

5. Proof of Theorem 3.1. We need a little preparation first. A matrix $Z = (z_{ij}) \in \mathbb{R}^{n \times n}$ is *doubly stochastic* if all $z_{ij} \geq 0$ and

$$\sum_{k=1}^{n} z_{ik} = \sum_{k=1}^{n} z_{kj} = 1 \quad \text{for } i, j = 1, 2, \dots, n.$$

Using a Birkhoff theorem [4] (see also [18, pp. 527–528]) and the technique of Hoffman and Wielandt [17] (see also [35, p. 190]), we can prove the following.

LEMMA 5.1. *Let $Z = (z_{ij})$ be an $n \times n$ doubly stochastic matrix, and let $M = (m_{ij}) \in \mathbb{C}^{n \times n}$. Then there exists a permutation τ of $\{1, 2, \dots, n\}$ such that*

$$\sum_{i,\,j=1}^{n} |m_{ij}| z_{ij} \geq \sum_{i=1}^{n} |m_{i\tau(i)}|.$$

For $X \in \mathbb{C}^{m \times n}$, we introduce the following notation for a $k \times \ell$ submatrix of $X = (x_{ij})$:

$$X \begin{pmatrix} i_1 \dots i_k \\ j_1 \dots j_\ell \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} x_{i_1 j_1} & x_{i_1 j_2} & \cdots & x_{i_1 j_\ell} \\ x_{i_2 j_1} & x_{i_2 j_2} & \cdots & x_{i_2 j_\ell} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i_k j_1} & x_{i_k j_2} & \cdots & x_{i_k j_\ell} \end{pmatrix}, \tag{5.1}$$

where $1 \leq i_1 < \cdots < i_k \leq n$ and $1 \leq j_1 < \cdots < j_\ell \leq n$. The following lemma is due to Li [22, pp. 207–208]

LEMMA 5.2 (see Li [22]). *Suppose that $X \in \mathbb{C}^{n \times n}$ is nonsingular, and $1 \leq i_1 < \cdots < i_k \leq n$ and $1 \leq j_1 < \cdots < j_\ell \leq n$, and $k + \ell > n$. Then*

$$\left\| X \begin{pmatrix} i_1 \cdots i_k \\ j_1 \cdots j_\ell \end{pmatrix} \right\|_2 \geq \|X^{-1}\|_2^{-1}.$$

Moreover, if X is unitary, then

$$\left\| X \begin{pmatrix} i_1 \cdots i_k \\ j_1 \cdots j_\ell \end{pmatrix} \right\|_2 = 1.$$

Proof of Theorem 3.1. We shall prove (3.3) first. Due to the argument we made right after Theorem 3.1, it suffices for us to prove (3.3a). Let the eigen decompositions of A and $\widetilde{A}$ be

$$A = U \Lambda U^* \quad \text{and} \quad \widetilde{A} = \widetilde{U} \widetilde{\Lambda} \widetilde{U}^*,$$

where U and $\widetilde{U}$ are unitary and $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $\widetilde{\Lambda} = \operatorname{diag}(\widetilde{\lambda}_1, \widetilde{\lambda}_2, \ldots, \widetilde{\lambda}_n)$. Notice that

$$A - \widetilde{A} = A - D^*AD = A - AD + AD - D^*AD = A(I - D) + (D^{-*} - I)\widetilde{A}.$$

Pre- and postmultiply the equations by U^* and $\widetilde{U}$, respectively, to get

$$\Lambda U^*\widetilde{U} - U^*\widetilde{U}\widetilde{\Lambda} = \Lambda U^*(I - D)\widetilde{U} + U^*(D^{-*} - I)\widetilde{U}\widetilde{\Lambda}. \tag{5.2}$$

Set

$$Q \stackrel{\text{def}}{=} U^*\widetilde{U} = (q_{ij}), \quad E \stackrel{\text{def}}{=} U^*(I - D)\widetilde{U} = (e_{ij}), \quad \widetilde{E} \stackrel{\text{def}}{=} U^*(D^{-*} - I)\widetilde{U} = (\widetilde{e}_{ij}).$$

Then (5.2) reads $\Lambda Q - Q\widetilde{\Lambda} = \Lambda E + \widetilde{E}\widetilde{\Lambda}$, or componentwise $\lambda_i q_{ij} - q_{ij}\widetilde{\lambda}_j = \lambda_i e_{ij} + \widetilde{e}_{ij}\widetilde{\lambda}_j$, so

$$|(\lambda_i - \widetilde{\lambda}_j)q_{ij}|^2 = |\lambda_i e_{ij} + \widetilde{e}_{ij}\widetilde{\lambda}_j|^2 \le (|\lambda_i|^2 + |\widetilde{\lambda}_j|^2)(|e_{ij}|^2 + |\widetilde{e}_{ij}|^2),$$

which yields[6] $[\varrho_2(\lambda_i, \widetilde{\lambda}_j)]^2 |q_{ij}|^2 \le |e_{ij}|^2 + |\widetilde{e}_{ij}|^2$. Hence

$$\begin{aligned}\sum_{i,\,j=1}^{n} \left[\varrho_2(\lambda_i, \widetilde{\lambda}_j)\right]^2 |q_{ij}|^2 &\le \|U^*(I - D)\widetilde{U}\|_{\mathrm{F}}^2 + \|U^*(D^{-*} - I)\widetilde{U}\|_{\mathrm{F}}^2 \\ &= \|I - D\|_{\mathrm{F}}^2 + \|D^{-*} - I\|_{\mathrm{F}}^2.\end{aligned}$$

The matrix $(|q_{ij}|^2)_{n\times n}$ is a doubly stochastic matrix. The above inequality and Lemma 5.1 imply that

$$\sum_{i=1}^{n} \left[\varrho_2(\lambda_i, \widetilde{\lambda}_{\tau(i)})\right]^2 \le \|I - D\|_{\mathrm{F}}^2 + \|D^{-*} - I\|_{\mathrm{F}}^2$$

for some permutation τ of $\{1, 2, \ldots, n\}$. This is (3.3a).

We now prove (3.4) and (3.5). Suppose that A is nonnegative definite. There is a matrix $B \in \mathbb{C}^{n\times n}$ such that $A = B^*B$. With this B, $\widetilde{A} = D^*AD = D^*B^*BD = \widetilde{B}^*\widetilde{B}$, where $\widetilde{B} = BD$. Let SVDs of B and $\widetilde{B}$ be

$$B = U\Lambda^{1/2}V^* \quad \text{and} \quad \widetilde{B} = \widetilde{U}\widetilde{\Lambda}^{1/2}\widetilde{V}^*,$$

where $\Lambda^{1/2} = \operatorname{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$ and $\widetilde{\Lambda}^{1/2} = \operatorname{diag}\left(\sqrt{\widetilde{\lambda}_1}, \sqrt{\widetilde{\lambda}_2}, \ldots, \sqrt{\widetilde{\lambda}_n}\right)$. In what follows, we actually work with BB^* and $\widetilde{B}\widetilde{B}^*$, rather than $A = B^*B$ and $\widetilde{A} = \widetilde{B}^*\widetilde{B}$ themselves. We have

$$\widetilde{B}\widetilde{B}^* - BB^* = \widetilde{B}D^*B^* - \widetilde{B}D^{-1}B^* = \widetilde{B}(D^* - D^{-1})B^*.$$

Pre- and postmultiply the above equations by $\widetilde{U}^*$ and U, respectively, to get

$$\widetilde{\Lambda}\widetilde{U}^*U - \widetilde{U}^*U\Lambda = \widetilde{\Lambda}^{1/2}\widetilde{V}^*(D^* - D^{-1})V\Lambda^{1/2}. \tag{5.3}$$

Write $Q \stackrel{\text{def}}{=} \widetilde{U}^*U = (q_{ij})$. Equation (5.3) implies

$$\|D^* - D^{-1}\|_{\mathrm{F}}^2 = \|\widetilde{V}^*(D^* - D^{-1})V\|_{\mathrm{F}}^2 = \sum_{i,\,j=1}^{n} \frac{|\widetilde{\lambda}_i - \lambda_j|}{\sqrt{\widetilde{\lambda}_i\lambda_j}} |q_{ij}|^2.$$

[6]This inequality still holds even if $\lambda_i = \widetilde{\lambda}_j = 0$ because of our convention $0/0 = 0$; see section 2.

Since $(|q_{ij}|^2)_{n\times n}$ is a doubly stochastic matrix, an application of Lemma 5.1 and Proposition 2.4 concludes the proof of (3.5). To confirm (3.4), let k be the index such that

$$\eta \stackrel{\text{def}}{=} \max_{1\le i\le n} \chi(\lambda_i, \widetilde{\lambda}_i) = \chi(\lambda_k, \widetilde{\lambda}_k).$$

If $\eta = 0$, no proof is necessary. Assume $\eta > 0$. Also assume, without loss of generality, that

$$\lambda_k > \widetilde{\lambda}_k \ge 0.$$

Partition U, V, $\widetilde{U}$, $\widetilde{V}$ as follows:

$$U = \begin{array}{c} \begin{array}{cc} \scriptstyle k & \scriptstyle n-k \end{array} \\ \begin{pmatrix} U_1 & U_2 \end{pmatrix} \end{array}, \; V = \begin{array}{c} \begin{array}{cc} \scriptstyle k & \scriptstyle n-k \end{array} \\ \begin{pmatrix} V_1 & V_2 \end{pmatrix} \end{array}, \; \widetilde{U} = \begin{array}{c} \begin{array}{cc} \scriptstyle k-1 & \scriptstyle n-k+1 \end{array} \\ \begin{pmatrix} \widetilde{U}_1 & \widetilde{U}_2 \end{pmatrix} \end{array}, \; \widetilde{V} = \begin{array}{c} \begin{array}{cc} \scriptstyle k-1 & \scriptstyle n-k+1 \end{array} \\ \begin{pmatrix} \widetilde{V}_1 & \widetilde{V}_2 \end{pmatrix} \end{array},$$

and write $\Lambda = \operatorname{diag}(\Lambda_1, \Lambda_2)$ and $\widetilde{\Lambda} = \operatorname{diag}(\widetilde{\Lambda}_1, \widetilde{\Lambda}_2)$, where $\Lambda_1 \in \mathbb{R}^{k\times k}$ and $\widetilde{\Lambda}_1 \in \mathbb{R}^{(k-1)\times(k-1)}$. It follows from (5.3) that

$$\widetilde{\Lambda}_2 \widetilde{U}_2^* U_1 - \widetilde{U}_2^* U_1 \Lambda_1 = \widetilde{\Lambda}_2^{1/2} \widetilde{V}_2^* (D^* - D^{-1}) V_1 \Lambda_1^{1/2}.$$

Postmultiply this equation by Λ_1^{-1} to get

$$\widetilde{\Lambda}_2 \widetilde{U}_2^* U_1 \Lambda_1^{-1} - \widetilde{U}_2^* U_1 = \widetilde{\Lambda}_2^{1/2} \widetilde{V}_2^* (D^* - D^{-1}) V_1 \Lambda_1^{-1/2}. \tag{5.4}$$

Lemma 5.2 implies that $\|\widetilde{U}_2^* U_1\|_2 = 1$ since $\widetilde{U}_2^* U_1$ is an $(n-k+1)\times k$ submatrix of unitary $\widetilde{U}^* U$ and $k + (n-k+1) = n+1 > n$. Bearing in mind that $\|\widetilde{\Lambda}_2\|_2 = \widetilde{\lambda}_k = \|\widetilde{\Lambda}_2^{1/2}\|_2^2$ and $\|\Lambda_1^{-1}\|_2 = 1/\lambda_k = \|\Lambda_1^{-1/2}\|_2^2$, we have

$$\begin{aligned}
1 - \frac{\widetilde{\lambda}_k}{\lambda_k} &= \left\|\widetilde{U}_2^* U_1\right\|_2 - \|\widetilde{\Lambda}_2\|_2 \left\|\widetilde{U}_2^* U_1\right\|_2 \|\Lambda_1^{-1}\|_2 \\
&\le \left\|\widetilde{U}_2^* U_1\right\|_2 - \left\|\widetilde{\Lambda}_2 \widetilde{U}_2^* U_1 \Lambda_1^{-1}\right\|_2 \\
&\le \left\|\widetilde{U}_2^* U_1 - \widetilde{\Lambda}_2 \widetilde{U}_2^* U_1 \Lambda_1^{-1}\right\|_2 \\
&= \left\|\widetilde{\Lambda}_2^{1/2} \widetilde{V}_2^* (D^* - D^{-1}) V_1 \Lambda_1^{-1/2}\right\|_2 \quad \text{(by (5.4))} \\
&\le \|\widetilde{\Lambda}_2^{1/2}\|_2 \left\|\widetilde{V}_2^* (D^* - D^{-1}) V_1\right\|_2 \|\Lambda_1^{-1/2}\|_2 \\
&= \sqrt{\frac{\widetilde{\lambda}_k}{\lambda_k}} \left\|\widetilde{V}_2^* (D^* - D^{-1}) V_1\right\|_2 \\
&\le \sqrt{\frac{\widetilde{\lambda}_k}{\lambda_k}} \|D^* - D^{-1}\|_2,
\end{aligned}$$

an immediate consequence of which is (3.4). ∎

6. Proofs of Theorems 4.1 and 4.3. We need the following lemma regarding the relative distance χ.

LEMMA 6.1.

1. *If* $0 \le \alpha \le \beta \le \widetilde{\beta} \le \widetilde{\alpha}$, *then* $\chi(\alpha, \widetilde{\alpha}) \ge \chi(\beta, \widetilde{\beta})$.

2. *If* α, $\widetilde{\alpha} \geq 0$, *then* $2\,\chi(\alpha, \widetilde{\alpha}) \leq \chi(\alpha^2, \widetilde{\alpha}^2)$.
3. *For* α, β, $\gamma \geq 0$, *we have*

$$\chi(\alpha, \gamma) \leq \chi(\alpha, \beta) + \chi(\beta, \gamma) + \frac{1}{8}\chi(\alpha, \beta)\chi(\beta, \gamma)\chi(\alpha, \gamma). \tag{6.1}$$

Thus if $\chi(\alpha, \beta)\chi(\beta, \gamma) < 8$ *also, then*

$$\chi(\alpha, \gamma) \leq \frac{\chi(\alpha, \beta) + \chi(\beta, \gamma)}{1 - \frac{1}{8}\chi(\alpha, \beta)\chi(\beta, \gamma)}.$$

Proof. To prove the first inequality, we notice that function $\frac{1}{x} - x$ is monotonically decreasing for $0 \leq x \leq 1$, and that $0 \leq \alpha/\widetilde{\alpha} \leq \beta/\widetilde{\beta} \leq 1$. Thus

$$\chi(\alpha, \widetilde{\alpha}) = \frac{1}{\sqrt{\alpha/\widetilde{\alpha}}} - \sqrt{\alpha/\widetilde{\alpha}} \geq \frac{1}{\sqrt{\beta/\widetilde{\beta}}} - \sqrt{\beta/\widetilde{\beta}} = \chi(\beta, \widetilde{\beta}),$$

as was to be shown. If α, $\widetilde{\alpha} \geq 0$, then

$$\chi(\alpha^2, \widetilde{\alpha}^2) = \chi(\alpha, \widetilde{\alpha})\frac{|\alpha + \widetilde{\alpha}|}{\sqrt{|\alpha\widetilde{\alpha}|}} = \chi(\alpha, \widetilde{\alpha})\frac{\alpha + \widetilde{\alpha}}{\sqrt{\alpha\widetilde{\alpha}}} \geq \chi(\alpha, \widetilde{\alpha})\frac{2\sqrt{\alpha\widetilde{\alpha}}}{\sqrt{\alpha\widetilde{\alpha}}} = 2\chi(\alpha, \widetilde{\alpha}),$$

which confirms the second inequality.

For the third inequality (6.1), without loss of generality, we may assume $0 \leq \alpha \leq \gamma$. Now if $\beta \leq \alpha$ or $\gamma \leq \beta$, we have by the first inequality

$$\chi(\alpha, \gamma) \leq \begin{cases} \chi(\beta, \gamma) \leq \chi(\alpha, \beta) + \chi(\beta, \gamma), & \text{if } \beta \leq \alpha, \\ \chi(\alpha, \beta) \leq \chi(\alpha, \beta) + \chi(\beta, \gamma), & \text{if } \gamma \leq \beta, \end{cases}$$

so (6.1) holds. Consider the case $0 \leq \alpha \leq \beta \leq \gamma$. It can be verified that

$$\chi(\alpha, \gamma) = \chi(\alpha, \beta) + \chi(\beta, \gamma) + \chi(\sqrt{\alpha}, \sqrt{\beta})\chi(\sqrt{\beta}, \sqrt{\gamma})\chi(\sqrt{\alpha}, \sqrt{\gamma}).$$

Inequality (6.1) follows by applying the second inequality. $\square$

Proofs of Theorems 4.1 *and* 4.3. Set $\widehat{B} = BD_2$ and denote its singular values by $\widehat{\sigma}_1 \geq \widehat{\sigma}_2 \geq \cdots \geq \widehat{\sigma}_n$. Apply Theorem 3.1 to B^*B and $\widehat{B}^*\widehat{B} = D_2^*B^*BD_2$ to get

$$\max_{1 \leq i \leq n} \chi(\sigma_i^2, \widehat{\sigma}_i^2) \leq \|D_2^* - D_2^{-1}\|_2 \quad \text{and} \quad \sqrt{\sum_{i=1}^{n} [\chi(\sigma_i^2, \widehat{\sigma}_i^2)]^2} \leq \|D_2^* - D_2^{-1}\|_{\mathrm{F}}.$$

Now apply the second inequality of Lemma 6.1 to obtain

$$\max_{1 \leq i \leq n} \chi(\sigma_i, \widehat{\sigma}_i) \leq \frac{1}{2}\|D_2^* - D_2^{-1}\|_2 \quad \text{and} \quad \sqrt{\sum_{i=1}^{n} [\chi(\sigma_i, \widehat{\sigma}_i)]^2} \leq \frac{1}{2}\|D_2^* - D_2^{-1}\|_{\mathrm{F}}. \tag{6.2}$$

Similarly for $\widehat{B} = BD_2$ and $\widetilde{B} = D_1^*BD_2 = D_1^*\widehat{B}$, we have

$$\max_{1 \leq i \leq n} \chi(\widehat{\sigma}_i, \widetilde{\sigma}_i) \leq \frac{1}{2}\|D_1^* - D_1^{-1}\|_2 \quad \text{and} \quad \sqrt{\sum_{i=1}^{n} [\chi(\widehat{\sigma}_i, \widetilde{\sigma}_i)]^2} \leq \frac{1}{2}\|D_1^* - D_1^{-1}\|_{\mathrm{F}}. \tag{6.3}$$

The first inequalities in (6.2) and (6.3), and the assumptions of Theorem 4.1, imply

$$\chi(\sigma_i,\widehat{\sigma}_i)\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i) \le \frac{1}{4}\|D_1^* - D_1^{-1}\|_2\|D_2^* - D_2^{-1}\|_2 < \frac{1}{4}\times 32 = 8.$$

By Lemma 6.1, we have

$$\begin{aligned}
\chi(\sigma_i,\widetilde{\sigma}_i) &\le \frac{\chi(\sigma_i,\widehat{\sigma}_i)+\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i)}{1-\frac{1}{8}\chi(\sigma_i,\widehat{\sigma}_i)\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i)}\\
&\le \frac{1}{2}\cdot\frac{\|D_1^* - D_1^{-1}\|_2 + \|D_2^* - D_2^{-1}\|_2}{1-\frac{1}{32}\|D_1^* - D_1^{-1}\|_2\|D_2^* - D_2^{-1}\|_2},\\
\sqrt{\sum_{i=1}^n [\chi(\sigma_i,\widetilde{\sigma}_i)]^2} &\le \sqrt{\sum_{i=1}^n \left[\frac{\chi(\sigma_i,\widehat{\sigma}_i)+\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i)}{1-\frac{1}{8}\chi(\sigma_i,\widehat{\sigma}_i)\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i)}\right]^2}\\
&\le \frac{\sqrt{\sum_{i=1}^n [\chi(\sigma_i,\widehat{\sigma}_i)]^2} + \sqrt{\sum_{i=1}^n [\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i)]^2}}{1-\frac{1}{8}\max_{1\le i\le n}\chi(\sigma_i,\widehat{\sigma}_i)\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i)}\\
&\le \frac{1}{2}\cdot\frac{\|D_1^* - D_1^{-1}\|_{\mathrm{F}} + \|D_2^* - D_2^{-1}\|_{\mathrm{F}}}{1-\frac{1}{32}\|D_1^* - D_1^{-1}\|_2\|D_2^* - D_2^{-1}\|_2},
\end{aligned}$$

as expected. This completes the proof of Theorem 4.1. To prove Theorem 4.3, we notice that

$$\begin{aligned}
\varrho_p(\sigma_i,\widetilde{\sigma}_i) &\le \varrho_p(\sigma_i,\widehat{\sigma}_i) + \varrho_p(\widehat{\sigma}_i,\widetilde{\sigma}_i) && (\varrho_p \text{ is a metric on } \mathbb{R})\\
&\le 2^{-1/p}\chi(\sigma_i,\widehat{\sigma}_i) + 2^{-1/p}\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i) && (\text{by Proposition 2.2})\\
&\le 2^{-1-1/p}\left(\|D_2^* - D_2^{-1}\|_2 + \|D_1^* - D_1^{-1}\|_2\right) && (\text{by (6.2) and (6.3)})
\end{aligned}$$

and

$$\begin{aligned}
\sqrt{\sum_{i=1}^n [\varrho_p(\sigma_i,\widetilde{\sigma}_i)]^2} &\le \sqrt{\sum_{i=1}^n [\varrho_p(\sigma_i,\widehat{\sigma}_i) + \varrho_p(\widehat{\sigma}_i,\widetilde{\sigma}_i)]^2} && (\varrho_p \text{ is a metric on } \mathbb{R})\\
&\le \sqrt{\sum_{i=1}^n [\varrho_p(\sigma_i,\widehat{\sigma}_i)]^2} + \sqrt{\sum_{i=1}^n [\varrho_p(\widehat{\sigma}_i,\widetilde{\sigma}_i)]^2}\\
&\le 2^{-1/p}\sqrt{\sum_{i=1}^n [\chi(\sigma_i,\widehat{\sigma}_i)]^2} + 2^{-1/p}\sqrt{\sum_{i=1}^n [\chi(\widehat{\sigma}_i,\widetilde{\sigma}_i)]^2}\\
& && (\text{by Proposition 2.2})\\
&\le 2^{-1-1/p}\left(\|D_2^* - D_2^{-1}\|_{\mathrm{F}} + \|D_1^* - D_1^{-1}\|_{\mathrm{F}}\right) && (\text{by (6.2) and (6.3)}).
\end{aligned}$$

These inequalities complete the proof of Theorem 4.3. ∎

7. Generalized eigenvalue problems and generalized singular value problems. In this section, we discuss perturbations for *scaled generalized eigenvalue problems* and *scaled generalized singular value problems*. As we shall see, the results in previous sections, as well as those in Li [25], can be applied to derive relative perturbation bounds for these problems.

- *The generalized eigenvalue problem:*
  $A_1 - \lambda A_2 \equiv S_1^* H_1 S_1 - \lambda S_2^* H_2 S_2$ and $\widetilde{A}_1 - \lambda \widetilde{A}_2 \equiv S_1^* \widetilde{H}_1 S_1 - \lambda S_2^* \widetilde{H}_2 S_2$, where H_1 and H_2 are positive definite; $\|H_j^{-1}\|_2 \|\widetilde{H}_j - H_j\|_2 < 1$ for $j = 1, 2$; S_1 and S_2 are some square matrices and one of them is nonsingular.[7]
- *The generalized singular value problem:*
  $\{B_1, B_2\} \equiv \{G_1 S_1, G_2 S_2\}$ and $\{\widetilde{B}_1, \widetilde{B}_2\} \equiv \{\widetilde{G}_1 S_1, \widetilde{G}_2 S_2\}$, where G_1 and G_2 are nonsingular; $\|G_j^{-1}\|_2 \|\widetilde{G}_j - G_j\|_2 < 1$ for $j = 1, 2$; S_1 and S_2 are some square matrices and one of them is nonsingular.

For the scaled generalized eigenvalue problem just mentioned, without loss of generality, we consider the case when S_2 is nonsingular. Then the generalized eigenvalue problem for $A_1 - \lambda A_2 \equiv S_1^* H_1 S_1 - \lambda S_2^* H_2 S_2$ is equivalent to the standard eigenvalue problem for

$$A \stackrel{\text{def}}{=} H_2^{-1/2} S_2^{-*} S_1^* H_1 S_1 S_2^{-1} H_2^{-1/2},$$

and the generalized eigenvalue problem for $\widetilde{A}_1 - \lambda \widetilde{A}_2 \equiv S_1^* \widetilde{H}_1 S_1 - \lambda S_2^* \widetilde{H}_2 S_2$ is equivalent to the standard eigenvalue problem for

$$\widetilde{A} \stackrel{\text{def}}{=} D_2^* H_2^{-1/2} S_2^{-*} S_1^* \widetilde{H}_1 S_1 S_2^{-1} H_2^{-1/2} D_2,$$

where

$$D_2 = D_2^* \stackrel{\text{def}}{=} \left(I + H_2^{-1/2} (\Delta H_2) H_2^{-1/2} \right)^{-1/2} \quad \text{and} \quad \Delta H_2 \stackrel{\text{def}}{=} \widetilde{H}_2 - H_2.$$

So, bounding relative distances between the eigenvalues of $A_1 - \lambda A_2$ and those of $\widetilde{A}_1 - \lambda \widetilde{A}_2$ is transformed to bounding relative distances between the eigenvalues of A and those of $\widetilde{A}$. The latter can be accomplished in two steps:

1. Bounding relative distances between the eigenvalues of A and those of

$$\widehat{A} \stackrel{\text{def}}{=} D_2^* H_2^{-1/2} S_2^{-*} S_1^* H_1 S_1 S_2^{-1} H_2^{-1/2} D_2 = D_2^* A D_2.$$

2. Bounding relative distances between the eigenvalues of $\widehat{A}$ and those of $\widetilde{A}$.

Denote and order the eigenvalues of A, $\widehat{A}$, and $\widetilde{A}$ as

$$\lambda_1 \geq \cdots \geq \lambda_n, \quad \widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_n, \quad \text{and} \quad \widetilde{\lambda}_1 \geq \cdots \geq \widetilde{\lambda}_n.$$

Set

$$D_1 = D_1^* \stackrel{\text{def}}{=} \left(I + H_1^{-1/2} (\Delta H_1) H_1^{-1/2} \right)^{-1/2} \quad \text{and} \quad \Delta H_1 \stackrel{\text{def}}{=} \widetilde{H}_1 - H_1.$$

By Theorem 3.1 on A and $\widehat{A} = D_2^* A D_2$, Theorem 3.2 on $\widehat{A} = X^* H_1 X$, and $\widetilde{A} = X^* \widetilde{H}_1 X$, where $X = S_1 S_2^{-1} H_2^{-1/2} D_2$, we have

$$\chi(\lambda_i, \widehat{\lambda}_i) \leq \|D_2 - D_2^{-1}\|_2 \quad \text{and} \quad \chi(\widehat{\lambda}_i, \widetilde{\lambda}_i) \leq \|D_1 - D_1^{-1}\|_2 \tag{7.1}$$

and

$$\sqrt{\sum_{i=1}^n \left[\chi(\lambda_i, \widehat{\lambda}_i) \right]^2} \leq \|D_2 - D_2^{-1}\|_{\text{F}} \quad \text{and} \quad \sqrt{\sum_{i=1}^n \left[\chi(\widehat{\lambda}_i, \widetilde{\lambda}_i) \right]^2} \leq \|D_1 - D_1^{-1}\|_{\text{F}}. \tag{7.2}$$

[7]When S_2 is singular, both pencils will have the same number of the eigenvalue $+\infty$. For convenience, we define the relative differences by any measure introduced in section 2 to be 0.

By Lemma 6.1, we have that if $\|D_1 - D_1^{-1}\|_2 \|D_2 - D_2^{-1}\|_2 < 8$, then

$$\chi(\lambda_i, \widetilde{\lambda}_i) \leq \frac{\chi(\lambda_i, \widehat{\lambda}_i) + \chi(\widehat{\lambda}_i, \widetilde{\lambda}_i)}{1 - \frac{1}{8}\chi(\lambda_i, \widehat{\lambda}_i)\chi(\widehat{\lambda}_i, \widetilde{\lambda}_i)} \leq \frac{\|D_2 - D_2^{-1}\|_2 + \|D_1 - D_1^{-1}\|_2}{1 - \frac{1}{8}\|D_1 - D_1^{-1}\|_2 \|D_2 - D_2^{-1}\|_2}$$

and

$$\begin{aligned}
\sqrt{\sum_{i=1}^n \left[\chi(\lambda_i, \widetilde{\lambda}_i)\right]^2} &\leq \sqrt{\sum_{i=1}^n \left[\frac{\chi(\lambda_i, \widehat{\lambda}_i) + \chi(\widehat{\lambda}_i, \widetilde{\lambda}_i)}{1 - \frac{1}{8}\chi(\lambda_i, \widehat{\lambda}_i)\chi(\widehat{\lambda}_i, \widetilde{\lambda}_i)}\right]^2} \\
&\leq \frac{\sqrt{\sum_{i=1}^n \left[\chi(\lambda_i, \widehat{\lambda}_i)\right]^2} + \sqrt{\sum_{i=1}^n \left[\chi(\widehat{\lambda}_i, \widetilde{\lambda}_i)\right]^2}}{1 - \frac{1}{8}\max_{1\leq i\leq n} \chi(\lambda_i, \widehat{\lambda}_i)\chi(\widehat{\lambda}_i, \widetilde{\lambda}_i)} \\
&\leq \frac{\|D_2 - D_2^{-1}\|_{\mathrm{F}} + \|D_1 - D_1^{-1}\|_{\mathrm{F}}}{1 - \frac{1}{8}\|D_1 - D_1^{-1}\|_2 \|D_2 - D_2^{-1}\|_2}.
\end{aligned}$$

Notice also that for $j = 1, 2$ and for any unitarily invariant norm $\|\!|\cdot|\!\|$,

$$\|\!|D_j - D_j^{-1}|\!\| \leq \frac{\|H_j^{-1}\|_2 \|\!|\Delta H_j|\!\|}{\sqrt{1 - \|H_j^{-1}\|_2 \|\Delta H_j\|_2}}.$$

So we have proved the following.

THEOREM 7.1. *Let* $A_1 - \lambda A_2 \equiv S_1^* H_1 S_1 - \lambda S_2^* H_2 S_2$ *and* $\widetilde{A}_1 - \lambda \widetilde{A}_2 \equiv S_1^* \widetilde{H}_1 S_1 - \lambda S_2^* \widetilde{H}_2 S_2$, *where* H_1 *and* H_2 *are* $n \times n$, *positive definite, and* $\|H_j^{-1}\|_2 \|\widetilde{H}_j - H_j\|_2 < 1$ *for* $j = 1, 2$. S_1 *and* S_2 *are some square matrices and one of them is nonsingular. Let the generalized eigenvalues of* $A_1 - \lambda A_2$ *and* $\widetilde{A}_1 - \lambda \widetilde{A}_2$ *be*

$$\lambda_1 \geq \cdots \geq \lambda_n \quad \textit{and} \quad \widetilde{\lambda}_1 \geq \cdots \geq \widetilde{\lambda}_n.$$

If $\theta_1 \theta_2 \|\Delta H_1\|_2 \|\Delta H_2\|_2 < 8$, *then*

$$\begin{aligned}
\max_{1\leq i\leq n} \chi(\lambda_i, \widetilde{\lambda}_i) &\leq \frac{\theta_1 \|\Delta H_1\|_2 + \theta_2 \|\Delta H_2\|_2}{1 - \frac{1}{8}\theta_1\theta_2 \|\Delta H_1\|_2 \|\Delta H_2\|_2}, \\
\sqrt{\sum_{i=1}^n \left[\chi(\lambda_i, \widetilde{\lambda}_i)\right]^2} &\leq \frac{\theta_1 \|\Delta H_1\|_{\mathrm{F}} + \theta_2 \|\Delta H_2\|_{\mathrm{F}}}{1 - \frac{1}{8}\theta_1\theta_2 \|\Delta H_1\|_2 \|\Delta H_2\|_2},
\end{aligned}$$

where $\theta_j \stackrel{\text{def}}{=} \|H_j^{-1}\|_2 \Big/ \sqrt{1 - \|H_j^{-1}\|_2 \|\Delta H_j\|_2}$ *for* $j = 1, 2$.

On the other hand, from (7.1), (7.2), and Proposition 2.2, we get

$$\varrho_p(\lambda_i, \widehat{\lambda}_i) \leq 2^{-1/p} \|D_2 - D_2^{-1}\|_2 \quad \text{and} \quad \varrho_p(\widehat{\lambda}_i, \widetilde{\lambda}_i) \leq 2^{-1/p} \|D_1 - D_1^{-1}\|_2$$

and

$$\sqrt{\sum_{i=1}^n \left[\varrho_p(\lambda_i, \widehat{\lambda}_i)\right]^2} \leq 2^{-1/p} \|D_2 - D_2^{-1}\|_{\mathrm{F}} \quad \text{and} \quad \sqrt{\sum_{i=1}^n \left[\varrho_p(\widehat{\lambda}_i, \widetilde{\lambda}_i)\right]^2} \leq 2^{-1/p} \|D_1 - D_1^{-1}\|_{\mathrm{F}}.$$

Since ϱ_p is a metric on $\mathbb{R}$, we have

$$\varrho_p(\lambda_i, \widetilde{\lambda}_i) \le \varrho_p(\lambda_i, \widehat{\lambda}_i) + \varrho_p(\widehat{\lambda}_i, \widetilde{\lambda}_i) \le 2^{-1/p}\left(\|D_2 - D_2^{-1}\|_2 + \|D_1 - D_1^{-1}\|_2\right)$$

and

$$\begin{aligned}\sqrt{\sum_{i=1}^n \left[\varrho_p(\lambda_i, \widetilde{\lambda}_i)\right]^2} &\le \sqrt{\sum_{i=1}^n \left[\varrho_p(\lambda_i, \widehat{\lambda}_i) + \varrho_p(\widehat{\lambda}_i, \widetilde{\lambda}_i)\right]^2} \\ &\le \sqrt{\sum_{i=1}^n \left[\varrho_p(\lambda_i, \widehat{\lambda}_i)\right]^2} + \sqrt{\sum_{i=1}^n \left[\varrho_p(\widehat{\lambda}_i, \widetilde{\lambda}_i)\right]^2} \\ &\le 2^{-1/p}\left(\|D_2 - D_2^{-1}\|_{\mathrm{F}} + \|D_1 - D_1^{-1}\|_{\mathrm{F}}\right).\end{aligned}$$

THEOREM 7.2. *Let all conditions of Theorem* 7.1, *except* $\|D_1 - D_1^{-1}\|_2\|D_2 - D_2^{-1}\|_2 < 8$, *which is no longer necessary, hold. Then*

$$\max_{1\le i\le n} \varrho_p(\lambda_i, \widetilde{\lambda}_i) \le 2^{-1/p}(\theta_1\|\Delta H_1\|_2 + \theta_2\|\Delta H_2\|_2),$$

$$\sqrt{\sum_{i=1}^n \left[\varrho_p(\lambda_i, \widetilde{\lambda}_i)\right]^2} \le 2^{-1/p}(\theta_1\|\Delta H_1\|_{\mathrm{F}} + \theta_2\|\Delta H_2\|_{\mathrm{F}}).$$

As to the scaled generalized singular value problem mentioned above, we shall consider instead its corresponding generalized eigenvalue problem [21, 36, 37] for

$$S_1^* G_1^* G_1 S_1 - \lambda S_2^* G_2^* G_2 S_2 \quad \text{and} \quad S_1^* \widetilde{G}_1^* \widetilde{G}_1 S_1 - \lambda S_2^* \widetilde{G}_2^* \widetilde{G}_2 S_2. \tag{7.3}$$

THEOREM 7.3. *Let* $\{B_1, B_2\} \equiv \{G_1S_1, G_2S_2\}$ *and* $\{\widetilde{B}_1, \widetilde{B}_2\} \equiv \{\widetilde{G}_1S_1, \widetilde{G}_2S_2\}$, *where* G_1 *and* G_2 *are* $n \times n$ *and nonsingular;* $\|G_j^{-1}\|_2\|\widetilde{G}_j - G_j\|_2 < 1$ *for* $j = 1, 2$*;* S_1 *and* S_2 *are some square matrices and one of them is nonsingular. Let the generalized singular values of* $\{B_1, B_2\}$ *and* $\{\widetilde{B}_1, \widetilde{B}_2\}$ *be*

$$\sigma_1 \ge \cdots \ge \sigma_n \quad \textit{and} \quad \widetilde{\sigma}_1 \ge \cdots \ge \widetilde{\sigma}_n.$$

If $\delta_{12}\delta_{22} < 32$, *where*

$$\delta_{jt} = \left\|\left(I + (\Delta G_j)G_j^{-1}\right)^* - \left(I + (\Delta G_j)G_j^{-1}\right)^{-1}\right\|_t \quad \textit{for } j = 1, 2 \textit{ and } t = 2, \mathrm{F},$$

then

$$\max_{1\le i\le n} \chi(\sigma_i, \widetilde{\sigma}_i) \le \frac{1}{2}\cdot\frac{\delta_{12} + \delta_{22}}{1 - \frac{1}{32}\delta_{12}\delta_{22}},$$

$$\sqrt{\sum_{i=1}^n [\chi(\sigma_i, \widetilde{\sigma}_i)]^2} \le \frac{1}{2}\cdot\frac{\delta_{1\mathrm{F}} + \delta_{2\mathrm{F}}}{1 - \frac{1}{32}\delta_{12}\delta_{22}}.$$

It can be proved that for $j = 1, 2$ *and* $t = 2, \mathrm{F}$,

$$\begin{aligned}\delta_{jt} &\le \left(\frac{\|(\Delta G_j)G_j^{-1} + G_j^{-*}(\Delta G_j)^*\|_t}{\|(\Delta G_j)G_j^{-1}\|_t} + \frac{\|(\Delta G_j)G_j^{-1}\|_2}{1 - \|(\Delta G_j)G_j^{-1}\|_2}\right)\|(\Delta G_j)G_j^{-1}\|_t \\ &\le \left(1 + \frac{1}{1 - \|G_j^{-1}\|_2\|\Delta G_j\|_2}\right)\|G_j^{-1}\|_2\|\Delta G_j\|_t.\end{aligned}$$

Proof. Consider the case when S_2 is nonsingular. (The case when S_1 is nonsingular can be handled analogously.) By (7.3), we know that the singular values of $B \stackrel{\text{def}}{=} G_1 S_1 S_2^{-1} G_2^{-1}$ and $\widetilde{B} \stackrel{\text{def}}{=} \widetilde{G}_1 S_1 S_2^{-1} \widetilde{G}_2^{-1}$ are $\sigma_1 \geq \cdots \geq \sigma_n$ and $\widetilde{\sigma}_1 \geq \cdots \geq \widetilde{\sigma}_n$, respectively. Set

$$D_1 = I + (\Delta G_1) G_1^{-1},\ \Delta G_1 = \widetilde{G}_1 - G_1, \text{ and } D_2 = I + (\Delta G_2) G_2^{-1},\ \Delta G_2 = \widetilde{G}_2 - G_2;$$

then $\widetilde{B} = D_1 B D_2^{-1}$. By Theorem 4.1, we have

$$\max_{1 \leq i \leq n} \chi(\sigma_i, \widetilde{\sigma}_i) \leq \frac{1}{2} \frac{\|D_1^* - D_1^{-1}\|_2 + \|D_2^{-*} - D_2\|_2}{1 - \frac{1}{32}\|D_1^* - D_1^{-1}\|_2 \|D_2^{-*} - D_2\|_2},$$

$$\sqrt{\sum_{i=1}^{n} [\chi(\sigma_i, \widetilde{\sigma}_i)]^2} \leq \frac{1}{2} \frac{\|D_1^* - D_1^{-1}\|_{\mathrm{F}} + \|D_2^{-*} - D_2\|_{\mathrm{F}}}{1 - \frac{1}{32}\|D_1^* - D_1^{-1}\|_2 \|D_2^{-*} - D_2\|_2},$$

as were to be shown. □

By the first half of the proof of Theorem 7.3 and by Theorem 4.3, we can prove the following.

THEOREM 7.4. *Let all conditions of Theorem* 7.3, *except* $\delta_{12}\delta_{22} < 32$, *which is no longer necessary, hold. Then*

$$\max_{1 \leq i \leq n} \varrho_p(\sigma_i, \widetilde{\sigma}_i) \leq \frac{1}{2^{1+1/p}} (\delta_{12} + \delta_{22}),$$

$$\sqrt{\sum_{i=1}^{n} [\varrho_p(\sigma_i, \widetilde{\sigma}_i)]^2} \leq \frac{1}{2^{1+1/p}} (\delta_{1\mathrm{F}} + \delta_{2\mathrm{F}}).$$

8. Conclusions. We have developed a relative perturbation theory for eigenvalue and singular value variations under multiplicative perturbations. In the theory, extensions of the celebrated Hoffman–Wielandt and Weyl–Lidskii theorems from the classical perturbation theory are made. Our extensions use two kinds of relative distance: ϱ_p and χ. Topologically, these new relative distances are equivalent to the classical measurement (2.1) for relative accuracy, but the new distances have better mathematical properties. It is proved that ϱ_p is indeed a metric on $\mathbb{R}$ while χ is not. Often it is the case that perturbation bounds using χ are sharper than bounds using ϱ_p.

Our unifying treatment in this paper covers many previously studied cases and yields bounds that are at least as sharp as existing ones. Our results are applicable to the computations of sharp error bounds in the Demmel–Kahan QR [8] algorithm and the Fernando–Parlett implementation of the Rutishauser QD algorithm [14]; see Li [23].

Previous approaches to building a relative perturbation theory are more or less along the lines of using the min-max principle for Hermitian matrix eigenvalue problems. Our approach in this paper, however, is through deriving the perturbation equations (5.2) and (5.3). A major advantage of this new approach is that these perturbation equations will lead to the successful extensions in [25] of Davis–Kahan $\sin\theta$ theorems [5] and Wedin $\sin\theta$ theorems [38].

Appendix A. Proofs of Propositions 2.3 and 2.4.

LEMMA A.1. *Let α, β, $\widetilde{\alpha}$, $\widetilde{\beta} \in \mathbb{R}$. If $\alpha \le \beta \le \widetilde{\beta} \le \widetilde{\alpha}$, then $\varrho_1(\alpha, \widetilde{\alpha}) \ge \varrho_1(\beta, \widetilde{\beta})$. If $\alpha \le \beta \le \widetilde{\beta} \le \widetilde{\alpha}$ and $\beta\widetilde{\beta} \ge 0$, then $\varrho_p(\alpha, \widetilde{\alpha}) \ge \varrho_p(\beta, \widetilde{\beta})$ for $p > 1$, and it is strict if either $\alpha < \beta$ or $\widetilde{\beta} < \widetilde{\alpha}$ holds.*

Proof. We consider function $f(\xi)$ defined by

$$f(\xi) \stackrel{\text{def}}{=} \frac{1-\xi}{\sqrt[p]{1+|\xi|^p}}, \quad \text{where } -1 \le \xi \le 1.$$

When $p = 1$,

$$f(\xi) = \begin{cases} 1, & \text{for } -1 \le \xi \le 0, \\ \frac{2}{1+\xi} - 1, & \text{for } 0 \le \xi \le 1, \end{cases}$$

so $f(\xi)$ decreases monotonically and decreases strictly monotonically for $0 \le \xi \le 1$. We are about to prove that when $p > 1$, function $f(\xi)$ so defined is strictly monotonically decreasing. This is true if $p = \infty$. When $1 < p < \infty$, set $h(\xi) \stackrel{\text{def}}{=} [f(\xi)]^p$ and $g(\xi) \stackrel{\text{def}}{=} [f(-\xi)]^p$. Since, for $0 < \xi < 1$,

$$h'(\xi) = -\frac{p(1-\xi)^{p-1}(1+\xi^{p-1})}{(1+\xi^p)^2} < 0 \quad \text{and} \quad g'(\xi) = \frac{p(1+\xi)^{p-1}(1-\xi^{p-1})}{(1+\xi^p)^2} > 0,$$

for $0 < \xi < 1$, $h(\xi)$ is strictly monotonically decreasing and $g(\xi)$ is strictly monotonically increasing. Thus function $f(\xi)$ is strictly monotonically decreasing for $p > 1$.

There are four cases to deal with. Assume that at least one of $\alpha \le \beta$ and $\widetilde{\beta} \le \widetilde{\alpha}$ is strict.

1. $0 \le \alpha \le \beta \le \widetilde{\beta} \le \widetilde{\alpha}$, then $0 \le \alpha/\widetilde{\alpha} < \beta/\widetilde{\beta} \le 1$; thus
$$\varrho_p(\alpha, \widetilde{\alpha}) = f(\alpha/\widetilde{\alpha}) > f(\beta/\widetilde{\beta}) = \varrho_p(\beta, \widetilde{\beta}).$$
2. $\alpha \le 0 \le \beta \le \widetilde{\beta} \le \widetilde{\alpha}$ or $\alpha \le \beta \le \widetilde{\beta} \le 0 \le \widetilde{\alpha}$; then
$$\varrho_p(\alpha, \widetilde{\alpha}) \ge 1 \ge \varrho_p(\beta, \widetilde{\beta}).$$
It is easy to verify that the equalities in the two inequality signs cannot be satisfied simultaneously.
3. $\alpha \le \beta \le 0 \le \widetilde{\beta} \le \widetilde{\alpha}$. Only $p = 1$ shall be considered:
$$\varrho_1(\alpha, \widetilde{\alpha}) = 1 = \varrho_1(\beta, \widetilde{\beta}).$$
4. $\alpha \le \beta \le \widetilde{\beta} \le \widetilde{\alpha} \le 0$, then $0 \le \widetilde{\alpha}/\alpha < \widetilde{\beta}/\beta \le 1$; thus
$$\varrho_p(\alpha, \widetilde{\alpha}) = f(\widetilde{\alpha}/\alpha) > f(\widetilde{\beta}/\beta) = \varrho_p(\beta, \widetilde{\beta}).$$

The proof is completed. ∎

Remark A.1. In Lemma A.1, assumption $\beta\widetilde{\beta} \ge 0$ for the case $p > 1$ is essential. A *counterexample* is the following: let $\xi > \zeta > 0$, and let $\alpha = -\zeta \le \beta = -\zeta < \widetilde{\beta} = \zeta < \widetilde{\alpha} < \xi$. Then

$$\varrho_p(\alpha, \widetilde{\alpha}) = \frac{\xi + \zeta}{\sqrt[p]{\xi^p + \zeta^p}} < 2^{1-1/p} = \varrho_p(\beta, \widetilde{\beta}).$$

Proof of Proposition 2.3. For any permutation τ of $\{1, 2, \ldots, n\}$, the idea of our proof is to construct $n+1$ permutations τ_j such that

$$\tau_0 = \tau, \qquad \tau_n = \text{identity permutation},$$

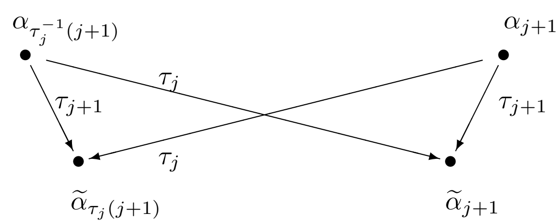
and for $j = 0, 1, 2, \ldots, n-1$,

$$\max_{1 \le i \le n} \varrho_p(\alpha_i, \widetilde{\alpha}_{\tau_j(i)}) \ge \max_{1 \le i \le n} \varrho_p(\alpha_i, \widetilde{\alpha}_{\tau_{j+1}(i)}).$$

The construction of these τ_j's goes as follows. Set $\tau_0 = \tau$. Given τ_j, if $\tau_j(j+1) = j+1$, set $\tau_{j+1} = \tau_j$; otherwise, define

$$\tau_{j+1}(i) = \begin{cases} \tau_j(i), & \text{for } \tau_j^{-1}(j+1) \ne i \ne j+1, \\ j+1, & \text{for } i = j+1, \\ \tau_j(j+1), & \text{for } i = \tau_j^{-1}(j+1). \end{cases}$$

In this latter case, τ_j and τ_{j+1} differ only at two indices as shown in the following picture (notice that $\tau_j^{-1}(j+1) > j+1$ and $\tau_j(j+1) > j+1$):



With Lemma A.1, it is easy to prove that

$$\begin{aligned} &\max\left\{\varrho_p(\alpha_{j+1}, \widetilde{\alpha}_{\tau_j(j+1)}),\ \varrho_p(\alpha_{\tau_j^{-1}(j+1)}, \widetilde{\alpha}_{j+1})\right\} \\ &\ge \max\left\{\varrho_p(\alpha_{j+1}, \widetilde{\alpha}_{j+1}),\ \varrho_p(\alpha_{\tau_j^{-1}(j+1)}, \widetilde{\alpha}_{\tau_j(j+1)})\right\}. \end{aligned}$$

Thus τ_j's so constructed have the desired properties. ∎

A proof of Proposition 2.4 can be given analogously with the help of the first inequality of Lemma 6.1 and the following lemma.

Lemma A.2. *Let* $\alpha_1 \ge \alpha_2 > 0$ *and* $\widetilde{\alpha}_1 \ge \widetilde{\alpha}_2 > 0$. *Then*

$$[\chi(\alpha_1, \widetilde{\alpha}_1)]^2 + [\chi(\alpha_2, \widetilde{\alpha}_2)]^2 \le [\chi(\alpha_1, \widetilde{\alpha}_2)]^2 + [\chi(\alpha_2, \widetilde{\alpha}_1)]^2,$$

and the equality holds if and only if either $\alpha_1 = \alpha_2$ *or* $\widetilde{\alpha}_1 = \widetilde{\alpha}_2$.

Proof. It can be verified that

$$\begin{aligned} &\frac{(\widetilde{\alpha}_1 - \alpha_1)^2}{\widetilde{\alpha}_1 \alpha_1} + \frac{(\widetilde{\alpha}_2 - \alpha_2)^2}{\widetilde{\alpha}_2 \alpha_2} - \frac{(\widetilde{\alpha}_2 - \alpha_1)^2}{\widetilde{\alpha}_2 \alpha_1} - \frac{(\widetilde{\alpha}_1 - \alpha_2)^2}{\widetilde{\alpha}_1 \alpha_2} \\ &\quad = -\frac{(\alpha_1 - \alpha_2)(\widetilde{\alpha}_1 - \widetilde{\alpha}_2)(\widetilde{\alpha}_1 \widetilde{\alpha}_2 + \alpha_1 \alpha_2)}{\widetilde{\alpha}_1 \alpha_1 \widetilde{\alpha}_2 \alpha_2} \le 0, \end{aligned}$$

and the equality holds if and only if either $\alpha_1 = \alpha_2$ or $\widetilde{\alpha}_1 = \widetilde{\alpha}_2$. ∎

Appendix B. ϱ_p is a metric on $\mathbb{R}$. Throughout this appendix, we will be working with real numbers. The definition (2.2) of ϱ_p immediately implies that

1. $\varrho_p(\alpha, \widetilde{\alpha}) \ge 0$; and $\varrho_p(\alpha, \widetilde{\alpha}) = 0$ if and only if $\alpha = \widetilde{\alpha}$.

2. $\varrho_p(\alpha,\widetilde{\alpha}) = \varrho_p(\widetilde{\alpha},\alpha)$.

So it remains to show that ϱ_p satisfies the triangle inequality

$$\varrho_p(\alpha,\gamma) \le \varrho_p(\alpha,\beta) + \varrho_p(\beta,\gamma) \quad \text{for } \alpha,\ \beta,\ \gamma \in \mathbb{R} \tag{B.1}$$

to conclude that the following holds.

THEOREM B.1. *ϱ_p is a metric on $\mathbb{R}$.*

We strongly conjecture that ϱ_p is a metric on $\mathbb{C}$. Unfortunately, we are unable to prove it at this point.

Since ϱ_p is symmetric with respect to its two arguments, we may assume, without loss of generality, that from now on

$$\alpha \le \gamma. \tag{B.2}$$

There are three possible positions for β:

$$\beta \le \alpha \quad \text{or} \quad \alpha < \beta \le \gamma \quad \text{or} \quad \gamma < \beta. \tag{B.3}$$

The hardest part of our proof is to show that (B.1) holds for the second position of β in (B.3). We state it in the following lemma whose proof is postponed to the end of this section.

LEMMA B.2. *Inequality* (B.1) *holds for $\alpha \le \beta \le \gamma$, and the equality holds if and only if $\beta = \alpha$ or $\beta = \gamma$.*

With this lemma, we are now ready to prove (B.1).

Proof of (B.1). The proof is divided into two different cases.

- *The case $\alpha\gamma \ge 0$.* Lemma B.2 says that (B.1) is true if $\alpha \le \beta \le \gamma$. If either $\beta < \alpha$ or $\gamma < \beta$, by Lemma A.1, we have

$$\varrho_p(\alpha,\gamma) \le \begin{cases} \varrho_p(\alpha,\beta) \le \varrho_p(\alpha,\beta) + \varrho_p(\beta,\gamma), & \text{if } \gamma \le \beta, \\ \varrho_p(\beta,\gamma) \le \varrho_p(\alpha,\beta) + \varrho_p(\beta,\gamma), & \text{if } \beta \le \alpha. \end{cases}$$

- *The case $\alpha\gamma < 0$.* We may assume $\alpha < 0$ and $\gamma > 0$ (see (B.2)). Consider the three possible positions (B.3) for β.
  1. $\beta \le \alpha < 0$. In this subcase, $1/\alpha \le 1/\beta < 0 < 1/\gamma$. By Lemma B.2, we have
  $$\varrho_p(\alpha,\gamma) = \varrho_p(1/\alpha,1/\gamma) \le \varrho_p(1/\alpha,1/\beta)+\varrho_p(1/\beta,1/\gamma) = \varrho_p(\alpha,\beta)+\varrho_p(\beta,\gamma).$$
  2. $\alpha \le \beta \le \gamma$. This subcase has been taken care of by Lemma B.2.
  3. $0 < \gamma \le \beta$. In this subcase, $1/\alpha < 0 < 1/\beta \le 1/\gamma$. The rest is the same as in subcase 1 above.

The proof is completed. $\square$

Proof of Lemma B.2. Since both swapping α and γ and multiplying α, β, γ all by -1 lose no generality, we may further assume that

$$\alpha \le |\alpha| \le \gamma. \tag{B.4}$$

Inequality (B.1) clearly holds if one of α, β, γ is zero or if $\beta = \alpha$, $\beta = \gamma$, or $\alpha = \gamma$. So from now on we assume

$$\alpha < \beta < \gamma \quad \text{and} \quad \alpha \ne 0,\ \beta \ne 0,\ \gamma \ne 0.$$

For $1 \le p < \infty$,

$$\begin{aligned}
\varrho_p(\alpha,\gamma) &= \frac{\gamma-\alpha}{\sqrt[p]{\gamma^p+|\alpha|^p}} = \frac{\gamma-\beta+\beta-\alpha}{\sqrt[p]{\gamma^p+|\alpha|^p}} = \frac{\gamma-\beta}{\sqrt[p]{\gamma^p+|\alpha|^p}} + \frac{\beta-\alpha}{\sqrt[p]{\gamma^p+|\alpha|^p}} \\
&= \frac{\gamma-\beta}{\sqrt[p]{\gamma^p+|\beta|^p}} + \frac{\beta-\alpha}{\sqrt[p]{|\beta|^p+|\alpha|^p}} \\
&\quad +(\gamma-\beta)\left(\frac{1}{\sqrt[p]{\gamma^p+|\alpha|^p}} - \frac{1}{\sqrt[p]{\gamma^p+|\beta|^p}}\right) \\
&\quad +(\beta-\alpha)\left(\frac{1}{\sqrt[p]{\gamma^p+|\alpha|^p}} - \frac{1}{\sqrt[p]{|\alpha|^p+|\beta|^p}}\right) \\
&= \varrho_p(\alpha,\beta)+\varrho_p(\beta,\gamma)+h,
\end{aligned}$$

where

$$\begin{aligned}
h = &\frac{(\gamma-\beta)(|\beta|^p-|\alpha|^p)}{\sqrt[p]{\gamma^p+|\alpha|^p}\,\sqrt[p]{\gamma^p+|\beta|^p}} \cdot \frac{\sqrt[p]{\gamma^p+|\beta|^p}-\sqrt[p]{\gamma^p+|\alpha|^p}}{|\beta|^p-|\alpha|^p} \\
&+\frac{(\beta-\alpha)(|\beta|^p-\gamma^p)}{\sqrt[p]{\gamma^p+|\alpha|^p}\,\sqrt[p]{|\alpha|^p+|\beta|^p}} \cdot \frac{\sqrt[p]{|\alpha|^p+|\beta|^p}-\sqrt[p]{\gamma^p+|\alpha|^p}}{|\beta|^p-\gamma^p}.
\end{aligned}$$

The second factors of the two summands in h are always nonnegative. Now if $\alpha < \beta \le |\alpha| \le \gamma$, then $|\beta|^p - |\alpha|^p \le 0$ and $|\beta|^p - \gamma^p < 0$, and thus $h < 0$. Hence $\varrho_p(\alpha,\gamma) < \varrho_p(\alpha,\beta)+\varrho_p(\beta,\gamma)$. Consider now $|\alpha| < \beta < \gamma$. Then

$$\begin{aligned}
h = \frac{(\gamma-\beta)(\beta-|\alpha|)}{\sqrt[p]{\gamma^p+|\alpha|^p}} &\left(\frac{1}{\sqrt[p]{\gamma^p+\beta^p}} \cdot \frac{\beta^p-|\alpha|^p}{\beta-|\alpha|} \cdot \frac{\sqrt[p]{\gamma^p+\beta^p}-\sqrt[p]{\gamma^p+|\alpha|^p}}{\beta^p-|\alpha|^p}\right. \\
&\left. - \frac{1}{\sqrt[p]{|\alpha|^p+\beta^p}} \cdot \frac{\gamma^p-\beta^p}{\gamma-\beta} \cdot \frac{\sqrt[p]{|\alpha|^p+\beta^p}-\sqrt[p]{\gamma^p+|\alpha|^p}}{\beta^p-\gamma^p}\right) \\
< 0.&
\end{aligned}$$

The last inequality is true because $\sqrt[p]{\gamma^p+\beta^p} > \sqrt[p]{|\alpha|^p+\beta^p} \Rightarrow \frac{1}{\sqrt[p]{\gamma^p+\beta^p}} < \frac{1}{\sqrt[p]{|\alpha|^p+\beta^p}}$ and

$$\begin{aligned}
&0 < \frac{\beta^p-|\alpha|^p}{\beta-|\alpha|} \le \frac{\gamma^p-\beta^p}{\gamma-\beta}, \\
&0 < \frac{\sqrt[p]{\gamma^p+\beta^p}-\sqrt[p]{\gamma^p+|\alpha|^p}}{(\gamma^p+\beta^p)-(\gamma^p+|\alpha|^p)} \le \frac{\sqrt[p]{|\alpha|^p+\beta^p}-\sqrt[p]{\gamma^p+|\alpha|^p}}{(|\alpha|^p+\beta^p)-(\gamma^p+|\alpha|^p)}
\end{aligned}$$

by Lemma B.3, since for $1 < p < \infty$, $f(x) = x^p$ is convex and $g(x) = \sqrt[p]{x}$ is concave. So we also have $\varrho_p(\alpha,\gamma) < \varrho_p(\alpha,\beta)+\varrho_p(\beta,\gamma)$ for $|\alpha| < \beta < \gamma$. The proof for the case $p < \infty$ is completed.

When $p = \infty$, (B.4) and $\alpha < \beta < \gamma$ imply $|\gamma| > \max\{|\alpha|, |\beta|\}$. So

$$\begin{aligned}
\varrho_\infty(\alpha,\gamma) &= \frac{\gamma-\alpha}{\gamma} = \frac{\gamma-\beta}{\gamma} + \frac{\beta-\alpha}{\gamma} \\
&= \frac{\gamma-\beta}{\gamma} + \frac{\beta-\alpha}{\max\{|\alpha|,|\beta|\}} + (\beta-\alpha)\left(\frac{1}{\gamma} - \frac{1}{\max\{|\alpha|,|\beta|\}}\right) \\
&< \varrho_\infty(\alpha,\beta)+\varrho_\infty(\beta,\gamma),
\end{aligned}$$

as was to be shown. ∎

LEMMA B.3. *Suppose functions $f(x)$ and $g(x)$ are defined on the interval $[a,b]$, and suppose $f(x)$ is convex and $g(x)$ concave. Let $x,y,z \in [a,b]$ and $x \le y \le z$. Then*

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(y)}{z-y} \quad \text{and} \quad \frac{g(y)-g(x)}{y-x} \ge \frac{g(z)-g(y)}{z-y}.$$

A proof of this lemma can be found in most mathematical analysis books; see, e.g., [31, section 1.4.4].

Acknowledgments. I thank Professor W. Kahan for his consistent encouragement and support, Professor J. Demmel for helpful discussions on open problems in this research area, and Professor B. N. Parlett for drawing my attention to Ostrowski's theorem. Thanks also go to Professor I. C. F. Ipsen for sending me the report [13]. Professor R. Bhatia and the referees' constructive comments, which improve the presentation considerably, are greatly appreciated.

REFERENCES

[1] J. Barlow and J. Demmel, *Computing accurate eigensystems of scaled diagonally dominant matrices*, SIAM J. Numer. Anal., 27 (1990), pp. 762–791.
[2] F. L. Bauer and C. T. Fike, *Norms and exclusion theorems*, Numer. Math., 2 (1960), pp. 137–141.
[3] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics 169, Springer-Verlag, New York, 1996.
[4] G. D. Birkhoff, *Tres observaciones sobre el algebra lineal*, Univ. Nac. de Tucuman Rev., Ser. A, 5 (1946), pp. 147–151.
[5] C. Davis and W. Kahan, *The rotation of eigenvectors by a perturbation.* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.
[6] P. Deift, J. Demmel, L.-C. Li, and C. Tomei, *The bidiagonal singular value decomposition and Hamiltonian mechanics*, SIAM J. Numer. Anal., 28 (1991), pp. 1463–1516.
[7] J. Demmel and W. Gragg, *On computing accurate singular values and eigenvalues of matrices with acyclic graphs*, Linear Algebra Appl., 185 (1993), pp. 203–217.
[8] J. Demmel and W. Kahan, *Accurate singular values of bidiagonal matrices*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 873–912.
[9] J. Demmel and K. Veselić, *Jacobi's method is more accurate than QR*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1204–1245.
[10] G. DiLena, R. I. Peluso, and G. Piazza, *Results on the relative perturbation of the singular values of a matrix*, BIT, 33 (1993), pp. 647–653.
[11] S. C. Eisenstat and I. C. F. Ipsen, *Relative perturbation bounds for eigenspaces and singular vector subspaces*, in Proceedings of the Fifth SIAM Conference on Applied Linear Algebra, J. G. Lewis, ed., SIAM, Philadelphia, PA, 1994, pp. 62–66.
[12] S. C. Eisenstat and I. C. F. Ipsen, *Relative perturbation techniques for singular value problems*, SIAM J. Numer. Anal., 32 (1995), pp. 1972–1988.
[13] S. C. Eisenstat and I. C. F. Ipsen, *Relative Perturbation Results for Eigenvalues and Eigenvectors of Diagonalisable Matrices*, Technical Report CRSC-TR96-6, Department of Mathematics, North Carolina State University, Raleigh, NC, 1996.
[14] K. V. Fernando and B. N. Parlett, *Accurate singular values and differential qd algorithms*, Numer. Math., 67 (1994), pp. 191–229.
[15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, MD, 1989.
[16] M. Gu and S. C. Eisenstat, *Relative Perturbation Theory for Eigenproblems*, Research Report YALEU/DCS/RR–934, Department of Computer Science, Yale University, New Haven, CT, 1993.
[17] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J., 20 (1953), pp. 37–39.
[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[19] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[20] W. Kahan, *Accurate Eigenvalues of a Symmetric Tridiagonal Matrix*, Technical Report CS41, Computer Science Department, Stanford University, Stanford, CA, 1966 (revised June 1968).

[21] R.-C. Li, *Bounds on perturbations of generalized singular values and of associated subspaces*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 195–234.

[22] R.-C. Li, *Norms of certain matrices with applications to variations of the spectra of matrices and matrix pencils*, Linear Algebra Appl., 182 (1993), pp. 199–234.

[23] R.-C. Li, *On Deflating Bidiagonal Matrices*, manuscript, Department of Mathematics, University of California, Berkeley, CA, 1994.

[24] R.-C. Li, *Relative Perturbation Theory:* (I) *Eigenvalue and Singular Value Variations*, Technical Report UCB//CSD-94-855, Computer Science Division, Department of EECS, University of California at Berkeley, 1994; LAPACK working note 85 (revised January 1996) available online at http://www.netlib.org/lapack/lawns/lawn84.ps

[25] R.-C. Li, *Relative Perturbation Theory:* (II) *Eigenspace and Singular Subspace Variations*, Technical Report UCB//CSD-94-856, Computer Science Division, Department of EECS, University of California at Berkeley, 1994; LAPACK working note 85 (revised January 1996 and April 1996), available at http://www.netlib.org/lapack/lawns/lawn85.ps.

[26] R.-C. Li, *Relative Perturbation Theory:* (III) *More Bounds on Eigenvalue Variation*, Linear Algebra Appl., 266 (1996), pp. 337–345.

[27] V. B. Lidskii, *The proper values of the sum and product of symmetric matrices*, Dokl. Akad. Nauk SSSR, 75 (1950), pp. 769–772 (in Russian). Translation by C. Benster available from the National Translation Center of the Library of Congress.

[28] R. Mathias, *Spectral perturbation bounds for positive definite matrices*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 959–980.

[29] R. Mathias and G. W. Stewart, *A block QR algorithm and the singular value decomposition*, Linear Algebra Appl., 182 (1993), pp. 91–100.

[30] L. Mirsky, *Symmetric gauge functions and unitarily invariant norms*, Quart. J. Math., 11 (1960), pp. 50–59.

[31] D. S. Mitrinovic, *Analytic Inequalities*, Springer-Verlag, New York, 1970.

[32] A. M. Ostrowski, *A quantitative formulation of Sylvester's law of inertia*, Proc. Nat. Acad. Sci. USA, 45 (1959), pp. 740–744.

[33] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[34] I. Slapničar, *Accurate Symmetric Eigenreduction by a Jacobi Method*, Ph.D. thesis, Fernuniversität–Gesamthochschule–Hagen, Fachbereich Mathematik, 1992.

[35] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.

[36] J.-G. Sun, *Perturbation analysis for the generalized singular value decomposition*, SIAM J. Numer. Anal., 20 (1983), pp. 611–625.

[37] C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal., 13 (1976), pp. 76–83.

[38] P.-Å. Wedin, *Perturbation bounds in connection with singular value decomposition*, BIT, 12 (1972), pp. 99–111.

[39] H. Wielandt, *An extremum property of sums of eigenvalues*, Proc. Amer. Math. Soc., 6 (1955), pp. 106–110.

[40] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.