Numer. Math. (1999) 81: 377–413

Numerische
Mathematik

The Lidskii-Mirsky-Wielandt theorem – additive and multiplicative versions*

Chi-Kwong Li, Roy Mathias

Department of Mathematics, The College of William and Mary, Williamsburg, Virginia 23187, USA; e-mail: ckli@math.wm.edu, mathias@math.wm.edu

Received November 20, 1996 / Revised version received January 27, 1998

Summary. We use a simple matrix splitting technique to give an elementary new proof of the Lidskii-Mirsky-Wielandt Theorem and to obtain a multiplicative analog of the Lidskii-Mirsky-Wielandt Theorem, which we argue is the fundamental bound in the study of relative perturbation theory for eigenvalues of Hermitian matrices and singular values of general matrices. We apply our bound to obtain numerous bounds on the matching distance between the eigenvalues and singular values of matrices. Our results strengthen and generalize those in the literature.

Mathematics Subject Classification (1991): 65F15

1. Introduction

Given an $n \times n$ Hermitian matrix A let $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$ denote its ordered eigenvalues. The *singular values* of an $m \times n$ matrix A, are defined by

$$\sigma_i(A) = \sqrt{\lambda_i(A^*A)}, \quad i = 1, 2, \ldots, q \equiv \min\{m, n\}.$$

An $m \times n$ matrix can have at most $q \equiv \min\{m, n\}$ non-zero singular values so we shall only consider its largest q singular values.

The Lidskii-Mirsky-Wielandt theorem (e.g., see [19, Theorem IV.4.8] or [2, Theorem 9.4]) states:

* Both authors were supported by grants from the National Science Foundation

Correspondence to: R. Mathias

Theorem 1.1 *Let A and E be $n \times n$ Hermitian matrices. Then for any indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ we have*

$$\sum_{j=1}^{k}[\lambda_{i_j}(A+E) - \lambda_{i_j}(A)] \leq \sum_{j=1}^{k} \lambda_j(E). \tag{1.1}$$

This is a very useful result in matrix theory. It is a majorization relation that implies a bound on the matching distance between the eigenvalues of A and those of $A+E$, namely,

$$\sum_{j=1}^{k} |\lambda_{i_j}(A+E) - \lambda_{i_j}(A)| \ \leq \ \sum_{j=1}^{k} \sigma_j(E). \tag{1.2}$$

This result can also be stated in terms of norms. A norm $\|\cdot\|$ on $\mathbb{R}^n$ is called a *symmetric norm* if it is both permutation invariant, that is,

$$\|Px\| = \|x\|, \ \ \forall\, x \in \mathbb{R}^n, \ \text{permutation matrices} \ P$$

and absolute, that is,

$$\|(x_i)_{i=1}^n\| = \|(|x_i|)_{i=1}^n\|, \ \ \forall\, x \in \mathbb{R}^n.$$

Symmetric norms are also sometimes called symmetric gauge functions. For every symmetric norm $\|\cdot\|$ on $\mathbb{R}^n$ there is a corresponding unitarily invariant norm, which we shall denote by $|||\cdot|||$, on the space of $n \times n$ matrices. The correspondence is given by

$$|||A||| = \| \ (\sigma_i(A))_{i=1}^n \ \|. \tag{1.3}$$

For example, the norms corresponding to the ℓ_2 and ℓ_∞ norms on $\mathbb{R}^n$ are the the Frobenius norm (denoted $|||\cdot|||_{\mathrm{F}}$) and spectral (or 2-) norm (denoted $|||\cdot|||_2$) respectively. See [8, Sect. 7.4] for further information and background on the connection between symmetric norms and unitarily invariant norms.

With this notation, and the theory of majorization, the result (1.2) is equivalent to the statement that

$$\|(\lambda_i(A+E) - \lambda_i(A))_{i=1}^n\| \ \leq \ |||E||| \tag{1.4}$$

is valid for any Hermitian A and E and any symmetric norm. This bound is frequently stated and referred to when perturbation theory is studied in the context of matrix inequalities. If one specializes this to the ℓ_2 and ℓ_∞ norms on $\mathbb{R}^n$ then one obtains the Wielandt-Hoffman and the Lidskii-Weyl inequalities:

$$\left(\sum_{i=1}^{n}(\lambda_i(A+E) - \lambda_i(A))^2\right)^{1/2} \leq |||E|||_{\mathrm{F}} \tag{1.5}$$

$$\max_{i=1,\ldots,n} |\lambda_i(A+E) - \lambda_i(A)| \leq |||E|||_2 \tag{1.6}$$

In numerical linear algebra one typically considers only the Wielandt-Hoffman and Lidskii-Weyl bounds, even though they are both just special cases of the fundamental Lidskii-Wielandt perturbation bound (1.2), or its norm form (1.4).

In multiplicative (or relative) perturbation theory one wants to bound quantities such as

$$\left|\frac{\lambda_i(S^*AS)-\lambda_i(A)}{\sqrt{\lambda_i(A)\lambda_i(S^*AS)}}\right| \text{ or } \left|\log\frac{\lambda_i(S^*AS)}{\lambda_i(A)}\right| \tag{1.7}$$

in terms of the distance between S and the set of unitary matrices. Ipsen has surveyed the results on multiplicative perturbation theory for eigenvalues and singular values [10]. The paper [12] contains numerous references on the subject of multiplicative perturbation bounds.

Our purpose is to give a multiplicative analog of the fundamental Lidskii-Wielandt bound and to show that it implies all the other relative perturbation bounds for eigenvalues of Hermitian matrices and singular values of general matrices. We also give an elementary new proof of (1.1).

In Sect. 2 we give a simple new proof of (1.1) and its multiplicative analog. We then show how this multiplicative analog of the Lidskii-Mirsky-Wielandt inequality implies similar results for singular values, generalized eigenvalues and generalized singular values. These results give upper and lower bounds on the ratio of the unperturbed and perturbed eigenvalues. It is a small step from here to the bounds in Sect. 2.3 which are multiplicative bounds on the relative perturbation in the eigenvalues. These bounds are *completely analogous* to the perturbation bound (1.2). It is only because we want to state results in terms of norms that we have to continue the development in Sects. 3, 4 and 5. Notice that Theorem 2.3 is the only new theorem in this paper – all the other results in this paper may be viewed as corollaries of Theorem 2.3.

In Sect. 3 we briefly review the connection between majorization and symmetric norms, which were defined in the the introduction.

In Sect. 4 we take logarithms of the results in Sect. 2 and use the connection between majorization and symmetric norms to derive a multiplicative version of the matching bound (1.2) in terms of norms and the relative distance

$$\mathrm{rd}(\alpha,\beta) = |\log(\alpha/\beta)| = |\log|\alpha| - \log|\beta||$$

between two real numbers α and β of the same sign. This relative distance, while not new, has not been used before in the context of multiplicative perturbation theory for eigenvalues. As we shall see in Sects. 4 and 6, there are good reasons for us to use this measure in our study.

In Sect. 5 we derive a multiplicative version of the matching bound (1.2) in terms of the relative error

$$\chi(\alpha,\beta) = \left|\sqrt{\frac{\alpha}{\beta}} - \sqrt{\frac{\beta}{\alpha}}\right| = \frac{|\alpha-\beta|}{\sqrt{\alpha\beta}}.$$

from the results in Sect. 2. In order to do this we need to prove a few scalar inequalities in Sect. 5.1. Our multiplicative perturbation bounds improve and generalize the work of R.-C. Li [12] who has extended many of the classical eigenvalue perturbation bounds from the additive context to the multiplicative context. The measure of relative difference $\chi(\cdot,\cdot)$ has been used by a number of authors but rd has much better mathematical properties and consequently the results in this section are harder to prove and are less clean than the results in the previous section.

In Sect. 6 we compare our results with those in the literature. There are many ways to state relative perturbation bounds. In this section, we also argue that the form used in Sect. 2.3, is the "right way" even though it is less familiar than the norm-wise results stated in Sects. 4 and 5. We end the section with a list of reasons why rd is a better measure of relative perturbation than χ

In Sect. 7 we use the matrix splitting technique to give a proof of Wielandt's min-max theorem. We also show that two natural multiplicative analogs of Wielandt's min-max theorem are false.

2. Matrix inequalities

The key ingredients of our proofs in this section are a matrix splitting technique and the following results of Weyl and Ostrowski:

Lemma 2.1 (Weyl's Inequality) [8, Corollary 4.3.3] *Let A and E be $n\times n$ Hermitian matrices with E positive semi-definite. Then*

(2.1) $$\lambda_i(A) \le \lambda_i(A+E) \qquad i = 1,2,\ldots,n.$$

Lemma 2.2 (Ostrowski's Inequality) [8, Theorem 4.5.9] *Let A be an $n\times n$ Hermitian matrix and let S be an $n\times n$ matrix with $\lambda_n(S^*S)\ge 1$. Then for $i=1,2,\ldots,n$, we have*

(2.2) $$|\lambda_i(A)| \le |\lambda_i(S^*AS)|,$$

or equivalently,

(2.3) $$1 \le \frac{\lambda_i(S^*AS)}{\lambda_i(A)}.$$

Here we use the fact that S^*AS and A have the same inertia and the convention that $0/0=1$ to ensure that the right side of (2.3) is always positive.

2.1. Proofs of Lidskii-Mirsky-Wielandt theorems

Let us prove (1.1). We may assume, without loss of generality, that $\lambda_k(E) = 0$; otherwise, we may replace E by $E - \lambda_k(E)I$ which reduces both sides of (1.1) by $k\lambda_k(E)$. Let $E = E_+ + E_-$ be the usual decomposition of E into positive and negative parts, i.e.,

$$E_+ = \sum_{i=1}^{n} \max\{0, \lambda_i(E)\} x_i x_i^* \quad \text{and} \quad E_- = \sum_{i=1}^{n} \min\{0, \lambda_i(E)\} x_i x_i^*,$$

where $E = \sum_{i=1}^{n} \lambda_i(E) x_i x_i^*$ is a spectral decomposition of E. Since $\lambda_k(E) = 0$ it follows that $\mathrm{trace}(E_+) = \sum_{i=1}^{k} \lambda_i(E)$, and hence we need only prove that

$$\sum_{j=1}^{k} [\lambda_{i_j}(A + E) - \lambda_{i_j}(A)] \leq \mathrm{trace}(E_+).$$

Using Weyl's inequality with the fact that $A + E \leq A + E_+$, we have $\lambda_{i_j}(A + E) \leq \lambda_{i_j}(A + E_+)$ and hence

$$\sum_{j=1}^{k} [\lambda_{i_j}(A + E) - \lambda_{i_j}(A)] \leq \sum_{j=1}^{k} [\lambda_{i_j}(A + E_+) - \lambda_{i_j}(A)]$$

Using Weyl's inequality again, this time with the fact that $E_+ \geq 0$, we have that $\lambda_j(A + E_+) - \lambda_j(A)$ is nonnegative for each $j = 1, \ldots, n$, and hence

$$\begin{aligned}\sum_{j=1}^{k} [\lambda_{i_j}(A + E_+) - \lambda_{i_j}(A)] &\leq \sum_{j=1}^{n} [\lambda_j(A + E_+) - \lambda_j(A)] \\ &= \mathrm{trace}(A + E_+) - \mathrm{trace}(A) \\ &= \mathrm{trace}(E_+),\end{aligned}$$

as desired.

Typically one derives the Lidskii-Mirsky-Wielandt bound from Wielandt's min-max representation of $\sum_{j=1}^{k} \lambda_{i_j}(A)$ [2] [19]. This is considerably more complicated than our approach.[1] In the last section we show that our technique can be used to prove Wielandt's min-max theorem. Now we use the same technique to prove a multiplicative analog of (1.1).

[1] The recent book [3] does contain two proofs of Lidskii-Mirsky-Wielandt (called Lidskii's inequality in [3]) that do not involve Wielandt's min-max representation. One uses induction, the other is somewhat similar to ours, but considerably more complicated. Thompson and Freede proved a generalization of the Lidskii-Wielandt inequalities (1.1) using rather elementary techniques and an inductive argument [21]. The fact that their result implies the Lidskii-Weilandt inequality appears to have been overlooked by many researchers.

Theorem 2.3 *Let A be an $n \times n$ Hermitian matrix and let $\tilde{A} = S^*AS$. Then for any indices $1 \le i_1 < i_2 < \cdots < i_k \le n$ such that $\lambda_{i_j}(A) \ne 0$ for $j = 1, \ldots, k$, we have*

$$\prod_{i=1}^{k} \lambda_{n+1-i}(S^*S) \le \prod_{j=1}^{k} \frac{\lambda_{i_j}(\tilde{A})}{\lambda_{i_j}(A)} \le \prod_{i=1}^{k} \lambda_i(S^*S). \tag{2.4}$$

Here we use the fact that the number of positive (respectively, negative) eigenvalues of $\tilde{A}$ is not larger than that of A to ensure that the middle expression in (2.4) is always nonnegative.

Proof. First assume that S is invertible. Let S have singular value decomposition $S = U \operatorname{diag}(d_1, d_2, \ldots, d_n)V$, where $d_1 \ge d_2 \ge \cdots \ge d_n > 0$ and U and V are unitary. Since eigenvalues are invariant under unitary similarities we may assume, without loss of generality, that $U = V = I$ and hence $S = \operatorname{diag}(d_1, d_2, \ldots, d_n)$. Furthermore, we may assume that $d_k = 1$; otherwise, we can replace the matrix S by S/d_k which will change all the left, middle and right quantities in (2.4) by the same multiple, namely, $1/d_k^{2k}$. (This is where we use the assumption that $\lambda_{i_j}(A) \ne 0$ for all j.)

Let $\tilde{D} = \operatorname{diag}(d_1, \ldots, d_k) \oplus I_{n-k}$. Then $\det(\tilde{D}^*\tilde{D}) = \prod_{i=1}^{k} \lambda_i(S^*S)$, and hence we need only to prove

$$\prod_{j=1}^{k} \frac{\lambda_{i_j}(\tilde{A})}{\lambda_{i_j}(A)} \le \det(\tilde{D}^*\tilde{D})$$

to get the upper bound in (2.4).

Using Ostrowski's inequality with the fact that $\lambda_n((S^{-1}\tilde{D})^*(S^{-1}\tilde{D})) \ge 1$, we have

$$\begin{aligned}\prod_{j=1}^{k} \frac{\lambda_{i_j}(S^*AS)}{\lambda_{i_j}(A)} &\le \prod_{j=1}^{k} \frac{\lambda_{i_j}((S^{-1}\tilde{D})^*S^*AS(S^{-1}\tilde{D}))}{\lambda_{i_j}(A)} \\ &= \prod_{j=1}^{k} \frac{\lambda_{i_j}(\tilde{D}^*A\tilde{D})}{\lambda_{i_j}(A)}. \end{aligned} \tag{2.5}$$

Using Ostrowski's inequality again, this time with the fact that $\lambda_n(\tilde{D}^*\tilde{D}) \ge 1$, we have $1 \le \frac{\lambda_i(\tilde{D}^*A\tilde{D})}{\lambda_i(A)}$ for all $i = 1, \ldots, n$, and hence

$$\prod_{j=1}^{k} \frac{\lambda_{i_j}(\tilde{D}^*A\tilde{D})}{\lambda_{i_j}(A)} \le \prod_{j=1}^{n} \frac{\lambda_i(\tilde{D}^*A\tilde{D})}{\lambda_i(A)} = \det(\tilde{D}^*\tilde{D}). \tag{2.6}$$

Combining (2.5) and (2.6), we get the desired upper bound.

Now, using the result established in the preceding paragraph, we have

$$\left(\prod_{j=1}^{k} \frac{\lambda_{i_j}(S^*AS)}{\lambda_{i_j}(A)}\right)^{-1} = \prod_{j=1}^{k} \frac{\lambda_{i_j}((S^{-1})^* S^* A S S^{-1})}{\lambda_{i_j}(S^*AS)}$$
$$\leq \prod_{j=1}^{k} \lambda_j(S^{-*}S^{-1})$$
$$= \left(\prod_{j=1}^{k} \lambda_{n+j-1}(S^*S)\right)^{-1} .$$

Taking the inverse of this inequality, we get the lower bound.

Now consider the singular case. The lower bound is trivially true in this case. Let $S(t) = S + tI$ with $0 < t < |\lambda|$ for any nonzero eigenvalue λ of S. Then $S(t)$ is invertible and $\lim_{t\downarrow 0} S(t) = S$. By the continuity of eigenvalues and (2.4) in the invertible case, we get the result for the singular case also. □

One may be tempted to remove the requirement in Theorem 2.3 that $\lambda_{i_j}(A) \neq 0$ and instead use the convention that $0/0 = 1$. To see that the resulting statement is not correct just consider the scalar example $A = 0$, $S = 2$.

We may summarize our proofs by saying that Weyl's qualitative bound (2.1) together with the linearity of the trace implies the much stronger bound (1.1), and that Ostrowski's qualitative bound (2.3) together with the multiplicativity of the determinant implies the much stronger bound (2.4). See [16] for another instance where a matrix splitting technique and Weyl's simple monotonicity result yield much stronger eigenvalue bounds.

To conclude this section, we remark that Theorem 2.3 provides some basic inequalities relating the eigenvalues of S^*AS and A. Similar to many other results in matrix inequalities, once some basic inequalities are available one can use the theory of majorization on real vectors to obtain a whole family of inequalities (e.g., see [9,15]). In fact, since we have a product inequality we may take positive powers of all the terms and obtain new inequalities. For example, in the development in Sect. 5 it is convenient to take the square root of the inequality of (2.11), but we could have equally well taken another power and obtained a different perturbation bound.

2.2. Related Lidskii-Mirsky-Wielandt inequalities

The inequalities (2.4) appear to be new for Hermitian matrices. In the special case that A is positive semi-definite (2.4) is equivalent to the singular value inequalities

$$\prod_{i=1}^{k} \frac{\sigma_{i_j}(A^{1/2}S)}{\sigma_{i_j}(A^{1/2})} \le \prod_{i=1}^{k} \sigma_i(S) \tag{2.7}$$

which is a special case of a more general family of inequalities [20] (or see [9, (3.3.53)] for the result without proof). The inequalities (2.7) were first proved by Gelfand and Naimark [7] (or see [3, Theorem III.4.5] or [9, (3.3.52)]).

Theorem 2.3 implies similar multiplicative inequalities for singular values, generalized eigenvalues, and generalized singular values.

Since the positive singular values of an $m \times n$ matrix A are the nonnegative square roots of the nonzero eigenvalues of AA^* or A^*A, two applications of (2.4) yield

Corollary 2.4 *Let A be an $m \times n$ matrix with q positive singular values. Suppose S is $m \times m$ and T is $n \times n$. Then*

$$\prod_{i=1}^{k} \sigma_{m+1-i}(S)\sigma_{n+1-i}(T) \le \prod_{j=1}^{k} \frac{\sigma_{i_j}(SAT)}{\sigma_{i_j}(A)} \le \prod_{i=1}^{k} \sigma_i(S)\sigma_i(T) \tag{2.8}$$

for any set of indices $1 \le i_1 < \cdots < i_k \le q$.

Proof. We shall prove the upper bound only since the lower bound can be proved in the same way. We have used the upper bound in (2.4) for the two inequalities below.

$$\begin{aligned}
\prod_{j=1}^{k} \frac{\sigma_{i_j}^2(SAT)}{\sigma_{i_j}^2(A)} &= \prod_{j=1}^{k} \frac{\lambda_{i_j}(SATT^*A^*S^*)}{\lambda_{i_j}(ATT^*A^*)} \frac{\lambda_{i_j}(ATT^*A^*)}{\lambda_{i_j}(AA^*)} \\
&\le \prod_{j=1}^{k} \lambda_i(SS^*) \prod_{j=1}^{k} \frac{\lambda_{i_j}(ATT^*A^*)}{\lambda_{i_j}(AA^*)} \\
&= \prod_{j=1}^{k} \sigma_i^2(S) \prod_{j=1}^{k} \frac{\lambda_{i_j}(T^*A^*AT)}{\lambda_{i_j}(A^*A)} \\
&\le \prod_{j=1}^{k} \sigma_i^2(S) \prod_{j=1}^{k} \lambda_i(T^*T) \\
&= \prod_{j=1}^{k} \sigma_i^2(S) \prod_{j=1}^{k} \sigma_i^2(T).
\end{aligned}$$

This is the square of the upper bound. □

Notice that when $m \ne n$, the pairing of the singular values of S and T in the upper and lower bounds is not the same.

Corollary 2.5 *Consider the generalized eigenvalue problems*

$$A_1 - \lambda A_2 \equiv S_1^* H_1 S_1^* - \lambda S_2^* H_2 S_2^*, \quad \text{and} \quad \tilde{A}_1 - \lambda \tilde{A}_2 \equiv S_1^* \tilde{H}_1 S_1^* - \lambda S_2^* \tilde{H}_2 S_2^*$$

where the H's and $\tilde{H}$'s are positive definite and S_2 is invertible. Let their eigenvalues be λ_i and $\tilde{\lambda}_i$. Let $F_1 = H_1^{-1/2} \tilde{H}_1 H_1^{-1/2}$ and let $F_2 = \tilde{H}_2^{-1/2} H_2 \tilde{H}_2^{-1/2}$. Then

$$\prod_{i=1}^{k} \lambda_{n+1-i}(F_1)\lambda_{n+1-i}(F_2) \le \prod_{j=1}^{k} \frac{\tilde{\lambda}_{i_j}}{\lambda_{i_j}} \le \prod_{i=1}^{k} \lambda_i(F_1)\lambda_i(F_2) \tag{2.9}$$

for any set of indices $1 \le i_1 < \cdots < i_k \le q$.

Corollary 2.6 *Consider the pairs of $n \times n$ matrices $(B_1, B_2) \equiv (G_1 S_1, G_2 S_2)$ and $(\tilde{B}_1, \tilde{B}_2) \equiv (\tilde{G}_1 S_1, \tilde{G}_2 S_2)$ where the G's and $\tilde{G}$'s and at least one of the S's is nonsingular. Let their generalized singular values be σ_i and $\tilde{\sigma}_i$, $i = 1, \ldots, n$. Then*

$$\begin{aligned} \prod_{i=1}^{k} \sigma_{n+1-i}(\tilde{G}_1 G_1^{-1})\sigma_{n+1-i}(G_2 \tilde{G}_2^{-1}) &\le \prod_{j=1}^{k} \frac{\tilde{\sigma}_{i_j}}{\sigma_{i_j}} \\ &\le \prod_{i=1}^{k} \sigma_i(\tilde{G}_1 G_1^{-1})\sigma_i(G_2 \tilde{G}_2^{-1}) \end{aligned} \tag{2.10}$$

for any set of indices $1 \le i_1 < \cdots < i_k \le q$.

2.3. Multiplicative matching bounds

In Theorem 2.3 we gave a multiplicative analog of (1.1). In this subsection we give a multiplicative analog of (1.2) that follows easily from Theorem 2.3. There are also matching bounds that follow from the results in the previous subsection.

Corollary 2.7 *Let A be an $n \times n$ Hermitian matrix and let $\tilde{A} = S^* A S$ where S is a invertible matrix. Using the fact that $S^* A S$ and A have the same inertia and the convention that $0/0 = 1$, we have*

$$\prod_{j=1}^{k} \max \left\{ \frac{\lambda_{i_j}(A)}{\lambda_{i_j}(\tilde{A})}, \frac{\lambda_{i_j}(\tilde{A})}{\lambda_{i_j}(A)} \right\} \le \prod_{j=1}^{k} \eta_j^2(S), \tag{2.11}$$

for any indices $1 \le i_1 < i_2 < \cdots < i_k \le n$, where

$$\eta_i(S) \equiv \max\{\sigma_{\tau(i)}(S), \sigma_{\tau(i)}^{-1}(S)\}, \tag{2.12}$$

where τ is a permutation such that $\eta_1 \ge \cdots \ge \eta_n$.

Proof. Let the indices $l_1 < \cdots < l_{k_1}$ be indices from $i_1 < \cdots < i_k$ for which

$$\frac{\lambda_{i_j}(A)}{\lambda_{i_j}(\tilde{A})} \leq \frac{\lambda_{i_j}(\tilde{A})}{\lambda_{i_j}(A)}.$$

Let $m_1 < \cdots < m_{k_2}$ be the remaining indices from $i_1 < \cdots < i_k$. Then, using the upper bound in (2.4) to bound the first product and the lower bound in (2.4) to bound the second, we have

$$\begin{aligned}
&\prod_{j=1}^{k} \max\left\{\frac{\lambda_{i_j}(A)}{\lambda_{i_j}(\tilde{A})}, \frac{\lambda_{i_j}(\tilde{A})}{\lambda_{i_j}(A)}\right\} \\
&= \prod_{j=1}^{k_1} \frac{\lambda_{l_j}(A)}{\lambda_{l_j}(\tilde{A})} \prod_{j=1}^{k_2} \frac{\lambda_{m_j}(\tilde{A})}{\lambda_{m_j}(A)} \\
&\leq \prod_{j=1}^{k_1} \lambda_j(S^*S) \prod_{j=1}^{k_2} \lambda_{n+1-j}^{-1}(S^*S) \\
&\leq \prod_{j=1}^{k_1} \max\{\lambda_j(S^*S), \lambda_j^{-1}(S^*S)\} \\
&\qquad \times \prod_{j=1}^{k_2} \max\{\lambda_{n+1-j}(S^*S), \lambda_{n+1-j}^{-1}(S^*S)\} \\
&\leq \prod_{j=1}^{k} \eta_j^2(S).
\end{aligned}$$

The final inequality follows from the definition of the η_i's–the penultimate line is the product of k of the η_i's, while in the last line we have the product of the largest k of the η_i's. □

The bound (2.11) is the multiplicative analog of (1.2). The use of η_i may seem rather unnatural because its definition is awkward, but in fact the η_i's are entirely analogous to the singular values of a Hermitian that appear on the right hand side of (1.2). To see this note that we could define the singular values of a *Hermitian* matrix E by

$$\sigma_i(E) = \max\{\lambda_{\tau(i)}(E), -\lambda_{\tau(i)}(E)\},$$

where the permutation τ is such that $\sigma_1(E) \geq \cdots \geq \sigma_n(E)$. Another analogy between the η_i's and singular values is that the singular values of a general matrix $m \times n$ matrix X may be defined as the largest $\min\{m, n\}$ eigenvalues of the Jordan-Wielandt matrix

$$\begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}.$$

The η_i's corresponding to an $n \times n$ invertible matrix S may be defined as the largest n singular values of the matrix

$$\begin{pmatrix} S & 0 \\ 0 & S^{-1} \end{pmatrix}.$$

The Jordan-Wielandt matrix has proved useful in studying singular values, perhaps this $2n \times 2n$ matrix will be useful in studying the η_i's.

We feel that (2.11) is the right analog of (1.2), and we shall discuss this in Sect. 6.2. One can use it to derive other matching bounds. In the next section, we shall derive one such bound that generalizes results in the literature.

One can prove a similar perturbation bound for singular values, generalized eigenvalues and generalized singular values. The results are straight forward and so we will only give the result for singular values. We shall assume that $m = n$ in order to simplify the notation.

Corollary 2.8 *Let A, S and T be $n \times n$ matrices. Assume that S and T are invertible. Then*

$$\prod_{j=1}^{k} \max\left\{ \frac{\sigma_{i_j}(SAT)}{\sigma_{i_j}(A)}, \frac{\sigma_{i_j}(A)}{\sigma_{i_j}(SAT)} \right\} \le \prod_{i=1}^{k} \eta_i \tag{2.13}$$

for any set of indices $1 \le i_1 < \cdots < i_k \le q$, where

$$\eta_i = \max\{\sigma_{\tau(i)}(S)\sigma_{\tau(i)}(T), \sigma_{\tau(i)}^{-1}(S)\sigma_{\tau(i)}^{-1}(T)\},$$

where τ is a permutation such that $\eta_1 \ge \cdots \ge \eta_n$.

Note that

$$\prod_{i=1}^{k} \eta_i \le \prod_{i=1}^{k} [\eta_i(S)\eta_i(T)].$$

The idea of using $\max\{\alpha/\beta, \beta/\alpha\}$ as the measure of the relative distance from α to β (when α and β have the same sign) has been suggested before. For example, Olver [18] proposes

$$|\log(\alpha/\beta)| = \log \max\{\alpha/\beta, \beta/\alpha\}$$

as the appropriate measure because it is a metric and it interacts with multiplication and division more neatly than, say, $|(\alpha - \beta)/\alpha|$.

The quantity $\max\{x, x^{-1}\}$ is the maximum of x and its multiplicative inverse is the multiplicative analog of

$$|x| \equiv \max\{x, -x\},$$

the (additive) absolute value. This multiplicative absolute value occurs in many places in this paper, especially in Sect. 5. Perhaps the presentation and the analogy with the additive case would have been clearer if we had defined a symbol to denote this multiplicative absolute value, say

$$|x|_\times \equiv \max\{x, x^{-1}\}.$$

3. Majorization and norms

In this section, we briefly review the aspects of majorization and symmetric norms that we shall use in the next two sections.

Recall that for two vectors $x, y \in \mathbb{R}^n$, we say that y *weakly majorizes* x (denoted by $x \prec_w y$) if the sum of the k largest entries of x is not larger than that of y for $k = 1, \ldots, n$. If in addition the sum of all the entries of x is equal to that of y then we say that y *majorizes* x. See [15] for further information on majorization and numerous applications.

The theory of majorization greatly simplifies our proofs because there are many functions that preserve weak majorization. The notation of majorization simplifies also the presentation of our results. For example, the fundamental additive perturbation bound for eigenvalues of Hermitian matrices (1.1) can be stated simply as

$$(\lambda_i(A+E) - \lambda_i(A))_{i=1}^n \prec_w (\lambda_i(E))_{i=1}^n. \tag{3.1}$$

The family of inequalities (1.1) states that (3.1) is a weak majorization. However, consideration of the trace shows that (3.1) is actually a majorization.

There is an intimate connection between symmetric norms and majorization. It is summarized in the following lemma. See e.g., [8, Theorem 7.4.45] for the usual proof of this fact or [11] for a new approach.

Lemma 3.1 *Let $x, y \in \mathbb{R}_+^n$. Then*

$$x \prec_w y$$

if and only if

$$\|x\| \leq \|y\|$$

for all symmetric norms $\|\cdot\|$ on $\mathbb{R}^n$.

In view of this result we can equivalently state results in terms of weak majorization of vectors or in terms of norm inequalities. We shall do the latter as it is the standard approach in numerical linear algebra.

In the context of the perturbation of singular values of rectangular matrices one needs to consider norms on spaces of different sizes. (The reader who is not interested in the strongest possible result in the rectangular case may assume that $m = n$ and skip the technicalities in this paragraph.) We shall use the following natural convention. Given $m \leq n$, the symmetric norm $\|\cdot\|$ on $\mathbb{R}^m$ induces a norm on $\mathbb{R}^n$ by

$$\|x\| = \max_{i_1 < \cdots < i_m} \|(x_{i_1}, \ldots, x_{i_m})\|.$$

(That is, $\|x\|$ is the largest norm of any vector consisting of m of the n components of x. It is easy to show that the maximum is obtained when one chooses the largest m components of x.) Consequently, a symmetric norm on $\mathbb{R}^m$ induces a unitarily invariant norm $|||\cdot|||$ on the space of $n \times n$ matrices. Note that the resulting unitarily invariant norm is not what one might at first expect. For example, if $\|\cdot\|$ is the Euclidean norm of $\mathbb{R}^2$, and we take $n = 4$ then symmetric norm induced on $\mathbb{R}^4$ is the square root of the sum of the squares of the two largest elements of the 4-vector and the unitarily invariant norm it induced on the space of say, 4×4 matrices is

$$|||X||| = (\sigma_1^2(X) + \sigma_2^2(X))^{1/2},$$

and not the Frobenius norm on the space of 4×4 matrices:

$$|||X|||_{\mathrm{F}} = (\sigma_1^2(X) + \sigma_2^2(X) + \sigma_3^2(X) + \sigma_4^2(X))^{1/2}.$$

Never-the-less, it is true that $|||X||| \leq |||X|||_{\mathrm{F}}$ and this is sufficient to deduce (5.11) from (5.10).

It is easy to prove the following natural lemma on the norm induced on $\mathbb{R}^n$ by $\|\cdot\|$ on $\mathbb{R}^m$ for $m \leq n$.

Lemma 3.2 *Let $m \leq n$ be positive integers. Let $\|\cdot\|$ be a symmetric norm on $\mathbb{R}^n$ and let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then for any indices $i_1 < \cdots < i_m$*

$$\|(x_{i_j})_{j=1}^m\| \;\leq\; \|(x_j)_{j=1}^n\|.$$

4. Perturbation bounds using $|\log(\alpha/\beta)|$

One can completely avoid the complications of the next section by using

$$\mathrm{rd}(\alpha, \beta) = |\log(\alpha/\beta)| = |\log|\alpha| - \log|\beta|\,|$$

as the measure of the relative distance between real numbers α and β with the same sign. In fact, if $\beta = \alpha\delta$ for some $\delta > 0$, then

$$rd(\alpha, \beta) = |\log \delta|.$$

As we shall see in the following discussion, if a matrix A is perturbed to a matrix $B = AS$, then the norm of the matrix $\log|S|$, where $|S| \equiv (S^*S)^{1/2}$, plays an important role in deriving perturbation bounds. In some sense, this can be viewed as a generalization of the scalar case. Also, notice that rd is a metric, unlike χ. The idea of using $\mathrm{rd}(\cdot,\cdot)$ as the measure of the relative distance from α to β (when α and β have the same sign) has been suggested before. For example, Olver [18] proposes it as the appropriate measure because it is a metric and it interacts with multiplication and division more neatly than, say, $|(\alpha - \beta)/\alpha|$.

Since the results are so straight forward we shall give the proofs of the first two results only. They contain all the elements essential to the other proofs.

Proposition 4.1 *Let A and $\tilde{A} = S^*AS$ be $n \times n$ Hermitian matrices, where S is invertible. Let λ_i and $\tilde{\lambda}_i$, $i = 1, 2, \ldots, n$ be the eigenvalues of A and $\tilde{A}$. Then*

$$\left(\, |\log(\lambda_i/\tilde{\lambda}_i)| \, \right)_{i=1}^n \prec_w (|\log \lambda_i(S^*S)|)_{i=1}^n \tag{4.1}$$

Let $\|\cdot\|$ be any symmetric norm and let $|||\cdot|||$ be the corresponding unitarily invariant norm. Then

$$\|(|\log(\lambda_i/\tilde{\lambda}_i)|)_{i=1}^n\| \le |||\log(S^*S)|||. \tag{4.2}$$

Proof. Taking logarithms of the multiplicative perturbation bound (2.11) gives the majorization (4.1). By Lemma 3.1 this majorization is equivalent to the corresponding norm inequality for all symmetric norms:

$$||(\, |\log(\lambda_i/\tilde{\lambda}_i)| \,)_{i=1}^n|| \le ||(\, \log(\lambda_i(S^*S)) \,)_{i=1}^n||.$$

Now using functional calculus and the correspondence between symmetric norms and unitarily invariant norms (1.3) we have

$$|| \, (\lambda_i(\log S^*S))_{i=1}^n \, || = ||| \, \log(S^*S) \, |||,$$

which combined with the previous inequality gives us (4.2). □

Notice how straightforwardly Proposition 4.1 follows from the multiplicative bounds in Sect. 2 and a standard majorization result involving symmetric norms. Compare this with the proof of Proposition 5.3 the corresponding result in Sect. 5 which depends on the preliminary result Lemma 5.1.

Proposition 4.2 *Let A and $\tilde{A}$ be $m \times n$ and have singular values σ_i and $\tilde{\sigma}_i$, respectively. Suppose that $\tilde{A} = SAT$ where S and T are invertible. Let $\|\cdot\|$ be any symmetric norm symmetric norm on $\mathbb{R}^q$, ($q = \min\{m, n\}$) and let $|||\cdot|||$ be the corresponding unitarily invariant norm. Then*

$$\|(|\log(\sigma_i/\tilde{\sigma}_i)|)_{i=1}^q\| \le \frac{1}{2} \left\{ \, |||\log(S^*S)||| + |||\log(T^*T)||| \, \right\}. \tag{4.3}$$

Since $\log(S^*S) = 2\log|S|$ we could have eliminated the factor of $1/2$ in (4.3) at the cost of replacing S^*S and T^*T by $|S|$ and $|T|$.

Proof. Let $\hat{A} = SA$, have singular values $\hat{\sigma}_i$. Then *using the fact that* $\mathrm{rd}(\cdot,\cdot)$ *is a metric* for the first inequality, the triangle inequality for the second, and Lemma 3.2 for the third, we have

$$\begin{aligned}
\|(|\log(\sigma_i/\tilde{\sigma}_i)|)_{i=1}^q\| &= \|(|\log(\sigma_i/\hat{\sigma}_i) + \log(\hat{\sigma}_i/\tilde{\sigma}_i)|)_{i=1}^q\| \\
&\le \|(|\log(\sigma_i/\hat{\sigma}_i)| + |\log(\hat{\sigma}_i/\tilde{\sigma}_i)|)_{i=1}^q\| \\
&\le \|(|\log(\sigma_i/\hat{\sigma}_i)|)_{i=1}^q\| + \|(|\log(\hat{\sigma}_i/\tilde{\sigma}_i)|)_{i=1}^q\|. \\
&\le \|(|\log(\sigma_i/\hat{\sigma}_i)|)_{i=1}^m\| + \|(|\log(\hat{\sigma}_i/\tilde{\sigma}_i)|)_{i=1}^n\|.
\end{aligned}$$

Now we need only bound the effect of S on the singular values of A and that of T on the singular values of SA separately. Each of these can be done by appealing to Proposition 4.1. For S:

$$\|(|\log(\sigma_i/\hat{\sigma}_i)|)_{i=1}^n\| = \frac{1}{2}\left\|\left(\log\left(\frac{\lambda_i(A^*A)}{\lambda_i(S^*A^*AS)}\right)\right)_{i=1}^n\right\| \\ \le \frac{1}{2}|||\log S^*S\ |||.$$

Using the same argument for T and adding the bounds gives (4.3). $\square$

Notice how much simpler the proof of Proposition 4.2 is than the proof of the corresponding result expressed in terms of χ – Corollary 5.5. Corollary 5.5 is proved from scratch, via the intermediate result Proposition 5.4, and the technical results Lemmata 5.1 and 5.2. We were not able to easily deduce it from Proposition 5.3 – the result in terms of χ for Hermitian matrices. R.-C. Li was able to deduce his singular value result from his eigenvalue result together with the weak triangle inequality for χ, i.e. (5.4), but his bound is weaker than ours.

When analyzing the accuracy of Jacobi's method [4, 17] applied to graded positive definite matrices one has a positive definite matrix written as $A = C^*HC$ which one perturbs to $\tilde{A} = C^*\tilde{H}C$ and one would like relative perturbation bounds on the eigenvalues in terms of $\Delta H = H - \tilde{H}$ and H rather than ΔA and A. Since

$$\lambda_i(A) = \sigma_i^2(H^{1/2}C),$$

and

$$\lambda_i(\tilde{A}) = \sigma_i^2((\tilde{H}^{1/2}H^{-1/2})H^{1/2}C).$$

Proposition 4.2 now yields

Proposition 4.3 *Let $A = C^*HC$ and $\tilde{A} = C^*\tilde{H}C$ be $n\times n$ positive definite matrices with eigenvalues λ_i, $\tilde{\lambda}_i$. Then for any symmetric norm $\|\cdot\|$ on $\mathbb{R}^n$ and the corresponding unitarily invariant norm $|||\cdot|||$*

$$\|(|\log(\lambda_i/\tilde{\lambda}_i)|)_{i=1}^n\| \le |||\log H^{-1/2}\tilde{H}H^{-1/2}|||, \tag{4.4}$$

or equivalently

$$\|(|\log(\lambda_i/\tilde{\lambda}_i)|)_{i=1}^n\| \le |||\log(I+E)||| \tag{4.5}$$

where $E = H^{-1/2}\Delta H H^{-1/2}$.

To see that the right hand side is indeed unchanged when the roles of H and $\tilde{H}$ are reversed notice that the eigenvalues of $H^{-1/2}\tilde{H}H^{-1/2}$ are the inverses of those of $\tilde{H}^{-1/2}H\tilde{H}^{-1/2}$.

Proposition 4.4 *Consider the generalized eigenvalue problems*

$$A_1 - \lambda A_2 \equiv S_1^* H_1 S_1 - \lambda S_2^* H_2 S_2, \ \text{ and } \ \tilde{A}_1 - \lambda \tilde{A}_2 \equiv S_1^* \tilde{H}_1 S_1 - \lambda S_2^* \tilde{H}_2 S_2$$

where the H's and $\tilde{H}$'s are positive definite and at least one of the S_i's is invertible. Let $||\cdot||$ be any symmetric norm on $\mathbb{R}^n$ and let $|||\cdot|||$ be the corresponding unitarily invariant norm. Then

$$\begin{aligned} \|(\log(\lambda_i/\tilde{\lambda}_i))_{i=1}^n)|| \le |||\log(H_1^{-1/2}\tilde{H}_1 H_1^{-1/2})||| \\ +|||\log(H_2^{-1/2}\tilde{H}_2 H_2^{-1/2})|||, \end{aligned} \tag{4.6}$$

or equivalently,

$$\|(\log(\lambda_i/\tilde{\lambda}_i))_{i=1}^n\| \le \frac{1}{2}\left[|||\log(I+E_1)||| + |||\log(I+E_2)|||\right], \tag{4.7}$$

where $E_i = H_i^{-1/2} \Delta H_i H_i^{-1/2}$.

The bound (4.6), which expresses the perturbation multiplicatively, is perhaps more natural than (4.7) which expresses the perturbation additively. In the rest of the paper we shall express perturbations additively as that is what is usually done in numerical linear algebra, but all of our perturbations could be expressed multiplicatively also.

Proposition 4.5 *Consider the pairs of $n \times n$ matrices $(B_1, B_2) \equiv (G_1 S_1, G_2 S_2)$ and $(\tilde{B}_1, \tilde{B}_2) \equiv (\tilde{G}_1 S_1, \tilde{G}_2 S_2)$ where the G's and $\tilde{G}$'s and at least one of the S's are nonsingular. Let the generalized singular values be σ_i and $\tilde{\sigma}_i$, $i = 1, \ldots, n$. Let $\|\cdot\|$ be any symmetric norm on $\mathbb{R}^n$ and let $|||\cdot|||$ be the corresponding unitarily invariant norm. Let $E_i = \Delta G_i (G_i)^{-1}$. Then*

$$\begin{aligned} \|(\chi(\sigma_1, \tilde{\sigma}_1), \ldots, \chi(\sigma_n, \tilde{\sigma}_n))\| \le |||\log|I+E_1|\,||| \\ +|||\log|I+E_2|\,|||. \end{aligned} \tag{4.8}$$

5. Perturbation bounds using χ

For nonzero real numbers α and β with $\alpha\beta > 0$, define

$$\chi(\alpha, \beta) = \left|\sqrt{\frac{\alpha}{\beta}} - \sqrt{\frac{\beta}{\alpha}}\right| = \frac{|\alpha - \beta|}{\sqrt{\alpha\beta}}$$

and define $\chi(0,0) = 0$. The function χ is a measure of relative separation. Notice that $\chi(\max\{x, x^{-1}\}, 1) = \chi(x, 1)$ for any scalar $x > 0$.

In this section, we combine the multiplicative bounds in Sect. 2 with the theory of majorization to obtain perturbation bounds in terms of χ that improve a number of results in the multiplicative perturbation theory literature.

5.1. Some scalar inequalities for χ

First, we establish some inequalities for χ which allow us to derive majorizations in terms of χ from the the product majorizations of Sect. 2 (or equivalently log majorizations like (5.1)).

Lemma 5.1 *Let x and y be positive n vectors such that ,*

$$(\log\max\{x_1,x_1^{-1}\},\dots,\log\max\{x_n,x_n^{-1}\}) \prec_w (\log\max\{y_1,y_1^{-1}\},\dots,\log\max\{y_n,y_n^{-1}\}). \tag{5.1}$$

Then

$$(|x_1-1/x_1|,\dots,|x_n-1/x_n|) \prec_w (|y_1-1/y_1|,\dots,|y_n-1/y_n|). \tag{5.2}$$

Consequently, for any symmetric norm

$$\|(|x_1-1/x_1|,\dots,|x_n-1/x_n|)\| \le \|(|y_1-1/y_1|,\dots,|y_n-1/y_n|)\|. \tag{5.3}$$

Proof. Since the function $g(t)=\mathrm{e}^t-\mathrm{e}^{-t}$ is convex and increasing on $[0,\infty)$, we have

$$(g(\log\max\{x_1,x_1^{-1}\}),\dots,g(\log\max\{x_n,x_n^{-1}\})) \prec_w (g(\log\max\{y_1,y_1^{-1}\}),\dots,g(\log\max\{y_n,y_n^{-1}\}))$$

by a result of Schur (e.g., see [15, Chapter 3,C.1]). This is the desired majorization since

$$g(\log\max\{t,t^{-1}\})=|t-t^{-1}|$$

for any positive number t.

By Lemma 3.1 this majorization implies the inequality (5.3). □

Let r,s be positive numbers. One might hope that the relative separation between rs and 1 is at most the sum of the relative separation between r and 1 and the relative separation between r and rs. That is

$$\chi(rs,1)\le\chi(r,1)+\chi(rs,r)=\chi(r,1)+\chi(s,1).$$

Unfortunately this inequality is true if and only if $(r-1)(s-1)\le 0$. When dealing with multiplicative perturbation bounds that contain two multiplicative perturbations, for example, the perturbation of singular values when one allows multiplicative perturbations on both sides, one would like to bound quantities like $\chi(rs,1)$ in terms of a function of r and a function of s. One way is to use the approach of R.-C. Li, [12, Lemma 6.1]:

$$\chi(rs,1)\le\frac{\chi(r,1)+\chi(s,1)}{1-\frac{1}{8}\chi(r,1)\chi(s,1)}. \tag{5.4}$$

Another is the following lemma.

Lemma 5.2 *Let $r, s > 0$. Then*

$$\chi(rs, 1) \le \frac{1}{2}\left[\chi(r^2, 1) + \chi(s^2, 1)\right] \tag{5.5}$$

The inequality (5.4) is an equality if and only if at least one of r and s is 1, the inequality (5.5) is an equality if and only if $r = s$. Thus neither is uniformly stronger than the other. In the limit as r and s approach 1, which is when one will usually use these bounds they are about the same.

A third way to deal with the problem is to decide to bound $\chi(\sqrt{rs}, 1)$ instead of $\chi(rs, 1)$. If we do this, then we can use Lemma 5.2 to conclude that

$$\chi(\sqrt{rs}, 1) \le \chi(r, 1) + \chi(s, 1).$$

Proof of Lemma 5.2. Without loss of generality we may assume that both r and s are greater than or equal to 1. Since if they are not, replace those that are less than one by their inverses. This can only increase the left hand side, but will leave the right hand side unchanged.

Now, without loss of generality we may assume that $r \ge s$. Now regard s as fixed and r as variable. When $r = s$ (5.5) is as equality. One can easily check that

$$\frac{d}{dr}\left[\frac{1}{2}(\chi(r^2, 1) + \chi(s^2, 1)) - \chi(rs, 1)\right]$$

is positive when ever $r > s$. This ensures that the inequality (5.5) holds, and indeed that it is a strict inequality unless $r = s$. □

5.2. Perturbation bounds

We are now ready to derive normwise perturbation bounds in terms of χ from the results in Sect. 2.

In numerical linear algebra one often states perturbation bounds in terms of the norm of the perturbing matrix (or some function of the matrix), rather than in terms of the singular values of the perturbing matrix. For this reason we state the bounds in this section in terms of the norm of the perturbing matrix. We could have equivalently stated them as weak majorizations (see for example (5.6)) or as a norm bound in terms of the singular values of the perturbing matrix (see for example (5.7)). In the interests of brevity we have given these different forms in Proposition 5.3 only.

Proposition 5.3 *Let A and $\tilde{A} = S^*AS$ be $n \times n$ Hermitian matrices, where S is invertible. Let λ_i and $\tilde{\lambda}_i$, $i = 1, 2, \ldots, n$ be the eigenvalues of A and $\tilde{A}$. Then*

$$(\chi(\lambda_1, \tilde{\lambda}_1), \ldots, \chi(\lambda_n, \tilde{\lambda}_n)) \prec_w (\chi(\lambda_1(S^*S), 1), \ldots, \chi(\lambda_n(S^*S), 1)). \tag{5.6}$$

Consequently, for any symmetric norm,

$$\begin{aligned} &\|(\chi(\lambda_1,\tilde{\lambda}_1),\ldots,\chi(\lambda_n,\tilde{\lambda}_n))\| \\ \text{(5.7)} \qquad &\quad \le \|(\chi(\lambda_1(S^*S),1),\ldots,\chi(\lambda_n(S^*S),1))\|. \end{aligned}$$

Equivalently, for any symmetric norm $\|\cdot\|$ *and corresponding unitarily invariant norm* $|||\cdot|||$

$$\text{(5.8)} \qquad \|(\chi(\lambda_1,\tilde{\lambda}_1),\ldots,\chi(\lambda_n,\tilde{\lambda}_n))\| \le |||S^* - S^{-1}|||.$$

In particular

$$\sqrt{\sum_{i=1}^{n}\chi^2(\lambda_i,\tilde{\lambda}_i)} \le |||S^* - S^{-1}|||_{\mathrm{F}}$$

and

$$\max_{i=1,\ldots,n}\chi(\lambda_i,\tilde{\lambda}_i) \le |||S^* - S^{-1}|||_2.$$

Proof. Since S is invertible, A and $\tilde{A}$ have the same inertia. Let $x, y \in \mathbb{R}^n$ be such that $x_i = \sqrt{\lambda_i(A)/\lambda_i(\tilde{A})}$ with the convention that $0/0 = 1$, and $y_i = \sigma_i(S)$, for $i = 1,\ldots,n$. Taking square root of (2.11) in Corollary 2.7, we see that the hypothesis of Lemma 5.1 holds. The first two inequalities now follow easily from Lemma 5.1 and the equality $\chi(\alpha,\beta) = \left|\sqrt{\alpha/\beta} - \sqrt{\beta/\alpha}\right|$.

Notice that the two unnumbered inequalities are just special cases of (5.8) and the left hand sides of (5.7) and (5.8) are the same. To establish the proposition we need only show that the right hand side of (5.7) is the same as that of (5.8). To see this note that the right hand side of (5.7) is the norm of the vector whose ith component is

$$\chi(\lambda_i(S^*S),1) = \left|\sqrt{\lambda_i(S^*S)} - \sqrt{\lambda_i^{-1}(S^*S)}\right| = |\sigma_i(S) - \sigma_i^{-1}(S)|.$$

Let $S = U\Sigma V^*$ be a singular value decomposition of S. Then

$$S^* - S^{-1} = V(\Sigma - \Sigma^{-1})U^*$$

has singular values

$$|\sigma_i(S) - \sigma_i^{-1}(S)|, \quad i = 1,\ldots,n.$$

Thus the vector $(\sigma_i(S^* - S^{-1}))_{i=1}^n$ is just a permutation of the vector $(\chi(\lambda_i(S^*S),1))_{i=1}^n$, and consequently they have the same norm. □

An $n \times n$ matrix S is unitary if and only if the eigenvalues of S^*S are all 1, or equivalently if and only if $S^* = S^{-1}$. The relative distance between $\lambda_i(S^*S)$ and 1 is $\chi(\lambda_i(S^*S),1)$. Thus (5.7) and (5.8) are the pleasing statement that, in any symmetric norm, the relative perturbation in the eigenvalues of A caused by the congruence S is bounded by the relative distance from S to the set of unitary matrices.

Notice that in the proof of Proposition 5.3 we took the square root of the multiplicative inequality (2.11). We could have taken some other positive power and obtained another inequality. In fact, we could have stated a pth power version of all our results in this section but we found the additional generality to be useful only in the context of Corollary 5.5.

Now we turn to singular values of possibly rectangular matrices. Proposition 5.4 is the strongest possible result in the rectangular case, but it is not at all satisfactory. Firstly the definition of the ρ's is rather complicated, especially in the rectangular case. Secondly, it would be nice to separate the effect of S and T on the right hand side. We give a number of bounds that do not involve ρ and that do separate the effects of S and T in Corollary 5.5. Readers may wish to omit the next technical result and skip straight to Corollary 5.5.

Proposition 5.4 *Let A and $\tilde{A}$ be $m \times n$ with singular values $\sigma_1 \geq \cdots \geq \sigma_q$ and $\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_q$, respectively, where $q = \min\{m, n\}$. Suppose that $\tilde{A} = SAT$ where S and T are invertible. Let $\rho_1, \ldots, \rho_q$ consist of the first q_1 terms, $0 \leq q_1 \leq q$, of the decreasing sequence $\{\sigma_j(S)\sigma_j(T)\}_{j=1}^q$ and the first $q - q_1$ terms of the increasing sequence $\{\sigma_{m-j+1}(S)\sigma_{n-j+1}(T)\}_{j=1}^q$ so that the q numbers*

$$\max\{\rho_i, \rho_i^{-1}\}, \qquad i = 1, \ldots, q,$$

are the largest q of the $2q$ numbers

$$\sigma_i(S)\sigma_i(T), \quad i = 1, \ldots, q,$$

and

$$\sigma_{m-i}^{-1}(S)\sigma_{n-i}^{-1}(T), \quad i = 1, \ldots, q.$$

Then

$$(\chi(\sigma_1, \tilde{\sigma}_1), \ldots, \chi(\sigma_q, \tilde{\sigma}_q)) \prec_w (\chi(\rho_1, 1), \ldots, \chi(\rho_q, 1)), \tag{5.9}$$

and for any symmetric norm

$$\|(\chi(\sigma_1, \tilde{\sigma}_1), \ldots, \chi(\sigma_q, \tilde{\sigma}_q))\| \leq \|(\chi(\rho_1, 1), \ldots, \chi(\rho_q, 1))\|$$

An alternative way to define q_1 and ρ_i's is the following. Let q_1 be the largest integer between 1 and q such that

$$\max\{\sigma_{q_1}(S)\sigma_{q_1}(T), (\sigma_{q_1}(S)\sigma_{q_1}(T))^{-1}\}$$
$$\geq \max\{\sigma_{m-q_1+1}(S)\sigma_{n-q_1+1}(T), (\sigma_{m-q_1+1}(S)\sigma_{n-q_1+1}(T))^{-1}\}.$$

If there is no such integer then set $q_1 = 0$. Now for $i = 1, \ldots, q$, define

$$\rho_i = \begin{cases} \sigma_i(S)\sigma_i(T) & \text{if} \quad i \le q_1, \\ \sigma_{m-(q_1-i)}(S)\sigma_{n-(q_1-i)}(T) & \text{otherwise} \end{cases} .$$

In any event, one can check that if $m = n$ then we may take $\rho_i = \sigma_i(S)\sigma_i(T)$ – a considerable simplification!

Proof of Proposition 5.4. Since S and T are invertible, A and $\tilde{A}$ have the same rank, i.e., the same number of nonzero singular values. Let $x, y \in \mathbb{R}^n$ be such that $x_i = \sqrt{\tilde{\sigma}_i/\sigma_i}$ with the convention that $0/0 = 1$, and $y_i = \sqrt{\rho}_i$, for $i = 1, \ldots, n$. Suppose the k largest entries of $\hat{x}$ are $x_{i_1}, \ldots, x_{i_{k_1}}$ and $x_{j_1}^{-1}, \ldots, x_{j_{k_2}}^{-1}$, where $x_{i_a} > 1$ for all $1 \le a \le k_1$, $x_{j_b} \le 1$ for all $1 \le b \le k_2$, and $k_1 + k_2 = k$. Taking the square root of (2.8) in Corollary 2.4, we see that $\prod_{a=1}^{k_1} x_{i_a}$ is not larger than the product of the first k_1 terms of the sequence $\{\sigma_1(S)\sigma_1(T)\}_{j=1}^q$ and $\prod_{b=1}^{k_2} x_{j_b}^{-1}$ is not larger than the product of the reciprocal of the first k_2 terms of the sequence $\{\sigma_{m-j+1}(S)\sigma_{n-j+1}(T)\}_{j=1}^q$. By the construction of ρ_i's, we see that the hypothesis of Lemma 5.1 holds. Using Lemma 5.1 and the equality $\chi(\alpha, \beta) = \left|\sqrt{\alpha/\beta} - \sqrt{\beta/\alpha}\right|$, we get (5.9). □

Corollary 5.5 *Let A and $\tilde{A}$ be $m \times n$ with singular values $\sigma_1 \ge \cdots \ge \sigma_q$ and $\tilde{\sigma}_1 \ge \cdots \ge \tilde{\sigma}_q$, respectively, where $q = \min\{m, n\}$. Suppose that $\tilde{A} = SAT$ where S and T are invertible. Let $\|\cdot\|$ be and any symmetric norm on $\mathbb{R}^q$ and let $|||\cdot|||$ be the corresponding unitarily invariant on the matrix space of the appropriate dimension. Then,*

$$\begin{aligned} &\|(\chi(\sigma_1, \tilde{\sigma}_1), \ldots, \chi(\sigma_q, \tilde{\sigma}_q))\| \\ &\le \tfrac{1}{2}|||S^* - S^{-1}||| + \tfrac{1}{2}|||T^* - T^{-1}|||, \end{aligned} \tag{5.10}$$

and in particular

$$\left(\sum_{i=1}^{q} \chi^2(\sigma_i, \tilde{\sigma}_i)\right)^{1/2} \le \frac{1}{2}|||S^* - S^{-1}|||_{\mathrm{F}} + \frac{1}{2}|||T^* - T^{-1}|||_{\mathrm{F}}, \tag{5.11}$$

and

$$\max_{i=1,\ldots,q} \chi(\sigma_i, \tilde{\sigma}_i) \le \frac{1}{2}|||S^* - S^{-1}|||_2 + \frac{1}{2}|||T^* - T^{-1}|||_2. \tag{5.12}$$

Furthermore, for any positive p we have

$$\begin{aligned} \|(\chi(\sigma_1^p, \tilde{\sigma}_1^p), \ldots, \chi(\sigma_q^p, \tilde{\sigma}_q^p))\| \le \frac{1}{2}|||\ |S|^p - |S|^{-p}||| \\ + \frac{1}{2}|||\ |T|^p - |T|^{-p}|||. \end{aligned} \tag{5.13}$$

If T is unitary then

$$(5.14)\qquad ||\chi(\sigma_i,\tilde\sigma_i)|| \le ||(\chi(\sigma_i(S),1)|| = |||\ |S|^{1/2} - |S|^{-1/2}|||.$$

We define the matrix absolute value in the usual way:

$$|X| \equiv (X^*X)^{1/2}.$$

When it is applicable the inequality (5.14) is slightly stronger than (5.11). One may wonder whether, when both S and T are merely nonsingular, it is true that

$$(5.15)\qquad ||\chi(\sigma_i,\tilde\sigma_i)|| \le |||\ |S|^{1/2} - |S|^{-1/2}||| + |||\ |T|^{1/2} - |T|^{-1/2}|||.$$

To see that it is not true one need look no further than the scalar example, $A = 1$, $S = T = 2$. This is just a manifestation of the fact that $\chi(rs,1) \le \chi(r,1) + \chi(s,1)$ if and only if $(r-1)(s-1) \le 0$. It is true that

$$||\chi(\sigma_i,\tilde\sigma_i)|| \le \frac{|||\ |S|^{1/2} - |S|^{-1/2}||| + |||\ |T|^{1/2} - |T|^{-1/2}|||}{1 - \frac{1}{8}|||\ |S|^{1/2} - |S|^{-1/2}|||_2\ |||\ |T|^{1/2} - |T|^{-1/2}|||_2}$$

provided that the denominator is positive. Since $\chi(r,1) \le \frac{1}{2}\chi(r^2,1)$, we have

$$(5.16)\qquad |||\ |S|^{1/2} - |S|^{-1/2}\ ||| \le \frac{1}{2}|||\ S^* - S^{-1}\ |||$$

so this bound is slightly stronger than R.-C. Li's bound [12, Theorem 4.1]. Alternatively, one could apply (5.13) with $p = 1/2$ and obtain the valid bound

$$||\chi(\sqrt{\sigma}_i,\sqrt{\tilde\sigma_i})|| \le |||\ |S|^{1/2} - |S|^{-1/2}||| + |||\ |T|^{1/2} - |T|^{-1/2}|||.$$

The bound (5.13) is perhaps a little hard to interpret because of the matrix absolute value. However, it is useful. We've seen one application with $p = 1/2$ in the previous display; the bound (5.10) is just (5.13) with $p = 1$; and in the context of the generalized eigenvalue problem we will see an application with $p = 2$.

Proof of Corollary 5.5. First we deduce (5.10) from (5.9). We know that $\rho_i = \sigma_{j_i}(S)\sigma_{k_i}(T)$, for each $i = 1,\ldots,q$, and that the indices $j_1,\ldots,j_q$ are distinct as are the indices $k_1,\ldots,k_q$. That is, each singular value of S and each singular value of T occurs at most once in the definition of the ρ_i's.

By Lemma 5.2 we have

$$\chi(\rho_i,1) \le \frac{1}{2}\left\{|\sigma_{j_i}(S) - \sigma_{j_i}^{-1}(S)| + |\sigma_{k_i}(T) - \sigma_{k_i}^{-1}(T)|\right\}$$

Using this entry-wise inequality for the first inequality below, the triangle inequality for the second, and Lemma 3.2 for the third, we have

$$\begin{aligned}
&\| \, (\chi(\rho_i, 1))_{i=1}^q \| \\
&\le \| \frac{1}{2} \left\{ |\sigma_{j_i}(S) - \sigma_{j_i}^{-1}(S)| + |\sigma_{k_i}(T) - \sigma_{k_i}^{-1}(T)| \right\})_{i=1}^q \| \\
&\le \frac{1}{2} \left\{ \|(\sigma_{j_i}(S) - \sigma_{j_i}^{-1}(S))_{i=1}^q\| + \|(\sigma_{k_i}(T) - \sigma_{k_i}^{-1}(T))_{i=1}^q\| \right\} \\
&\le \frac{1}{2} \left\{ \|(\sigma_i(S) - \sigma_i^{-1}(S))_{i=1}^m\| + \|(\sigma_i(T) - \sigma_i^{-1}(T))_{i=1}^n\| \right\} \\
&= \frac{1}{2} \left\{ |||\, S^* - S^{-1} \,||| + |||\, T^* - T^{-1} \,||| \right\}.
\end{aligned}$$

To prove (5.13) take the $p/2$ power of (2.8), instead of the square root as we did in Proposition 5.4, and prove the corresponding generalization of Proposition 5.4. Then use the proof of (5.10) above to deduce (5.13). □

We may apply the majorization result Lemma 5.1 to the singular value bound (2.8), without taking square roots to obtain:

Proposition 5.6 *Let $A = C^*HC$ and $\tilde{A} = C^*\tilde{H}C$ be $n \times n$ positive definite matrices with eigenvalues λ_i, $\tilde{\lambda}_i$. Let $E = H^{-1/2}\Delta H\, H^{-1/2}$. Then for any symmetric norm $\| \cdot \|$ on $\mathbb{R}^n$ and the corresponding unitarily invariant norm $||| \cdot |||$*

$$\|(\chi(\lambda_1, \tilde{\lambda}_1), \ldots, \chi(\lambda_n, \tilde{\lambda}_n))\| \le |||E(I+E)^{-1/2}||| \tag{5.17}$$

and in particular

$$\sqrt{\sum_{i=1}^n \chi^2(\lambda_i, \tilde{\lambda}_i)} \le |||E(I+E)^{-1/2}|||_{\mathrm{F}} \tag{5.18}$$

and

$$\max_{i=1,\ldots,n} \chi(\lambda_i, \tilde{\lambda}_i) \le |||E(I+E)^{-1/2}|||_2. \tag{5.19}$$

Now let us consider multiplicative perturbation bounds for the generalized eigenvalue and singular value problems. The conditions here are the same as those in [12] and are somewhat more restrictive than those for the standard eigenvalue and singular value problems.

Proposition 5.7 *Consider the generalized eigenvalue problems*

$$A_1 - \lambda A_2 \equiv S_1^* H_1 S_1^* - \lambda S_2^* H_2 S_2^*, \quad \text{and } \tilde{A}_1 - \lambda \tilde{A}_2 \equiv S_1^* \tilde{H}_1 S_1^* - \lambda S_2^* \tilde{H}_2 S_2^*$$

where the H's and $\tilde{H}$'s and at least one of the S's is invertible. Let $||\cdot||$ be any symmetric norm on $\mathbb{R}^n$ and let $|||\cdot|||$ be the corresponding unitarily invariant norm. Then

$$\|(\chi(\lambda_i,\tilde{\lambda}_i))_{i=1}^n|||\le \frac{1}{2}\left[|||(I+E_1)-(I+E_1)^{-1}||| + |||(I+E_2)-(I+E_2)^{-1}|||\right] \tag{5.20}$$

and

$$\|(\chi(\sqrt{\lambda_i},\sqrt{\tilde{\lambda}_i}))_{i=1}^n\|\le |||E_1(I+E_1)^{-1/2}||| + |||E_2(I+E_2)^{-1/2}|||. \tag{5.21}$$

Furthermore

$$\|(\chi(\lambda_i,\tilde{\lambda}_i))_{i=1}^n|||\le \frac{|||E_1(I+E_1)^{-1/2}|||+|||E_2(I+E_2)^{-1/2}|||}{1-\frac{1}{8}|||E_1(I+E_1)^{-1/2}|||_2\;|||E_2(I+E_2)^{-1/2}|||_2}, \tag{5.22}$$

provided that the denominator is positive.

The last bound (5.22) is the the generalization of [12, Theorem 7.1] to all unitarily invariant norms. We could of course have stated the special cases of (5.22) corresponding to the Frobenius norm and the 2-norm explicitly. Note that R.-C. Li weakened the result slightly by replacing $|||E_i|||$ by

$$\frac{|||H_i^{-1}|||_2|||\Delta H_i|||}{\sqrt{1-|||H_i^{-1}|||_2|||\Delta H_i|||_2}},$$

so as to avoid the use of E_i in the final bound. The first two bounds are somewhat cleaner than the last.

Neither of (5.20) and (5.22) is always stronger than the other. If $E_1=0$ and $E_2\neq 0$ then, by (5.16), (5.22) is stronger. On the other hand, in the limit as the denominator approaches 0 in (5.22), the bound (5.20) is stronger.

Proof. Without loss of generality S_2 is invertible.[2] Following R.-C. Li [12, Section 7], the eigenvalues of the unperturbed pencil are the same as the eigenvalues of the matrix

$$H_2^{-1/2}S_2^{-*}S_1^*H_1S_1S_2^{-1}H_2^{-1/2}$$

[2] If S_2 is singular then S_1 must be invertible. Let μ_i and $\tilde{\mu}_i$ be the eigenvalues of the pencils $A_2-\lambda A_1$. $\tilde{A}_2-\lambda\tilde{A}_1$. Then, if we define $\chi(\infty,\infty)=0$, $\chi(\mu_i,\tilde{\mu}_i)=\chi(\lambda_i,\tilde{\lambda}_i)$, and so the left hand side of (5.20) is unchanged by reversing the roles of A_1 and A_2. One can check that the right hand side is not changed either, and so it is sufficient to prove the result with A_1 and A_2 reversed.

which are the same as the squares of the singular values of the matrix

$$B = H_2^{-1/2} S_2^{-*} S_1^* H_1^{1/2}.$$

In the same way the eigenvalues of the perturbed pencil are the squares of the singular values of the matrix

$$\tilde{B} = \tilde{H}_2^{-1/2} S_2^{-*} S_1^* \tilde{H}_1^{1/2} = SBT$$

where

$$S = \tilde{H}_2^{-1/2} H_2^{1/2}, \text{ and } T = H_1^{-1/2} \tilde{H}_1^{1/2}.$$

Notice that $|T|^2 = T^*T = I + E_1$ and that $|S|^2 = (S^*S)^{-1} = I + E_2$. Now apply (5.13) with $p = 2$ we get

$$\begin{aligned} &\|(\chi(\lambda_i, \tilde{\lambda}_i)_{i=1}^n)\| \\ &\quad = \|(\chi(\sigma_i^2, \tilde{\sigma}_i^2)_{i=1}^n)\| \\ &\quad \le \frac{1}{2}(|||\,|T|^2 - |T|^{-2}||| + |||\,|S|^2 - |S|^{-2}|||) \\ &\quad = \frac{1}{2}\left[|||(I + E_1) - (I + E_1)^{-1}||| + |||(I + E_2) - (I + E_2)^{-1}|||\right] \end{aligned}$$

which is (5.20). Applying (5.13) with $p = 1/2$ and manipulating the resulting expression gives (5.21).

To deduce (5.22) one combines the scalar inequality (5.4) with the majorization

(5.23) $$(\chi(\sigma_i^2, \tilde{\sigma}_i^2))_{i=1}^n \prec_w (\chi(\sigma_i^2(S)\sigma_i^2(T), 1))_{i=1}^n.$$

In particular, from (5.4) we have

$$\begin{aligned} \chi(\sigma_i^2(S)\sigma_i^2(T), 1) &\le \frac{\chi(\sigma_i^2(S), 1) + \chi(\sigma_i^2(S), 1)}{1 - (1/8) \cdot \chi(\sigma_i^2(S), 1)\chi(\sigma_i^2(S), 1)} \\ &= \frac{|\sigma_i(S) - \sigma_i^{-1}(S)| + |\sigma_i(T) - \sigma_i^{-1}(T)|}{1 - (1/8) \cdot |\sigma_i(S) - \sigma_i^{-1}(S)||\sigma_i(T) - \sigma_i^{-1}(T)|} \\ &\le \frac{|\sigma_i(S) - \sigma_i^{-1}(S)| + |\sigma_i(T) - \sigma_i^{-1}(T)|}{1 - (1/8) \cdot |||S - S^{-1}|||_2|||T - T^{-1}|||_2} \\ &\equiv \gamma(|\sigma_i(S) - \sigma_i^{-1}(S)| + |\sigma_i(T) - \sigma_i^{-1}(T)|) \end{aligned}$$

Now using (5.23) for the first inequality and the component-wise bound on $\chi(\sigma_i^2(S)\sigma_i^2(T), 1)$ that we have just derived for the second inequality we have

$$\begin{aligned} &\|(\chi(\lambda_i, \tilde{\lambda}_i))_{i=1}^n||| \\ &\quad = \|(\chi(\sigma_i^2, \tilde{\sigma}_i^2))_{i=1}^n|| \\ &\quad \le \|(\chi(\sigma_i^2(S)\sigma_i^2(T), 1))_{i=1}^n|| \end{aligned}$$

$$\begin{aligned}
&\leq \gamma \| \, (|\sigma_i(S) - \sigma_i^{-1}(S)| + |\sigma_i(T) - \sigma_i^{-1}(T)|)_{i=1}^n || \\
&\leq \gamma (|| \, (|\sigma_i(S) - \sigma_i^{-1}(S)|)_{i=1}^n |||| \, (\sigma_i(T) - \sigma_i^{-1}(T)|)_{i=1}^n ||) \\
&= \gamma (|||S^* - S^{-1}||| + |||T^* - T^{-1}|||) \\
&= \gamma (||| \, |S| - |S|^{-1}||| + ||| \, |T| - |T|^{-1}|||).
\end{aligned}$$

Now note that $|T| = (T^*T)^{1/2} = (I + E_1)^{1/2}$. Substituting for $|S|$ and $|T|$ and manipulating the resulting expression yields the desired bound (5.22). □

Now let us consider the generalized singular value problem.

Proposition 5.8 *Consider the pairs of $n \times n$ matrices $(B_1, B_2) \equiv (G_1S_1, G_2S_2)$ and $(\tilde{B}_1, \tilde{B}_2) \equiv (\tilde{G}_1S_1, \tilde{G}_2S_2)$ where the G's and $\tilde{G}$'s and at least one of the S's is nonsingular. Let the generalized singular values be σ_i and $\tilde{\sigma}_i$, $i = 1, \ldots, n$. Let $\| \cdot \|$ be any symmetric norm on $\mathbb{R}^n$ and let $||| \cdot |||$ be the corresponding unitarily invariant norm. Let $E_i = \Delta G_i(G_i)^{-1}$. Then*

$$\begin{aligned}
\|(\chi(\sigma_1, \tilde{\sigma}_1), \ldots, \chi(\sigma_q, \tilde{\sigma}_q))\| \leq \frac{1}{2}|||(I + E_1)^* - (I + E_1)^{-1}||| \\
+\frac{1}{2}|||(I + E_2)^* - (I + E_2)^{-1}|||,
\end{aligned} \tag{5.24}$$

and in particular

$$\begin{aligned}
\left(\sum_{i=1}^q \chi^2(\sigma_i, \tilde{\sigma}_i)\right)^{1/2} \leq \frac{1}{2}|||(I + E_1)^* - (I + E_1)^{-1}|||_{\mathrm{F}} \\
+\frac{1}{2}|||(I + E_2)^* - (I + E_2)^{-1}|||_{\mathrm{F}},
\end{aligned} \tag{5.25}$$

and

$$\begin{aligned}
\max_{i=1,\ldots,q} \chi(\sigma_i, \tilde{\sigma}_i) \leq \frac{1}{2}|||(I + E_1)^* - (I + E_1)^{-1}|||_2 \\
+\frac{1}{2}|||(I + E_2)^* - (I + E_2)^{-1}|||_2.
\end{aligned} \tag{5.26}$$

In the case that either E_1 or E_2 is 0 one can strengthen the result (5.24) just as (5.14) is a strengthening of (5.10), the only problem is that the resulting right hand side rather cumbersome.

Proof. As in the previous result, without loss of generality S_2 is invertible. The the generalized singular values of the first pair are just the singular values of $B = G_1S_1S_2^{-1}G_2^{-1}$ and those of the second pair are the singular values of $\tilde{B} = (I + E_1)B(I + E_2)^{-1}$. The inequalities now follow from Corollary 5.5. □

One can also apply these ideas to the relative perturbation of eigenvalues of matrices that are known to have positive eigenvalues.

Proposition 5.9 *Let* $A = X \operatorname{diag}(\lambda_1, \ldots, \lambda_n) X^{-1}$ *and* $\tilde{A} = \tilde{X}\operatorname{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n)\tilde{X}^{-1}$*. Suppose that* $\tilde{A} = DA$*. Then*

$$(\chi(\lambda_1, \tilde{\lambda}_1), \ldots, \chi(\lambda_n, \tilde{\lambda}_n)) \prec_w \kappa(X^{-1}\tilde{X})(\chi(\sigma_1, 1), \ldots, \chi(\sigma_n, 1)),$$

where $\sigma_i = \sigma_i(X^{-1}DX)$*. Consequently for any symmetric norm*

$$\|(\chi(\lambda_1, \tilde{\lambda}_1), \ldots, \chi(\lambda_n, \tilde{\lambda}_n))\| \leq \kappa(X^{-1}\tilde{X})\|(\chi(\sigma_1, 1), \ldots, \chi(\sigma_n, 1))\|.$$

Proof. Let $B = X^{-1}DX$. Then

$$\tilde{\Lambda} = (X^{-1}\tilde{X})^{-1}B\Lambda(X^{-1}\tilde{X}).$$

Since both $\tilde{\Lambda}$ and Λ are positive definite their eigenvalues are the same as their singular values. Thus, from (2.4) we have

$$\prod_{i=1}^{k} \kappa^{-1}(X^{-1}\tilde{X})\sigma_{n+1-i} \leq \prod_{j=1}^{k} \frac{\tilde{\lambda}_{i_j}}{\lambda_{i_j}} \leq \prod_{i=1}^{k} \kappa(X^{-1}\tilde{X})\sigma_i. \tag{5.27}$$

One can derive the majorization result from this using a technique very similar to Lemma 5.1. □

We have not provided all the details in the proof of Proposition 5.9 (nor have we stated the result in its most general form, i.e., $\tilde{A} = D_1AD_2$). Our main point is that it is possible, using our methods, to prove a Lidskii-Wielandt type bound for diagonalizable matrices with real eigenvalues. Whether our bound (5.27) is good depends on whether $\kappa(X\tilde{X}^{-1})$ is small (as it will be in the limit as $\|D - I\| \to 0$). The norm on the right hand side of (5.27) is bounded, at least approximately, by

$$\kappa(X^{-1}\tilde{X})\|(\chi(\sigma_1(D - I), 1), \ldots, \chi(\sigma_n(D - I), 1)\|$$

since $X^{-1}DX = I + X^{-1}(D - I)X$.

6. Comparison with other research

First we compare our results with those in the literature. Then we compare our techniques and our approach with those in the literature and argue that our purely multiplicative result (Corollary 2.7) is the fundamental result and although it uses less familiar notation its use simplifies and strengthens the results based on it. Finally we compare the different measures of relative error.

6.1. Comparison of results

There are numerous papers with relative error bounds for eigenvalues and singular values – see [10]. The results due to R.-C. Li in [12] are the most general to date so, for the most part, we will confine our comparison to them.

First, one is usually interested in perturbation bounds when the perturbation is fairly small. In that case little is lost in taking first order approximations. One can check that in any given context all the applicable bounds in this paper (both those in terms of χ and those in terms of $|\log(\alpha/\beta)|$) as well as R.-C. Li's in [12] are the same to first order. Thus the differences in the results are mainly a matter of the generality of the bound and the simplicity of the statement and the proof of the bound.

Our perturbation bounds are considerably more general being for all symmetric norms while those in [12] are only for the two norms

$$(6.1)\quad l_\infty(x) = \max\{|x_1|, \ldots, |x_n|\}, \quad \text{and} \quad l_2(x) = \left\{\sum_{i=1}^n x_i^2\right\}^{1/2}.$$

That is, our results are multiplicative analogs of the more general Lidskii-Wielandt perturbation bound while those in [12] are multiplicative analogs of the Lidskii-Weyl (for l_∞) and Wielandt-Hoffman (for l_2) inequalities. The techniques in [12] do however extend to normal and (even diagonalizable) matrices and to analyzing the perturbation of invariant subspaces.

In the following, we illustrate how our results in Sect. 3, expressed in terms of χ improve those in [12] in various cases. The technical report [14], which is an earlier version of [12], contains more results and slightly stronger results that those in [12], but our results are also improvements over the corresponding results in [14].

First, when $A = A^*$ is positive definite, the results in [12, Theorem 3.1, part 2] are exactly the same as our Proposition 5.3 restricted to the two norms in (6.1).

Second, when $A = A^*$ is indefinite, [14, Equation (7.4)] is

$$\max_{i=1,\ldots,n} \chi(\lambda_i, \tilde{\lambda}_i) \le |||S^{-1}|||_2 |||I - SS^*|||_2$$

while our result (5.7) is

$$\max_{i=1,\ldots,n} \chi(\lambda_i, \tilde{\lambda}_i) \le |||S^{-1} - S^*|||_2,$$

which is stronger because

$$|||S^{-1} - S^*|||_2 = |||S^{-1}(I - SS^*)|||_2 \le |||S^{-1}|||_2 |||I - SS^*|||_2.$$

Third, the Wielandt-Hoffman type bound for possibly indefinite Hermitian matrices in [12, Theorem 3.1, part 1] states that if A is Hermitian and $\tilde{A} = S^*AS$ there is a permutation τ such that

$$\sum_{i=1}^{n} \frac{(\lambda_i - \tilde{\lambda}_{\tau(i)})^2}{\lambda_i^2 + \tilde{\lambda}_{\tau(i)}^2} \;\leq\; \sum_{i=1}^{n}(1-\sigma_i(S))^2 + \sum_{i=1}^{n}(1-\sigma_i(S)^{-1})^2.$$

This inequality, and the additional fact that the permutation τ can be taken to be the identity follows from our result:

$$\begin{aligned}\sum_{i=1}^{n} \frac{(\lambda_i - \tilde{\lambda}_i)^2}{\lambda_i^2 + \tilde{\lambda}_i^2} &\leq \sum_{i=1}^{n} \frac{(\lambda_i - \tilde{\lambda}_i)^2}{2\lambda_i\tilde{\lambda}_i} \\ &\leq \sum_{i=1}^{n} \frac{(\sigma_i(S) - \sigma_i(S)^{-1})^2}{2} \\ &\leq \sum_{i=1}^{n}(1-\sigma_i(S))^2 + \sum_{i=1}^{n}(1-\sigma_i(S)^{-1})^2.\end{aligned}$$

The first and third inequalities are easily verified because the individual summand satisfy the same inequalities. The second inequality is the square of our bound (5.8) in the case of the norm l_2.

Note that in general, our bounds and proofs are the same in both the positive definite and the general Hermitian case, whereas in [12] the results in the indefinite case are weaker. Note also, that our bound (5.8) is stronger than the Wielandt Hoffman bound [13, Corollary 2.1] even though the latter contains an additional condition which may be hard to verify.

Using a slightly different notion of relative error R.-C. Li (see [10, Corollary 5.1]) states that there is a permutation τ such that

$$\sqrt{\sum_{i=1}^{n}\left(\frac{|\lambda_i - \tilde{\lambda}_{\tau(i)}|}{\lambda_i}\right)^2} \leq |||S|||_2 \cdot |||S^* - S^{-1}|||_{\mathrm{F}}.$$

Our (5.8) implies this result and the fact that τ can be taken to be the identity. We discuss where this unknown permutation comes from in the next subsection.

Fourth, if $\tilde{A} = SAT$ the result in [12, Theorem 4.1] asserts that

$$\|(\chi(\sigma_i(A), \sigma_i(\tilde{A}))_{1\leq i\leq n}\| \leq \frac{\gamma}{2}\left\{|||S^* - S^{-1}||| + |||T^* - T^{-1}|||\right\},$$

with

$$\gamma = \frac{32}{32 - |||S^* - S^{-1}|||_2|||T^* - T^{-1}|||_2},$$

if $\|\cdot\|$ is l_2 or l_∞. By Corollary 5.5, we have the inequality for all symmetric norms with $\gamma = 1$. In the next subsection we explain why the factor γ arises in R.-C. Li's bounds. Using similar methods as when comparing eigenvalues one can show that our singular value bounds are stronger and more general than those in [12, Theorem 4.2]. We omit the details for the sake of brevity.

Fifth, Proposition 5.6, on graded matrices, is a generalization of [12, Theorem 3.2, bounds (3.9) and (3.11)] to all symmetric norms. We can also generalize [12, Theorem 5.5] to all symmetric norms in exactly the same way.

Sixth, our bounds (5.20) and (5.21) for the generalized eigenvalue problem are valid for a larger range of perturbations than (5.22), which is the type of bound obtained in [12, Theorem 7.1].

Finally, our bound for the generalized singular value problem is stronger than [12, Theorem 7.3] by the same factor γ above.

In conclusion, we have generalized all the perturbation bounds in [12] involving either the (possibly generalized) eigenvalues of Hermitian matrices or the (possibly generalized) singular values of general matrices to all symmetric norms. Our approach is to first prove a multiplicative majorization like those in Sect. 2 and then use the majorization result Lemma 5.1 to deduce a bound on the norm of the vector of relative perturbations.

6.2. Comparison of techniques

Now we look at the difference between our proof techniques and existing techniques, and we argue that (2.11) is the correct generalization of additive perturbation bound (1.2).

The proofs in [12] and rest of the literature are almost entirely at the matrix level–none use majorization. Our proofs move very quickly from matrices to scalars, that is, to singular values and eigenvalues.

Our approach to proving the fundamental inequalities

$$\sum_{j=1}^{k}[\lambda_{i_j}(A+E)-\lambda_{i_j}(A)] \le \sum_{j=1}^{k}\lambda_j(E) \tag{6.2}$$

and its multiplicative analog (2.11) is to deduce them from the much simpler monotonicity properties of eigenvalues of Hermitian matrices expressed in Weyl's and Ostrowskii's Inequalities. In this sense our approach is somewhat similar to most[3] of the other work in this area that is based on min-max characterizations or monotonicity properties of eigenvalues of Hermitian matrices. R.-C. Li's work [12] is different in that it uses a perturbation equation

[3] Eisenstat and Ipsen use both monotonicity principles and a perturbation equation in [5], and using R.-C. Li's technique they do derive a multiplicativeWielandt-Hoffman type bound for diagonalizable matrices [6, Section 6].

and then the fact that the permutation matrices are the extreme point of the doubly stochastic as in the standard proof of the Wielandt-Hoffman theorem. By using this approach R.-C. Li was able to obtain simultaneous bounds on all the eigenvalues and thus give the first relative Wielandt-Hoffman type result. The weakness of this approach is that it just states that "there is a matching of the λ's and the $\tilde{\lambda}$'s" such that the norm of the differences is small. It does not specify the matching. In the positive definite case one can use a separate argument to show that the best matching is match λ_i with $\tilde{\lambda}_i$ $i = 1, \ldots, n$, this is not the case in the indefinite case [12, Propositions 2.3 and 2.4 and Remarks 2.1, 2.2 and 2.3]. Our approach shows that in the indefinite case, even though the identity permutation may not give the smallest norm of the differences, the norm of the differences is still smaller than the right hand side. That is, even though the identity permutation may not be the best it is good enough.

Let us see why (2.11) is the right analog of (1.2). If we take logarithms of (2.11) and state the result as a weak majorization we have:

$$(6.3) \qquad (|\log \lambda_i(S^* AS) - \log \lambda_i(A)|)_{i=1}^n \prec_w (|\log \lambda_i(S^* S)|)_{i=1}^n$$

while we may state (1.2) as the weak majorization

$$(6.4) \qquad (|\lambda_i(A) - \lambda_i(A+E)|)_{i=1}^n \prec_w (|\lambda_i(E)|)_{i=1}^n .$$

The analogy is immediate–thus we have a multiplicative analog of the fundamental additive perturbation bound for eigenvalues of Hermitian matrices.

One may wonder why we feel that (2.11) is the fundamental result rather than the weak majorization (5.6) or the equivalent result statement in terms of norms (5.7). The reason is that (2.11) is stronger than (5.6) and (5.7). One cannot deduce (2.11) from (5.6) and (5.7). Another way to think of this is that in Lemma 5.1 (5.1) implies (5.2) but not conversely.

This objection to the relative distance χ does not apply to the relative distance rd. One *can* derive (2.11) (or its logarithmic version (6.3)) from the weak majorization (4.1) which is expressed in terms of rd. To see this, observe that since multiplying S by $t > 0$ merely adds $2 \log t$ to each component of the left hand side and right hand side of (6.3), and so it is sufficient to prove (6.3) in the case where $\sigma_n(S) \geq 1$. However, in this case the components of the vectors on either side of (4.1) are nonnegative even without the absolute value sign. That is, if $\sigma_n(S) \geq 1$ then (4.1) *is* (6.3).

Notice also that we do not derive Proposition 5.4 and Corollary 5.5, the norm-wise form of the relative singular value bound from Proposition 5.3, the norm-wise form of the relative eigenvalue bound, rather we combine the singular value bound (2.8) from Sect. 2, and the key Lemma 5.1 that states that log majorization implies majorization in terms of the χ's. Another way to see that the fundamental result for the relative perturbation of singular values is (2.8) and not Proposition 5.4 is that if we raise the σ's and ρ's in the majorization Proposition 5.4 to any positive power p we get the statement

$$(\chi(\sigma_1^p, \tilde{\sigma}_1^p), \ldots, \chi(\sigma_n^p, \tilde{\sigma}_n^p)) \prec_w (\chi(\rho_1^p, 1), \ldots, \chi(\rho_n^p, 1)). \tag{6.5}$$

This is indeed a valid inequality, but it cannot be deduced from Proposition 5.4 , which is the special case when $p = 1$. It does however follow easily from the singular value bound (2.8) in Sect. 2–just take the pth power of (2.8) and then apply Lemma 5.1 as in the proof of Proposition 5.4. The inequality (6.5) with p other than 1 is not just a curiosity, we we have used (6.5) with $p = 1/2$ and $p = 2$.

Yet another reason for preferring (2.11) is that it deals very easily with the case when we have two (or more) multiplicative perturbations. In this situation the natural way to proceed would be to bound the effect of one perturbation and then the effect of other and then add the two bounds, and this is what we do in the proof of Corollary 2.4. However, once we have a bound in terms of χ it is much harder to deal with two multiplicative perturbations since this approach does not work because it is not true that $\chi(a, c) \leq \chi(a, b) + \chi(b, c)$ for all $a, b, c > 0$. One has then to resort to using the more cumbersome inequality

$$\chi(a, c) \leq \frac{\chi(a, b) + \chi(b, c)}{1 - \frac{1}{8}\chi(a, b)\chi(b, c)} \tag{6.6}$$

which is valid whenever the denominator is positive [12, Lemma 6.1]. This is why R.-C. Li's relative Wielandt-Hoffman bound has an extra factor

$$\gamma = \frac{32}{32 - |||S^* - S^{-1}|||_2 |||T^* - T^{-1}|||_2}.$$

Notice that this factor is 1 if, and only if, one of S or T is unitary. The reason is that in this case the one that is unitary does not change the singular values of SAT and so we need only bound the effect of the other. It is no longer a two step process and so there is no need to use the bound (6.6).

6.3. Which relative distance?

A number of different measures of the relative difference between two numbers α and β of the same sign have been proposed in the context of multiplicative perturbation bounds for eigenvalues and singular values. We shall not attempt a survey here, but will argue that

$$\mathrm{rd}(\alpha, \beta) \equiv |\log \alpha/\beta| = |\log|\alpha| - \log|\beta||,$$

even though it is perhaps less intuitive that some of the others, has the best mathematical properties in this context:

1. rd is a metric, and so two step proofs are easier (compare the proofs of Corollary 5.5 and Proposition 4.2)
2. rd is symmetric in its arguments.
3. rd has a connection with the standard (additive) absolute value.
4. As explained in the previous subsection, the weak majorization (4.1) in terms of rd implies the fundamental bound (5.3). This is not the case for the weak majorization (5.6) which is expressed in terms of χ.
5. Taking powers of the basic inequality (2.11) before applying majorization techniques doesn't give new inequalities, thus there is no question of which power to use.

In Subsect. 6.2 we argued that one should not express multiplicative perturbation bounds in terms of norms. However, if one has to use norms, perhaps it would be best to use rd as the measure of relative error and use the norm bounds in Sect. 4.

7. Remarks and related inequalities

As mentioned in Sect. 2, typically one derives the Lidskii-Mirsky-Wielandt bound from a result of Wielandt. In this section, we give a proof for the result of Wielandt using our technique, and show that some other well-known matrix inequalities also follow readily from our proof. Furthermore, we give examples showing that it is impossible to get a multiplicative analog of Wielandt to prove Theorem 5.3.

We shall use $\{e_1, \ldots, e_n\}$ to denote the standard basis of $\mathbb{C}^n$ in our discussion.

Theorem 7.1 [19, Chapter IV, Theorem 4.5] *Let A be an $n \times n$ Hermitian matrix. For any indices $1 \le i_1 < i_2 < \cdots < i_k \le n$,*

$$\sum_{j=1}^{k} \lambda_{i_j}(A) = \max_{\substack{W_1 \subset W_2 \subset \cdots \subset W_k \\ \dim W_j = i_j}} \; \min_{\substack{Y \in \mathbb{C}^{n\times k} \\ Y e_j \in W_j,\ Y^*Y = I_k}} \mathrm{tr}(Y^*AY), \tag{7.1}$$

where $W_1 \subset \cdots \subset W_k$ are subspaces of $\mathbb{C}^n$.

Proof. Suppose A has spectral decomposition $\sum_{j=1}^{n} \lambda_j(A) x_j x_j^*$. If $W_j = \mathrm{span}\{x_1, \ldots, x_j\}$ for $j = 1, \ldots, i_k$, then

$$\sum_{j=1}^{k} \lambda_{i_j}(A) = \min_{\substack{Y e_j \in W_j \\ Y^*Y = I_k}} \mathrm{tr}(Y^*AY).$$

Thus one can focus on proving that the left side is not less than the right side of (7.1). To this end, let $W_1 \subset \cdots \subset W_k$ be subspaces of $\mathbb{C}^n$. We shall show that there exists $Y \in \mathbb{C}^{n\times k}$ satisfying $Ye_j \in W_j$ and $Y^*Y = I_k$ such that $\sum_{j=1}^k \lambda_{i_j}(A) \geq \mathrm{tr}\,(Y^*AY)$.

We prove by induction on n. The result is trivial when $n = 1$. Suppose $n \geq 2$ and we know that the result is true for Hermitian matrices of order $(n-1)$. We consider three cases:

(a) $k = 1$;

(b) $i_j = j$ for $j = 1, \ldots, k$;

(c) none of (a) or (b) hold.

In all cases, we may replace A by

$$\tilde{A} = \sum_{j<i_k} \lambda_j(A) x_j x_j^* + \lambda_{i_k}(A) \sum_{j\geq i_k} x_j x_j^*.$$

If we can find $Y \in \mathbb{C}^{n\times k}$ satisfying $Ye_j \in W_j$ and $Y^*Y = I_k$ such that $\sum_{j=1}^k \lambda_{i_j}(\tilde{A}) \geq \mathrm{tr}\,(Y^*\tilde{A}Y)$, then the result will follow from the facts that $\sum_{j=1}^k \lambda_{i_j}(A) = \sum_{j=1}^k \lambda_{i_j}(\tilde{A})$ and $\mathrm{tr}\,(Y^*\tilde{A}Y) \geq \mathrm{tr}\,(Y^*AY)$.

For case (a), let W_1 have dimension $i_1 = p$. Since the null space of $\tilde{A} - \lambda_p(A)I$ has dimension $n - p + 1$, there exists a unit vector $y \in W_1$ such that $\tilde{A}y = \lambda_p(A)y$, and hence $\lambda_p(\tilde{A}) = y^*\tilde{A}y$.

For case (b), let $Y \in \mathbb{C}^{n\times k}$ satisfy $Ye_j \in W_j$ and $Y^*Y = I_k$. Then there exists an $n \times (n-k)$ matrix Z such that $[Y \mid Z]$ is unitary. Since $\tilde{A} \geq \lambda_k(A)I$, we have

$$\sum_{j=1}^k \lambda_j(\tilde{A}) = \mathrm{tr}\,\tilde{A} - \mathrm{tr}\,(Z^*\lambda_k(A)\,Z) \geq \mathrm{tr}\,\tilde{A} - \mathrm{tr}\,(Z^*\tilde{A}Z) = \mathrm{tr}\,(Y^*\tilde{A}Y).$$

For case (c), we may assume that $i_k = n$. Otherwise, we can add the terms $i_k + 1, i_k + 2, \ldots, n = i_{\tilde{k}}$ to the sequence, and enlarge W_k to get $W_{k+1}, W_{k+2}, \ldots$, until we get $\mathbb{C}^n = W_{\tilde{k}}$. Suppose the modified problem is solved, i.e., we can find $Y \in \mathbb{C}^{n\times\tilde{k}}$ satisfying $Ye_j \in W_j$ and $Y^*Y = I_{\tilde{k}}$ such that $\sum_{j=1}^{\tilde{k}} \lambda_{i_j}(\tilde{A}) \geq \mathrm{tr}\,(Y^*\tilde{A}Y)$. One can write $Y = [Y_1 \mid Y_2]$ so that $Y_1 \in \mathbb{C}^{n\times k}$. Using the fact that $\tilde{A} \geq \lambda_{i_k}(A)I$, we have

$$\begin{aligned}\sum_{j=1}^k \lambda_{i_j}(\tilde{A}) &= \sum_{j=1}^{\tilde{k}} \lambda_{i_j}(\tilde{A}) - \mathrm{tr}\,(Y_2^*\lambda_{i_k}(A)IY_2)\\ &\geq \mathrm{tr}\,(Y^*AY) - \mathrm{tr}\,(Y_2^*\tilde{A}Y_2) = \mathrm{tr}\,(Y_1^*\tilde{A}Y_1).\end{aligned}$$

(The rest of the proof uses the idea of Wielandt and the results of the special cases we developed. We present it for the sake of completeness.)

Now suppose $i_k = n$. Since case (a) does not hold, there exists a largest integer ℓ such that $i_\ell + 1 < i_{\ell+1}$. Let $i_\ell = p$, $i_{\ell+1} = q$, and let W be an

$(n-1)$ dimensional subspace in $\mathbb{C}^n$ containing W_{i_ℓ} and the eigenvectors x_j, $j = p+2, \ldots, n$. Define $\tilde{W}_j = W_j \cap W$ for $j = 1, \ldots, k$. Let B be the compression of the matrix A on W, i.e., $B = Z^*AZ$ for some $Z \in \mathbb{C}^{n\times(n-1)}$ with $Z^*Z = I_{n-1}$ and the columns of Z span W. By construction, the $n-p-1$ smallest eigenvalues of A are the same as that of B. By induction assumption, we can find $\tilde{Y} \in \mathbb{C}^{(n-1)\times k}$ such that $\tilde{Y}e_j \in \tilde{W}_j$, $\tilde{Y}^*\tilde{Y} = I_k$, and

$$\sum_{j=1}^{\ell} \lambda_{i_j}(B) + \sum_{j=q-1}^{n-1} \lambda_{i_j}(B) \geq \operatorname{tr}(\tilde{Y}^*B\tilde{Y}).$$

Let $Y = Z\tilde{Y} \in \mathbb{C}^{n\times k}$. Then $Ye_j \in W_j$ for $j = 1, \ldots, k$, $Y^*Y = I_k$, and $\operatorname{tr}(Y^*AY) = \operatorname{tr}(\tilde{Y}^*B\tilde{Y})$. The proof is done if we can show that

$$\sum_{j=1}^{k} \lambda_{i_j}(A) \geq \sum_{j=1}^{\ell} \lambda_{i_j}(B) + \sum_{j=q-1}^{n-1} \lambda_j(B). \tag{7.2}$$

To this end, note that $\sum_{j=q-1}^{n-1} \lambda_{i_j}(B) = \sum_{j=q}^{n} \lambda_j(A) = \sum_{j=\ell+1}^{k} \lambda_{i_j}(A)$ by construction, and for $j = 1, \ldots, \ell$, $\lambda_{i_j}(B)$ can be viewed as $\min\{y^*Ay : y \in V_j\}$, where V_j is the i_j-dimensional subspace of $\mathbb{C}^n$ containing $Z\tilde{x}_1, \ldots, Z\tilde{x}_{i_j}$, where $\tilde{x}_1, \ldots, \tilde{x}_{i_j}$ are the eigenvectors of B corresponding to the eigenvalues $\lambda_1(B), \ldots, \lambda_{i_j}(B)$. By the special case (b), we have $\lambda_{i_j}(A) \geq \lambda_{i_j}(B)$. Thus (7.2) follows. □

We note the proofs of the special cases (a) and (b) are very short, and one can actually deduce several well-known matrix inequalities from them.

First, if one considers an $(n-1)\times(n-1)$ principal submatrix B of A, and applies the result of case (a) to W_1, which is spanned by the eigenvectors of B corresponding to the k largest eigenvalues, then we have $\lambda_k(A) \geq \lambda_k(B)$ Applying the result to $-A$ and $-B$, we see that $\lambda_k(B) \geq \lambda_{k+1}(A)$ for $k = 1, \ldots, n-1$. The result is the Cauchy interlacing theorem.

Second, if one applies the result of case (b) to W_j, which is spanned by j standard basis vectors, then we see that the sum of any k diagonal entries of A is not larger than $\sum_{j=1}^{k} \lambda_j(A)$ for all $k = 1, \ldots, n$. Clearly, the equality holds when $k = n$ by the trace condition. This is the well-known fact that the eigenvalues of a Hermitian matrix majorize its diagonal entries.

For positive semi-definite matrices, we have the following multiplicative analog of Theorem 7.1 due to Hoffman.

Theorem 7.2 [1, 2.16] *Let A be an $n\times n$ positive semi-definite matrix. For any indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,*

$$\prod_{j=1}^{k} \lambda_{i_j}(A) = \max_{\substack{W_1\subset W_2\subset\cdots\subset W_k \\ \dim W_j = i_j}} \min_{\substack{Y\in\mathbb{C}^{n\times k} \\ Ye_j\in W_j,\ Y^*Y=I_k}} \det(Y^*AY), \tag{7.3}$$

where $W_1 \subset \cdots \subset W_k$ *are subspaces of* $\mathbb{C}^n$.

One may wonder whether we can extend Theorem 7.2 to Hermitian matrices by replacing $\prod_{j=1}^{k} \lambda_{i_j}(A)$ by $|\prod_{j=1}^{k} \lambda_{i_j}(A)|$ and $\det(Y^*AY)$ by $|\det(Y^*AY)|$. Unfortunately, as shown by the following example, the resulting statement is false.

Example 7.3 Let $A = \mathrm{diag}(3, -1, -2, -3)$. Then

$$\prod_{i=1}^{3} \lambda_i(A) = 6 < 12 = \det(\mathrm{diag}(3, -2, -3)).$$

If A is positive semidefinite then $\lambda_i(A) = \sigma_i(A)$, and we may regard 7.2 as a statement about singular values. We do have the following representation of the product of the k largest singular values of general and normal matrices.

Theorem 7.4 *Let A be an $n \times n$ matrix, and let $1 \le k \le n$. Suppose $i_j = j$ for $j = 1, 2, \ldots, k$. Then*

$$\left|\prod_{j=1}^{k} \sigma_{i_j}(A)\right| = \max_{\substack{U_1 \subset U_2 \subset \cdots \subset U_k \\ \dim U_j = i_j \\ W_1 \subset W_2 \subset \cdots \subset W_k \\ \dim W_j = i_j}} \min_{\substack{X, Y \in \mathbb{C}^{n\times k} \\ Xe_j \in U_j,\ Ye_j \in W_j \\ X^*X = Y^*Y = I_k}} |\det(X^*AY)|, \tag{7.4}$$

If A is normal, then

$$\left|\prod_{j=1}^{k} \sigma_{i_j}(A)\right| = \max_{\substack{W_1 \subset W_2 \subset \cdots \subset W_k \\ \dim W_j = i_j}} \min_{\substack{Y \in \mathbb{C}^{n\times k} \\ Ye_j \in W_j,\ Y^*Y = I_k}} |\det(Y^*AY)|. \tag{7.5}$$

The following example shows that Theorem 7.1 is false without the condition $i_j = j$.

Example 7.5 Let A be the Hermitian matrix $\mathrm{diag}(1, 1, -1, -1)$. Then for $i_1 = 3$, we have $\sigma_3(A) = 1$. However, for any subspace W_1 with $\dim W_1 = 3$, the compression of A on W_1 is indefinite. Thus we can find a unit vector y such that $y^*Ay = 0$. It follows that

$$\min_{\substack{y \in W_1 \\ y^*y = 1}} |y^*Ay| = 0,$$

and it is also clear that for any other 3 dimensional subspace U_1,

$$\min_{\substack{z \in U_1, y \in W_1 \\ y^*y = z^*z = 1}} |y^*Az| = 0.$$

Acknowledgement. Thanks are due to the referees for their comments that helped improve the presentation.

References

1. Ali Amir-Moez, N. (1990): Extreme Properties of Linear Transformations. Polygonal Publishing, New Jersey
2. Bhatia, R. (1987): Perturbation Bounds for Matrix Eigenvalues. Pitman Research Notes in Mathematics 162. Longman Scientific and Technical, New York
3. Bhatia, R. (1997): Matrix Analysis. Springer-Verlag, New York
4. Demmel, J., Veselić, K. (1992): Jacobi's method is more accurate than QR. SIAM J. Matrix Anal. Appl. **13**:1204–1245
5. Eisenstat, S.C., Ipsen, I.C.F. (1994): Relative perturbation techniques for singular value problems. SIAM J. Num. Anal. **32**:1972–1988
6. Eisenstat, S.C., Ipsen, I.C.F. (1997): Absolute perturbation bounds for matrix eigenvalues imply relative bounds. TR 97-16, Center for Research in Scientific Computation, Department of Mathematics, North Carolina State University, Raleigh, NC.
7. Gelfand, I.M., Naimark, M. (1950): The relations between the unitary represention of the complex unimodular group and its unitary subgroup. Izv. Akad. Nauk SSSR Ser. Mat., 14:239–260
8. Horn, R.A., Johnson, C.R. (1985): Matrix Analysis. Cambridge University Press, New York
9. Horn, R.A., Johnson, C.R. (1991): Topics in Matrix Analysis. Cambridge University Press, New York
10. Ipsen, I.C.F. (1998): Relative perturbation results for matrix eigenvalues and singular values. Acta Numerica **7**:151–201
11. Li, C.-K., Mathias, R. (1998): Generalizations of Ky Fan's dominance theorem. SIAM J. Matrix Anal. Appl. **19**:99–106
12. Li., R.-C. (1994): Relative perturbation theory: (I) eigenvalue and singular value variations. to appear in SIMAX, also published (in an earlier version with different theorem numbers) as LAPACK working note 84
13. Li., R.-C. (1997): Relative perturbation theory: (III) more bounds on eigenvalue variation. Linear Algebra Appl. **266**:337–345
14. Li., R.-C. (1994): Relative perturbation theory: (I) eigenvalue and singular value variations. LAPACK working note 84.
15. Marshall, A.W., Olkin., I. (1979): Inequalities: Theory of Majorization and its Applications. Academic Press, London
16. Mathias, R. (1998): Quadratic residual bounds for the Hermitian eigenvalue problem. SIAM J. Matrix Anal. Appl. **19**:541–559
17. Mathias, R. (1995): Accurate eigensystem computations by Jacobi methods. SIAM J. Matrix Anal. Appl., 16(3):977–1003
18. Olver, F.W.J. (1978): A new approach to error arithmetic. SIAM. J. Num. Anal., 15:368–393
19. Stewart, G., Sun, J.-G. (1990): Matrix Perturbation Theory. Academic Press, Boston
20. Thompson, R.C. (1973): On the singular values of a matrix product-II. Scripta Math., 29:111–114
21. Thompson, R.C., Freede, Linda (1971): On the eigenvalues of sums of hermitian matrices. Linear Algebra Appl., 4:369–376