NORMS AND INEQUALITIES FOR CONDITION NUMBERS

Albert W. Marshall and Ingram Olkin
NORMS AND INEQUALITIES FOR CONDITION NUMBERS

ALBERT W. MARSHALL AND INGRAM OLKIN

The condition number $c_\varphi$ of a nonsingular matrix $A$ is defined by $c_\varphi(A) = \varphi(A)\varphi(A^{-1})$ where ordinarily $\varphi$ is a norm. It was proved by O. Taussky-Todd that (c) $c_\varphi(A) \leq c_\varphi(AA^*)$ when $\varphi(A) = (\text{tr } AA^*)^{1/2}$ and when $\varphi(A)$ is the maximum absolute characteristic root of $A$. It is shown that (c) holds whenever $\varphi$ is a unitarily invariant norm, i.e., whenever $\varphi$ satisfies $\varphi(A) > 0$ for $A \neq 0$; $\varphi(\alpha A) = |\alpha| \varphi(A)$ for complex $\alpha$; $\varphi(A + B) \leq \varphi(A) + \varphi(B)$; $\varphi(A) = \varphi(AU) = \varphi(A)$ for all unitary $U$. If in addition, $\varphi(E_{ij}) = 1$, where $E_{ij}$ is the matrix with one in the $(i, j)$th place and zeros elsewhere, then $c_\varphi(A) \geq c_\varphi(AA^*)^{1/2}$. Generalizations are obtained by exploiting the relation between unitarily invariant norms and symmetric gauge functions. However, it is shown that (c) is independent of the usual norm axioms.

1. Introduction. The genesis of this study is the proposition that under certain conditions, the matrix $AA^*$ is more "ill-conditioned" than $A$. More precisely, the condition number $c_\varphi(A)$ is defined for nonsingular matrices $A$ as

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

where ordinarily $\varphi$ is a norm. The statement concerning ill-conditioning of $AA^*$ is the inequality

$$(c) \quad c_\varphi(A) \leq c_\varphi(AA^*) .$$

Where $\varphi(A)$ is the maximum absolute characteristic root of $A$ and where $\varphi(A) = (\text{tr } AA^*)^{1/2}$, inequality (c) was proved by O. Taussky-Todd [7]. This raises the question of whether (c) is true for all norms. In this paper, we show that quite the contrary is true; (c) is independent of the usual norm axioms. However, we also prove that (c) does hold for a quite general class of norms.

In the course of proving these results, we obtain some inequalities for symmetric gauge functions, which may be of independent interest.

2. Gauge functions and matrix norms. We call $\varphi$ a matrix norm if

$$(\text{ai}) \quad \varphi(A) > 0 \quad \text{when } A \neq 0 ,$$

Received March 10, 1964.
In addition to these basic axioms, various other conditions are sometimes imposed:

(aIV) \( \varphi(E_{ij}) = 1 \),

where \( E_{ij} \) is the matrix with one in the \((i,j)\)th position and zero elsewhere,

(aV) \( \varphi(AB) \leq \varphi(A)\varphi(B) \),

(aVI) \( \varphi(A) = \varphi(UA) = \varphi(AU) \) for all unitary matrices \( U \).

If \( \varphi \) satisfies aI, aII, aIII, and aVI, \( \varphi \) is called a unitarily invariant norm.

There is an important connection between unitarily invariant norms and symmetric gauge functions. A function \( \Phi \) on a complex vector space is called a gauge function if

(bI) \( \Phi(u) > 0 \) when \( u \neq 0 \),

(bII) \( \Phi(\alpha u) = |\alpha| \Phi(u) \) for complex \( \alpha \),

(bIII) \( \Phi(u + v) \leq \Phi(u) + \Phi(v) \).

Often it is convenient to assume, in addition, that

(bIV) \( \Phi(e_i) = 1 \),

where \( e_i \) is the vector with one in the \( i \)th place and zero elsewhere. If, in addition to bI, bII, and bIII,

(bV) \( \Phi(u_1, \cdots, u_n) = \Phi(\varepsilon_1u_{i_1}, \cdots, \varepsilon_nu_{i_n}) \)

whenever \( \varepsilon_j = \pm 1 \) and \((i_1, \cdots, i_n)\) is a permutation of \((1, \cdots, n)\), then \( \Phi \) is called a symmetric gauge function.

It was noted by Von Neumann [8] that a norm \( \varphi \) is unitarily invariant if and only if there exists a symmetric gauge function \( \Phi \) such that \( \varphi(A) = \Phi(\alpha) \) for all \( A \), where \( \alpha_1, \cdots, \alpha_n \) are the eigenvalues of \( AA^* \).

If \( \Phi \) is a symmetric gauge function and \( u, v \) satisfy \( u_i \leq v_i, i = 1, \cdots, n \), then it follows [6, p. 85] that

\[
(2.1) \quad \Phi(u_1, \cdots, u_n) \leq \Phi(v_1, \cdots, v_n).
\]

If \( \Phi \) is a symmetric gauge function satisfying bIV, then [6, p. 86]

\[
(2.2) \quad \max_i |u_i| \leq \Phi(u_1, \cdots, u_n) \leq \sum_{i=1}^n |u_i|.
\]
If \( \varphi \) is the unitarily invariant matrix norm determined by \( \Phi \) as above, then it follows that

\[
\frac{\varphi(AB)}{\varphi(A)\varphi(B)} \leq \frac{\sum_{i=1}^{n} \lambda_i(ABB^*A^*)}{[\max_i \lambda_i(AA^*)][\max_j \lambda_j(BB^*)]}
\leq \frac{n \max_i \lambda_i(BB^*A^*A)}{[\max_i \lambda_i(AA^*)][\max_j \lambda_j(BB^*)]} \leq n,
\]

where \( \lambda_i(M) \) are the eigenvalues of \( M \). Thus, for any \( k \geq n, k\varphi \) is a unitarily invariant matrix norm also satisfying aIV. Of course, \( \varphi \) itself satisfies aIV (since \( \Phi \) satisfies bIV), and this property is destroyed by the renormalization.

3. The condition number inequality.

**Theorem 3.1.** If \( \varphi \) is a unitarily invariant norm, then

\[ c_\varphi(A) \leq c_\varphi(AA^*) . \]

If \( \Phi \) is a symmetric gauge function which determines \( \varphi \), then we may rewrite (c) in the form

\[ \Phi(\alpha_1, \ldots, \alpha_n)\Phi(\alpha_1^{-1}, \ldots, \alpha_n^{-1}) \leq \Phi(\alpha_1^2, \ldots, \alpha_n^2)\Phi(\alpha_1^{-2}, \ldots, \alpha_n^{-2}) . \]

Thus, Theorem 3.1 is a very special case of

**Theorem 3.2.** If \( \Phi \) is a symmetric gauge function, then

\[ \Phi(\alpha_1, \ldots, \alpha_n)\Phi(\alpha_1^{-r}, \ldots, \alpha_n^{-r}) \text{ is increasing in } r > 0, \text{ where } \alpha_i > 0. \]

The proof of Theorem 3.2 is embodied in the lemmas below.

Following [2] we say \((a_1, \ldots, a_n)\) is majorized by \((b_1, \ldots, b_n)\), written \((a) < (b)\), if

(i) \[ a_1 \geq \cdots \geq a_n > 0, \quad b_1 \geq \cdots \geq b_n > 0, \]

(ii) \[ \sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i, \quad k = 1, \ldots, n - 1, \]

(iii) \[ \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i . \]

**Lemma 3.3.** If \((a) < (b)\), and \( \Phi \) is a symmetric gauge function, then

\[ \Phi(a_1, \ldots, a_n) \leq \Phi(b_1, \ldots, b_n) , \]

\[ \Phi(a_1^{-r}, \ldots, a_n^{-r}) \leq \Phi(b_1^{-r}, \ldots, b_n^{-r}) . \]

**Proof.** Proofs of (3.1) have been given by Fan [1] and Ostrowski.
First, note that we can assume for $h$ and $j$ fixed, $h < j$,

\[ a_h = \alpha b_h + (1 - \alpha) b_j, \quad a_j = (1 - \alpha) b_h + \alpha b_j, \quad a_i = b_i, \quad i \neq h, j. \]

That this is true follows from the fact that if $(a) < (b)$, then $a$ can be derived from $b$ by successive applications of a finite number of transformations of the form (3.3) (see [2, p. 47]).

Let $\vec{b} = (b_1, \ldots, b_{h-1}, b_j, b_{h+1}, \ldots, b_{j-1}, b_h, b_{j+1}, \ldots, b_n)$, so that $\Phi(b_1, \ldots, b_n) = \Phi(\vec{b}_1, \ldots, \vec{b}_n)$. By convexity,

\[ (ab_i + (1 - \alpha) \vec{b}_i)^{-1} \leq ab_i^{-1} + (1 - \alpha) \vec{b}_i^{-1}. \]

Then using (2.1) and the convexity of $\Phi$, it follows that

\[ \Phi(a_1^{-1}, \ldots, a_n^{-1}) = \Phi[(ab_1 + (1 - \alpha) \vec{b}_1)^{-1}, \ldots, (ab_n + (1 - \alpha) \vec{b}_n)^{-1}] \leq \Phi[(ab_1^{-1} + (1 - \alpha) \vec{b}_1^{-1}, \ldots, (ab_n^{-1} + (1 - \alpha) \vec{b}_n^{-1}) \leq \alpha \Phi(b_1^{-1}, \ldots, b_n^{-1}) + (1 - \alpha) \Phi(\vec{b}_1^{-1}, \ldots, \vec{b}_n^{-1}). \]

As a consequence of Lemma 3.3., we have that if $(a) < (b)$ then

\[ \Phi(a_1, \ldots, a_n) \Phi(a_1^{-1}, \ldots, a_n^{-1}) \leq \Phi(b_1, \ldots, b_n) \Phi(b_1^{-1}, \ldots, b_n^{-1}). \]

The proof of Theorem 3.2 is completed by the following

**Lemma 3.4.** If $\alpha_1 \geq \cdots \geq \alpha_n > 0$ and $\alpha_r = \alpha_r^i / \Sigma \alpha_r^i$, $b_i = \alpha_i^r / \Sigma \alpha_i^r$, $0 < r < s$, then $(a) < (b)$.

**Proof.** We must show that for all $k$,

\[ \frac{\sum_{i=1}^{k} \alpha_i^r}{\sum_{i=1}^{n} \alpha_i^r} \leq \frac{\sum_{i=1}^{k} \alpha_i^s}{\sum_{i=1}^{n} \alpha_i^s}, \quad r < s, \]

which is true if and only if

\[ \sum_{i=1}^{k} \alpha_i^r \sum_{j=k+1}^{n} \alpha_j^s - \sum_{i=1}^{k} \alpha_i^s \sum_{j=k+1}^{n} \alpha_j^r = \sum_{i=1}^{k} \alpha_i^r \sum_{j=k+1}^{n} \alpha_j^r (\alpha_i^r - \alpha_j^r) \geq 0. \]

The latter follows from $\alpha_i \geq \alpha_j, \quad i < j$. ||

Observe that by (3.1) and Lemma 3.4, we have

\[ \frac{\Phi(\alpha_1^r, \ldots, \alpha_n^r)}{\Phi(\alpha_1^s, \ldots, \alpha_n^s)} \leq \frac{\Sigma \alpha_i^r}{\Sigma \alpha_i^s}. \]

In view of (2.2), it is perhaps natural to expect that

\[ \frac{\alpha_i^r}{\alpha_i^s} \leq \frac{\Phi(\alpha_1^r, \ldots, \alpha_n^r)}{\Phi(\alpha_1^r, \ldots, \alpha_n^s)} \leq \frac{\Sigma \alpha_i^r}{\Sigma \alpha_i^s}, \quad 0 < r < s, \quad \alpha_1 \geq \cdots \geq \alpha_n > 0, \]
for any symmetric gauge function $\Phi$. To see this we need only prove the left hand inequality, which may be written in the form

\[(3.5) \quad \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right) \leq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right),\]

and which is a consequence of (2.1).

An interesting counterpart to Theorem 3.2 can be obtained from (3.4).

**Theorem 3.5.** If $\Phi$ is a symmetric gauge function satisfying $\text{blIV}$, then $[\Phi(\alpha_1^r, \ldots, \alpha_n^r)]^{1/r}$ is decreasing in $r > 0$ whenever $\alpha_i > 0$, $i = 1, 2, \ldots, n$. Thus $[\Phi(\alpha_1^r, \ldots, \alpha_n^r)\Phi(\alpha_1^{-r}, \ldots, \alpha_n^{-r})]^{1/r}$ is decreasing in $r > 0$.

**Proof.** We have that

\[1 \leq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right) \leq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right),\]

the first inequality by $\text{blIV}$ and (2.1). The second inequality is (3.5). Thus

\[
\begin{align*}
\{\Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right)\}^r & \leq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right) \\
& \leq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right)^r,
\end{align*}
\]

so that

\[
\{\Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right)\}^{1/r} \leq \Phi\left(\left[\frac{\alpha_1}{\alpha_1}\right]^r, \ldots, \left[\frac{\alpha_n}{\alpha_1}\right]^r\right)^{1/r}.
\]

The theorem now follows from $\text{blII}$. \\

Theorem 3.5 can, of course, be specialized to yield a kind of converse to (c).

**Theorem 3.6.** If $\varphi$ is a unitarily invariant norm satisfying $\text{aiIV}$, then

\[(c^*) \quad [\varphi(\text{AA}^*)]^{1/2} \leq \varphi(A).
\]

Condition $(c^*)$ can also be obtained under somewhat different hypotheses. In particular, if $\varphi$ satisfies $\text{aiV}$, then

\[
\begin{align*}
\varphi(\text{AA}^*) &= \varphi(\text{AA}^*)\varphi((\text{AA}^*)^{-1}) \\
& \leq \varphi(A)\varphi(A^{-1})\varphi(A^*)\varphi(A^{-1}) = \varphi(A)\varphi(A^*).
\end{align*}
\]
If also \( \varphi(A) = \varphi(A^*) \), then \((c^*)\) follows. Of course, \( \varphi(A) = \varphi(A^*) \) if \( \varphi \) is unitarily invariant.

4. Independence of the norm axioms and \((c)\). It is our purpose here to show that the condition number inequality \((c)\) does not follow from the usual norm axioms \( aI - aV \). In fact, \( aI, aII, aIII, aIV, aV \) and \((c)\) are independent.

REMARK. It has been shown by Ostrowski [4] that \( aI \) is implied by \( aII, aIII, aV \), together with \( \varphi(A) \neq 0 \), so that \( aI \) is not included in the list of independent properties. Rella [5] has shown that \( aII, aIII, aIV \) and \( aV \) are independent, and we add \((c)\) to this list.

The results which prove the independence of \( aI - aV \) and \((c)\) are summarized in the following table, where \(+(-)\) indicates that a property is true (false).

<table>
<thead>
<tr>
<th>( \varphi(A) )</th>
<th>( aI )</th>
<th>( aII )</th>
<th>( aIII )</th>
<th>( aIV )</th>
<th>( aV )</th>
<th>( (c) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 1 )</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (\text{rank} \ A)(\text{tr} \ AA^*)^{1/2} )</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>( n \max</td>
<td>a_{ij}</td>
<td>)</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>( \max</td>
<td>a_{ij}</td>
<td>)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \Sigma</td>
<td>a_{ij}</td>
<td>)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

An example which serves in the last line of the table just as well as \( \Sigma |a_{ij}| \) is the norm \( \max_i \sum_j |a_{ij}| = \sup_x \Phi(xA)/\Phi(x) \), where \( \Phi(x) = \sum_i |x_i| \). Norms of this form are called "subordinate" or "lub" norms, and in this case \( \Phi \) is a symmetric guage function.

The remainder of this paper is devoted to proving the propositions indicated in the table.

The results for \( \varphi(A) = 1 \) are obvious, so we begin by considering \( \varphi(A) = (\text{rank} \ A)(\text{tr} \ AA^*)^{1/2} \). In this case, \( aII \) and \( aIV \) are obvious, and \((c)\) follows from Theorem 3.1, since \( (\text{tr} \ AA^*)^{1/2} \) is unitarily invariant. As is well known, \( (\text{tr} \ AA^*)^{1/2} \) satisfies \( aV \); this together with \( \text{rank} \ AB \leq (\text{rank} \ A)(\text{rank} \ B) \) yields \( aV \) for \( \varphi(A) = (\text{rank} \ A)(\text{tr} \ AA^*)^{1/2} \). That \( aIII \) is violated may be seen by taking \( A = I \) and \( B \) the matrix with a unit in the \((1,1)\)th place and zeros elsewhere.

For \( \varphi(A) = n \max_{i,j} |a_{ij}| \) and \( \max_{i,j} |a_{ij}| \) the first four columns of the table are well known, and we need only prove \((c)\). Let \( e_i \) be the row vector with one in the \( i \)th position and zero elsewhere. Denote \( M^{-1} = (m^{ij}) \) where \( M = (m_{ij}) \), and let \( U = AA^* \). By Cauchy’s inequality,
\[ |a_{ij}| |a^{\alpha\beta}| = |e_i A e_j^*| |e_\alpha A^{-1} e_\beta^*| \leq [(e_i U e_i^*) (e_\beta e_\beta^*) (e_\alpha U^{-1} e_\alpha^*)]^{1/2} \]
\[ = (u_{ii} u_{jj})^{1/2}. \]

Hence,
\[ \max_i |a_{ij}| \max_j |a^{\alpha\beta}| \leq (\max_i |u_{ii}| \max_j |u^{\alpha\beta}|)^{1/2}, \]
or
\[ c_\varphi(A) \leq [c_\varphi(A A^*)]^{1/2}. \]

Since \( U = A A^* \) is positive semi-definite,
\[ u_{ii} u_{jj} = (e_i U e_i^*) (e_j U^{-1} e_j^*) \geq (e_i e_i^*)^2 = 1, \]
and it follows that \( c_\varphi(A A^*) \geq 1 \). Thus, we have that
\[ (4.1) \quad c_\varphi(A) \leq [c_\varphi(A A^*)]^{1/2} \leq c_\varphi(A A^*), \]
which gives (c).

Note that the left inequality of (4.1) is a reversal of inequality (c*). That (4.1) also holds if \( \varphi(A) \) is the maximum of the absolute values of the characteristic values of \( A \) was proved by O. Taussky-Todd [6].

Since the first four columns of the table are well known for \( \varphi(A) = \sum |a_{ij}| \), we again need consider only (c). If \( A = \begin{pmatrix} B & 0 \\ 0 & 2I \end{pmatrix} \), where \( B = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \). Then (c) is violated. This same example shows that (c) is violated for \( \varphi(A) = \max_i \sum_j |a_{ij}|. \)

**Bibliography**

Donald Charles Benson, *Unimodular solutions of infinite systems of linear equations* ................................................................. 1
Richard Earl Block, *Transitive groups of collineations on certain designs* .............. 13
Dennison Robert Brown, *Topological semilattices on the two-cell* ......................... 35
Peter Southcott Bullen, *Some inequalities for symmetric means* ......................... 47
David Geoffrey Cantor, *On arithmetic properties of coefficients of rational functions* ........................................................................................................... 55
Luther Elic Claborn, *Dedekind domains and rings of quotients* ......................... 59
Allan Clark, *Homotopy commutativity and the Moore spectral sequence* ............. 65
Allen Devinatz, *The asymptotic nature of the solutions of certain linear systems of differential equations* ................................................................. 75
Robert E. Edwards, *Approximation by convolutions* ........................................ 85
Theodore William Gamelin, *Decomposition theorems for Fredholm operators* ....... 97
Edmond E. Granirer, *On the invariant mean on topological semigroups and on topological groups* ....................................................................... 107
Noel Justin Hicks, *Closed vector fields* .......................................................... 141
Charles Ray Hobby and Ronald Pyke, *Doubly stochastic operators obtained from positive operators* .............................................................................. 153
Robert Franklin Jolly, *Concerning periodic subadditive functions* ..................... 159
Tosio Kato, *Wave operators and unitary equivalence* ....................................... 171
Paul Katz and Ernst Gabor Straus, *Infinite sums in algebraic structures* ............. 181
Herbert Frederick Kreimer, Jr., *On an extension of the Picard-Vessiot theory* ........ 191
Radha Govinda Laha and Eugene Lukacs, *On a linear form whose distribution is identical with that of a monomial* ...................................................... 207
Donald A. Ludwig, *Singularities of superpositions of distributions* .................. 215
Albert W. Marshall and Ingram Olkin, *Norms and inequalities for condition numbers* ......................................................................................... 241
Horace Yomishi Mochizuki, *Finitistic global dimension for rings* .................... 249
Robert Harvey Oehmke and Reuben Sandler, *The collineation groups of division ring planes. II. Jordan division rings* ................................................. 259
George H. Orland, *On non-convex polyhedral surfaces in $E^3$* ....................... 267
Theodore G. Ostrom, *Collineation groups of semi-translation planes* .................. 273
Arthur Argyle Sagle, *On anti-commutative algebras and general Lie triple systems* ......................................................................................... 281
Laurent Siebenmann, *A characterization of free projective planes* .................... 293
Edward Silverman, *Simple areas* ........................................................................ 299
James McLean Sloss, *Chebyshev approximation to zero* .................................. 305
Robert S. Strichartz, *Isometric isomorphisms of measure algebras* ................. 315
Richard Joseph Turyn, *Character sums and difference sets* .............................. 319
L. E. Ward, *Concerning Koch’s theorem on the existence of arcs* .................... 347
Israel Zuckerman, *A new measure of a partial differential field extension* ........ 357