Norms and Inequalities for Condition Numbers, III*

ALBERT W. MARSHALL AND INGRAM OLKIN
University of Rochester
Rochester, New York
Stanford University
Stanford, California

Communicated by Olga Taussky-Todd

ABSTRACT

The main results provide comparisons between condition numbers (based on unitarily invariant norms) of (i) positive definite (Hermitian) matrices A, B and of $A + B$, (ii) a positive definite matrix and its principal submatrix, and (iii) a matrix and an augmented form of the matrix.

1. INTRODUCTION

The condition number c_φ of a nonsingular matrix A is defined by

$$c_\varphi(A) = \varphi(A)\varphi(A^{-1}),$$

where ordinarily φ is a norm. This definition can be extended to include singular and rectangular matrices by substituting the pseudoinverse A^+ for A^{-1}. Condition numbers arise in various contexts, and serve, e.g. as measures of the difficulty in solving a system of linear equations (see [1]).

For condition numbers based on norms that are unitarily invariant (i.e. $\varphi(A) = \varphi(AU) = \varphi(VA)$ for all unitary matrices U and V of appropriate order), we obtain the following comparisons.

PROPOSITION 1. *If $A : m \times q$ is of rank q and $(A, B) : m \times n$ is of rank n, then*

$$c_\varphi(A) \leqslant c_\varphi(A, B). \tag{1.1}$$

* Supported in part by the National Science Foundation Grant 17172, Stanford University.

PROPOSITION 2. *If $A : m \times n$ is of rank k, $H : n \times q$ is column orthonormal ($H^*H = I_q$), and rank $(A^*, H) =$ rank A^*, then*

$$c_\varphi(AH) \leqslant c_\varphi(A). \tag{1.2}$$

PROPOSITION 3. *If $A : m \times n$, and $\varepsilon > 0$, then*

$$c_\varphi(A + \varepsilon A^{*+}) \leqslant c_\varphi(A). \tag{1.3}$$

The inequalities are known for the case that $\varphi(A)$ is the maximum singular value of A. For this norm, (1.1) and (1.2) were obtained by Hanson and Lawson [2]; (1.3) was obtained by Klinger [3] when A is normal and nonsingular, and by Tewarson and Ramnath [8] without normality.

Unitarily invariant norms φ are monotone in the sense that if A and $B - A$ are positive semi-definite (Hermitian) then $\varphi(A) \leqslant \varphi(B)$. Marshall and Olkin [4, 5] and Marshall, Olkin, and Proschan [6] discuss the following propositions.

PROPOSITION 4. *If A is nonsingular and φ is unitarily invariant, then*

$$c_\varphi(A) \leqslant c_\varphi(AA^*). \tag{1.4}$$

PROPOSITION 5. *If A, B are positive definite and φ is a monotone norm, then*

$$c_\varphi(A + B) \leqslant \max[c_\varphi(A), c_\varphi(B)]. \tag{1.5}$$

We extend (1.4) to the case that A is singular or rectangular (Sec. 3), and show that no such general extension is possible for (1.5) (Sec. 4).

A reinterpretation of (1.1) yields the result that if

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$$

is a positive definite matrix, and φ is a unitarily invariant norm, then $c_\varphi(U^{1/2}) \geqslant c_\varphi(U_{11}^{1/2})$. In Sec. 5 we use results on majorization to show that

$$c_\varphi(U) \geqslant c_\varphi \begin{pmatrix} U_{11} & 0 \\ 0 & U_{22} \end{pmatrix} \geqslant c_\varphi(U_{11}).$$

2. PRELIMINARIES

In writing an inequality like $c_\varphi(A) \leqslant c_\varphi(A, B)$, or even in defining $c_\varphi(A) = \varphi(A)\varphi(A^+)$ we have acted as though φ is defined on matrices of various orders. If this is to be the case, we must be careful to specify what is meant by a norm because, e.g. the triangle inequality $\varphi(A + B) \leqslant \varphi(A) + \varphi(B)$ makes no sense if A and B are of different orders. However, if we assume that augmentation of a matrix by blocks of zeros to the right and below does not change its norm, i.e.

$$\varphi(A) = \varphi\left[\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}\right],$$

then we are free either to regard φ as defined on matrices of various orders, or to augment matrices by blocks of zeros to achieve a common order. To do this with impunity we must be sure that all norms φ defined on $m \times n$ matrices have, for all $i, j > 0$, the form

$$\varphi(A) = \hat{\varphi}\left[\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}\right], \tag{2.1}$$

for some norm $\hat{\varphi}$ defined on the $m + i \times n + j$ matrices. This is not difficult to demonstrate. But since we are concerned exclusively with unitarily invariant norms we want to be sure that $\hat{\varphi}$ can be found which is unitarily invariant whenever φ is unitarily invariant.

When φ is unitarily invariant, $\varphi(A)$ depends on A only through its singular values. More precisely, if $\alpha_1^2, \ldots, \alpha_n^2$ are the characteristic roots of A^*A,

$$\varphi(A) = \Phi(\alpha_1, \ldots, \alpha_n), \tag{2.2}$$

for some symmetric gauge function (SGF) Φ (see [7]). If we define $\hat{\Phi}(x_1, \ldots, x_{n+j}) = \Phi[x_{(1)}, \ldots, x_{(n)}]$ where $|x_{(1)}| \geqslant \cdots \geqslant |x_{(n+j)}|$ are obtained by reordering $x_1, \ldots, x_{n+j}$, then $\hat{\Phi}$ is an SGF which gives rise to a unitarily invariant norm $\hat{\varphi}$ satisfying (2.1).

In view of these remarks, one sees immediately that because (1.4) holds for nonsingular matrices A, it must also hold for singular and rectangular matrices.

Suppose A is an arbitrary matrix, $U = A^*A$ and $V = AA^*$. The *nonzero* singular values of A, $U^{1/2}$, $V^{1/2}$ are identical. When φ is unitarily invariant, $\varphi(A)$ depends only on these singular values, so that

$$c_\varphi(A) = c_\varphi(U^{1/2}) = c_\varphi(V^{1/2}). \tag{2.3}$$

3. AUGMENTED MATRICES

Using (2.3) and the notation $U = (A, B)^*(A, B)$, $U_{11} = A^*A$, we see that for any unitarily invariant norm φ, inequality (1.1) becomes

$$c_\varphi(U_{11}^{1/2}) \leqslant c_\varphi(U^{1/2}). \tag{3.1}$$

Here, U is positive definite because rank $(A, B) = n$.

Inequality (1.2) can be similarly rewritten: Since $H^*H = I_q$, there exist unitary matrices Γ and Δ such that

$$H = \Gamma\begin{pmatrix} I_q \\ 0 \end{pmatrix}\Delta.$$

Let $B = A\Gamma$, $U = B^*B$ and define U_{11} by

$$B^*\begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix}B = \begin{pmatrix} U_{11} & 0 \\ 0 & 0 \end{pmatrix}.$$

With this notation, (1.2) also becomes (3.1), but now U has rank k and is not positive definite unless $k = n$. However,

$$\operatorname{rank}(A^*, H) = \operatorname{rank}\left\{\Gamma^*(A^*, H)\begin{pmatrix} I & 0 \\ 0 & \Delta^* \end{pmatrix}\right\} = \operatorname{rank}\left[B^*, \begin{pmatrix} I_q \\ 0 \end{pmatrix}\right]$$

so that the condition $\operatorname{rank}(A^*, H) = \operatorname{rank} A^*$ of (1.2) is equivalent to

$$\operatorname{rank}\left[B^*, \begin{pmatrix} I_q \\ 0 \end{pmatrix}\right] = \operatorname{rank} B. \tag{3.2}$$

Of course, this means $q \leqslant k = \operatorname{rank} B$.

To complete the proof of (1.1) and (1.2), it remains to be shown that (3.1) holds. Denote the characteristic roots of $U = (u_{ij})_{i,j=1}^n$ by $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_n \geqslant 0$ and the characteristic roots of $U_{n-1} = (u_{ij})_{i,j=1}^{n-1}$ by $\gamma_1 \geqslant \cdots \geqslant \gamma_{n-1} \geqslant 0$. According to the separation theorem of Sturm,

$$\alpha_1 \geqslant \gamma_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \gamma_{n-1} \geqslant \alpha_n.$$

This shows that $\alpha_i \geqslant \gamma_i$, $i = 1, 2, \ldots, n-1$, and iteration of the argument yields $\alpha_i \geqslant \beta_i$, $i = 1, 2, \ldots, q$, where $\beta_1 \geqslant \cdots \geqslant \beta_q > 0$ are the characteristic roots of U_{11}. Providing that rank $U = k > q =$ rank U_{11}, it follows in a similar fashion that $\alpha_{k-j} \leqslant \beta_{q-j}$, $j = 0, 1, \ldots, q-1$. The monotonicity of the SGF Φ related to φ via (2.2), and $\alpha_i \geqslant \beta_i > 0$, $i = 1, 2, \ldots, q$, together yield

$$\Phi(\alpha_1, \ldots, \alpha_n) \geqslant \Phi(\alpha_1, \ldots, \alpha_q, 0, \ldots, 0) \geqslant \Phi(\beta_1, \ldots, \beta_q, 0, \ldots, 0). \quad (3.3)$$

The same monotonicity and $\beta_{q-j} \geqslant \alpha_{k-j} > 0$, $j = 0, 1, \ldots, q-1$ give

$$\Phi(\alpha_1^{-1}, \ldots, \alpha_n^{-1}) \geqslant \Phi(\alpha_k^{-1}, \alpha_{k-1}^{-1}, \ldots, \alpha_{k-q+1}^{-1}, 0, \ldots, 0)$$
$$\geqslant \Phi(\beta_q^{-1}, \beta_{q-1}^{-1}, \ldots, \beta_1^{-1}, 0, \ldots, 0). \quad (3.4)$$

The combination of (3.3) and (3.4) proves (3.1) under the condition that rank $U >$ rank U_{11}, as it is for (1.1) when U is positive definite.

It remains to be shown that (3.1) holds under the conditions (3.2) and $k = q$. With $k = q$, (3.2) implies that $B = (B_1, 0)$ where $B_1 : m \times q$. Consequently,

$$U = \begin{pmatrix} B_1^*B_1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} U_{11} & 0 \\ 0 & 0 \end{pmatrix}.$$

Since the nonzero roots of U and U_{11} coincide, (3.1) is trivial.

4. SUMS OF MATRICES

In considering the possibilities of extending (1.5) to matrices that are not positive definite, we begin with two simple counterexamples. The first of these shows that nonsingularity is insufficient for (1.5); the second shows that positive semidefiniteness is insufficient.

I. Let

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then $AA^* = 4I$ and $BB^* = I$, so that $c_\varphi(A) = c_\varphi(B) = 1$ whenever φ is unitarily invariant. On the other hand, $c_\varphi(A + B) > 1$, e.g. when φ is the spectral norm.

II. Let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Again, $c_\varphi(A) = c_\varphi(B) = 1$ when φ is unitarily invariant, but $c_\varphi(A+B) > 1$, e.g. when φ is the spectral norm.

The proof of (1.5) given by Marshall and Olkin [5] depends upon the convexity of the inverse function on the domain of positive definite matrices:

$$[\theta U_1 + (1-\theta)U_2]^{-1} \leqslant \theta U_1^{-1} + (1-\theta)U_2^{-1},$$

whenever $0 \leqslant \theta \leqslant 1$, U_1 and U_2 are positive definite, and where $A \leqslant B$ means $B - A$ is positive semidefinite. However, the pseudoinverse is not convex on the domain of positive semidefinite matrices. To see this, let U_1 be positive definite, $U_2 = 0$. Then for $0 < \theta < 1$,

$$[\theta U_1 + (1-\theta)U_2]^+ = \theta^{-1}U_1^{-1} > \theta U_1^{-1} = \theta U_1^+ + (1-\theta)U_2^+.$$

It is, however, possible to extend (1.5) in a rather trivial but useful way.

PROPOSITION 6. *If A, B are positive semidefinite,* rank A = rank B = rank(A, B), *and if φ is unitarily invariant, then* (1.5) *holds.*

To see this we take

$$B = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix},$$

where D is diagonal without loss of generality because φ is unitarily invariant. Then the rank condition and symmetry of A imply that

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

where A_{11} and D are of the same size. The application of (1.5) to A_{11} and D completes the proof.

This result can be used to show (1.3) for unitarily invariant norms as follows. Let $\gamma_1^2 \geqslant \cdots \geqslant \gamma_m^2$ be the characteristic roots of

$$[A + \varepsilon(A^*)^+][A + \varepsilon(A^*)^+]^* = AA^* + 2\varepsilon AA^+ + \varepsilon^2(AA^*)^+.$$

If we write

$$A = \Gamma \begin{pmatrix} D_\alpha & 0 \\ 0 & 0 \end{pmatrix} \Delta,$$

where $D_\alpha = \operatorname{diag}(\alpha_1,\ldots,\alpha_k)$ so that the α_i are the roots of $(AA^*)^{1/2}$, then

$$A^+ = \Delta^* \begin{pmatrix} D_\alpha^{-1} & 0 \\ 0 & 0 \end{pmatrix} \Gamma^*,$$

and γ_j^2 are the roots of

$$\Gamma \begin{pmatrix} D_\alpha^2 & 0 \\ 0 & 0 \end{pmatrix} \Gamma^* + 2\varepsilon \Gamma \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \Gamma^* + \varepsilon^2 \Gamma \begin{pmatrix} D_\alpha^{-2} & 0 \\ 0 & 0 \end{pmatrix} \Gamma^*$$

$$= \begin{pmatrix} D_\alpha^2 + 2\varepsilon I + D_\alpha^{-2} & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus $\gamma_j = \alpha_{i_j} + \varepsilon\alpha_{i_j}^{-1}$, $j = 1, 2,\ldots, k$, $\gamma_j = 0$, $j = k+1,\ldots, m$, for some permutation i_j. Consequently, with $W = (AA^*)^{1/2}$ inequality (1.3) can be written as

$$c_\varphi(W + \varepsilon W^+) \leqslant c_\varphi(W).$$

But $c_\varphi(W) = c_\varphi(\varepsilon W^+)$, so that this follows from the above generalization of (1.5).

As a very special case of (1.3), we have for positive definite matrices A that, for any $u_{-1}, u_1 > 0$,

$$c_\varphi(u_{-1}A^{-1} + u_1 A) \leqslant c_\varphi(A). \tag{4.1}$$

It is of interest to compare this with the following [5]: If A is positive definite, φ is unitarily invariant, $1 \leqslant v_1 \leqslant \cdots \leqslant v_l$ and $u_i \geqslant 0$, $0 \leqslant i \leqslant l$, then

$$c_\varphi(A) \leqslant c_\varphi(u_0 A + \cdots + u_l A^{v_l}) \leqslant c_\varphi(A^{v_l}),$$

$$c_\varphi(A^{-1}) \leqslant c_\varphi(u_0 A^{-1} + \cdots + u_l A^{-v_l}) \leqslant c_\varphi(A^{-v_l}).$$

One might be tempted to conjecture that (4.1) can be extended as follows:

$$c_\varphi(u_{-1}A^{-1} + u_1 A) \leqslant c_\varphi(u_{-l}A^{-v_l} + \cdots + u_{-1}A^{-1} + u_1 A + \cdots + u_j A^{z_j}).$$

This is false, as can be seen by taking $A = \operatorname{diag}(1, \frac{1}{2})$ and $\Phi(x_1, x_2) = \max(|x_1|, |x_2|)$. Then

$$c_\varphi(A^{-1} + A + A^2) = \frac{12}{11} < \frac{5}{4} = c_\varphi(A^{-1} + A).$$

5. AN APPLICATION OF MAJORIZATION

We have shown in Sec. 3 that for any unitarily invariant norm φ, and for positive definite and certain other matrices

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix},$$

$$c_\varphi(U^{1/2}) \geqslant c_\varphi(U_{11}^{1/2}).$$

Using different methods, we show here that

$$c_\varphi(U) \geqslant c_\varphi \begin{pmatrix} U_{11} & 0 \\ 0 & U_{22} \end{pmatrix} \geqslant c_\varphi(U_{11}) \tag{5.1}$$

whenever U is positive semidefinite and φ is unitarily invariant.

The second inequality of (5.1) is immediate from the fact that unitarily invariant norms are monotone (see Sec. 1). To prove the first inequality, let

$$\Gamma = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix},$$

be a unitary matrix such that $\Gamma_1 U_{11} \Gamma_1{}^* = \operatorname{diag}(\beta_1, \ldots, \beta_q) \equiv D_\beta$, and $\Gamma_2 U_{22} \Gamma_2{}^* = \operatorname{diag}(\delta_1, \ldots, \delta_{n-q}) \equiv D_\delta$. Of course, the characteristic roots $\alpha_1, \alpha_2, \ldots, \alpha_n$ of U are the same as the characteristic roots of

$$\Gamma U \Gamma^* = \begin{pmatrix} D_\beta & \Gamma_1 U_{12} \Gamma_2{}^* \\ \Gamma_2 U_{12}^* \Gamma_1{}^* & D_\delta \end{pmatrix}.$$

According to a result of I. Schur (e.g., see [6]), the vector $d = (\beta, \delta) = (d_1, d_2, \ldots, d_n)$ of diagonal elements of a positive semidefinite matrix is majorized by the vector $\alpha = (\alpha_1, \ldots, \alpha_n)$ of characteristic roots in the sense that, possibly after reordering components,

$$d_1 \geqslant \cdots \geqslant d_n, \qquad \alpha_1 \geqslant \cdots \geqslant \alpha_n$$

$$\sum_{i=1}^{l} d_i \leqslant \sum_{i=1}^{l} \alpha_i, \qquad l = 1, 2, \ldots, n-1, \qquad \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} \alpha_i.$$

This means [4, Lemma 3.3] that if k is defined by $d_k > 0$, $d_{k+1} = \cdots = d_n = 0$ (so also $\alpha_k > 0$, $\alpha_{k+1} = \cdots = \alpha_n = 0$), and if Φ is the SGF which corresponds to φ as in (2.2), then

$$\Phi(d_1,\ldots,d_k,0,\ldots,0) \leqslant \Phi(\alpha_1,\ldots,\alpha_k,0,\ldots,0),$$

$$\Phi(d_1^{-1},\ldots,d_k^{-1},0,\ldots,0) \leqslant \Phi(\alpha_1^{-1},\ldots,\alpha_k^{-1},0,\ldots,0).$$

These inequalities together prove the first inequality of (5.1).

From the fact that the characteristic roots of

$$\begin{pmatrix} U_{11} & 0 \\ 0 & U_{22} \end{pmatrix}$$

are majorized by the characteristic roots of U, one might conjecture that if

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}, \qquad \tilde{A} = \begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & A_{23} \\ 0 & A_{32} & A_{33} \end{bmatrix}$$

are positive definite, then the characteristic roots of $\tilde{A}$ are majorized by the characteristic roots of A. That this is false can be seen from the choice

$$A = \begin{pmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{pmatrix}, \qquad \tilde{A} = \begin{pmatrix} 1 & a & 0 \\ a & 1 & a \\ 0 & a & 1 \end{pmatrix}, \qquad |a| < \frac{1}{\sqrt{2}}.$$

The characteristic roots $(1 - a, 1 - a, 1 + 2a)$ of A, and the characteristic roots $(1 - a\sqrt{2}, 1, 1 + a\sqrt{2})$ of $\tilde{A}$ are not ordered either way by majorization.

Since we have obtained $c_\varphi(U_{11}) \leqslant c_\varphi(U)$ and $c_\varphi(U_{11}^{1/2}) \leqslant c_\varphi(U^{1/2})$, it is natural to inquire if one of the inequalities

$$\text{(i)} \quad c_\varphi(A) \leqslant c_\varphi(B), \qquad \text{(ii)} \quad c_\varphi(A^{1/2}) \leqslant c_\varphi(B^{1/2})$$

is implied by the other. If $A = \operatorname{diag}(625, 25, 1)$ and $B = \operatorname{diag}(325, 325, 1)$ then it is easily checked that with $\varphi(A) = (\operatorname{tr} AA^*)^{1/2}$, (i) is violated but (ii) holds. On the other hand, interchanging these special A and B shows that (i) can hold when (ii) is violated.

A comparison of (3.1) and (5.1) suggests the possibility that

$$c_\varphi(U^{1/2}) \geqslant c_\varphi \begin{pmatrix} U_{11}^{1/2} & 0 \\ 0 & U^{1/2} \end{pmatrix}$$

when U is positive definite. Whether or not this is true remains an open question.

REFERENCES

1 G. Forsythe and C. Moler, *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, Englewood Cliffs, N.J. (1967).
2 R. J. Hanson and C. L. Lawson, Extensions and applications of the Householder algorithm for solving linear least squares problems. *Math. Comp.* **23**(1969), 787–812.
3 A. Klinger, Approximate pseudoinverse solutions to ill-conditioned linear systems. *J. Optimization Theory Appl.* **2**(1968), 117–124.
4 A. W. Marshall and I. Olkin, Norms and inequalities for condition numbers. *Pacific J. Math.* **15**(1965), 241–247.
5 A. W. Marshall and I. Olkin, Norms and inequalities for condition numbers, II. *Linear Algebra and Appl.* **2**(1969), 167–172.
6 A. W. Marshall, I. Olkin, and F. Proschan, Monotonicity of ratios of means and other applications of majorization. *Inequalities, Proceedings of a Symposium* (O. Shisha, Ed.). Academic Press, New York (1967), pp. 177–190.
7 J. von Neumann, Some matrix-inequalities and metrization of matric-space, *Tomsk Univ., Rev.* **1**(1937), 286–300; *Collected Works*, vol. IV, Pergamon Press, New York (1962).
8 R. P. Tewarson and B. Ramnath, Some comments on the solution of linear equations. *BIT* **9**(1968), 167–173.

Received April, 1971