On matrices for which norm bounds are attained.

Hans Schneider*

Department of Mathematics
Van Vleck Hall
480 Lincoln Drive
University of Wisconsin-Madison
Madison WI 53706 USA

and

Hans F. Weinberger

School of Mathematics
University of Minnesota
107 Vincent Hall
206 Church Street
Minneapolis, MN, 55455 USA

January 1997

* Research supported by NSF Grant DMS-9424346

ABSTRACT

Let $\|A\|_{p,q}$ be the norm induced on the matrix A with n rows and m columns by the Hölder ℓ_p and ℓ_q norms on R^n and R^m (or C^n and C^m), respectively. It is easy to find an upper bound for the ratio $\|A\|_{r,s}/\|A\|_{p,q}$. In this paper we study the classes of matrices for which the upper bound is attained. We shall show that for fixed A, attainment of the bound depends only on the signs of $r-p$ and $s-q$. Various criteria depending on these signs are obtained. For the special case $p=q=2$, the set of all matrices for which the bound is attained is generated by means of singular value decompositions.

---

1. INTRODUCTION

Let A be a matrix with n rows and m columns. If A is considered as a complex transformation, let μ_1 and μ_2 be norms on C^m, and let ν_1 and ν_2 be norms on C^n. If A is real and is considered as a transformation from R^m to R^n, let the μ_i be norms on R^m and the ν_i be norms on R^n. Define the induced norms

$$\|A\|^{(i)} = \max_{\mathbf{x}} \nu_i(A\mathbf{x})/\mu_i(\mathbf{x})$$

for $i=1$ and 2, where the maximum is taken over either C^m or R^m, as is appropriate. It was shown in [SS] (see also [HJ, p.303]) that

$$\|A\|^{(2)} \le \max_{\mathbf{x}} \frac{\mu_1(\mathbf{x})}{\mu_2(\mathbf{x})} \ \max_{\mathbf{y}} \frac{\nu_2(\mathbf{y})}{\nu_1(\mathbf{y})} \ \|A\|^{(1)}, \tag{1.1}$$

and that equality is always attained for some $A \neq 0$. Here the maxima are taken over C^m and C^n if A is thought of as a complex transformation, and over R^m and R^n if A is a real and its action is confined to R^m.

In this work we shall be concerned with characterizing the set of all matrices A for which equality is attained in (1.1), at least in some cases.

We shall show that this set can be described by the following property.

Theorem 1. *If equality holds in the inequality (1.1), then every maximizer* $\mathbf{v}$ *of the ratio* $\nu_2(A\mathbf{x})/\mu_2(\mathbf{x})$ *has the properties that*

(i) $\mathbf{v}$ *is also a maximizer of the ratio* $\mu_1(\mathbf{x})/\mu_2(\mathbf{x})$,

(ii) $A\mathbf{v}$ *is a maximizer of the ratio* $\nu_2(\mathbf{y})/\nu_1(\mathbf{y})$, *and*

(iii) $\mathbf{v}$ *is also a maximizer of the ratio* $\nu_1(A\mathbf{x})/\mu_1(\mathbf{x})$.

Conversely, if there is one maximizer $\mathbf{v}$ *of* $\nu_1(A\mathbf{x})/\mu_1(\mathbf{x})$ *which has the properties (i) and (ii), then equality holds in (1.1).*

Theorem 1 can only provide useful information if the two maxima on the right and the corresponding maximizers are known. Both of these conditions apply when the norms involved are Hölder norms. We denote the ℓ_p norm by $\|\ \|_p$. For any p and q in the interval $[1,\infty]$ we define the induced norm

$$\|A\|_{p,q} := \max_{\mathbf{x}} \frac{\|A\mathbf{x}\|_q}{\|\mathbf{x}\|_p}. \tag{1.2}$$

The maximum of the ratio $\|\mathbf{x}\|_r/\|\mathbf{x}\|_p$ and the corresponding maximizers are well known. In order to state the result we recall that $\mathrm{sgn}(z)$ is defined to be 1 if $z > 0$, 0 if $z = 0$, and -1 if $z < 0$, and that $[z]_+$ is defined to be z if $z \geq 0$ and 0 if $z \leq 0$. We also define the three subsets of a real or complex vector space of m-tuples or n-tuples.

$$\begin{aligned} K_1 &= \{\mathbf{x} : \text{ all components of } \mathbf{x} \text{ have equal absolute values}\} \\ K_{-1} &= \{\mathbf{x} : \text{ at most one component of } \mathbf{x} \text{ differs from } 0\} \\ K_0 &= \text{the whole vector space.} \end{aligned} \tag{1.3}$$

The following result is found, e.g., in [HLP, p. 26 #16 and p. 29#19].

Proposition 1.

$$\|\mathbf{x}\|_r \leq m^{[(1/r)-(1/p)]_+} \|\mathbf{x}\|_p \qquad \text{for } p, r \in [1, \infty]. \tag{1.4}$$

Equality holds if and only if the m-vector $\mathbf{x}$ lies in $K_{\mathrm{sgn}(p-r)}$.

By inserting Proposition 1 into the inequality (1.1) and into Theorem 1, we immediately obtain the following special case for the Hölder spaces.

Proposition 2.

$$\|A\|_{r,s} \leq m^{[(1/p)-(1/r)]_+} n^{[(1/s)-(1/q)]_+} \|A\|_{p,q} \qquad \text{for } p, q, r, s \in [1, \infty]. \tag{1.5}$$

If equality holds in this inequality, then every maximizer $\mathbf{v}$ of the ratio $\|A\mathbf{x}\|_s/\|\mathbf{x}\|_r$ has the properties
(i) $\mathbf{v} \in K_{-\mathrm{sgn}(p-r)}$,
(ii) $A\mathbf{v} \in K_{\mathrm{sgn}(q-s)}$, *and*
(iii) $\mathbf{v}$ *is a maximizer of the ratio* $\|A\mathbf{x}\|_q/\|\mathbf{x}\|_p$.

Conversely, if there exists a maximizer $\mathbf{v}$ of the ratio $\|A\mathbf{x}\|_q/\|\mathbf{x}\|_p$ which has the properties (i) and (ii), then equality holds in (1.5).

For the case $m = n$, $q = p$, $s = r$, the inequality (1.5) was pointed out by Higham [H, p.124].

The inequality (1.5) simply states that for fixed s, $\|A\|_{r,s}$ is nondecreasing and $n^{1/r}\|A\|_{r,s}$ is nonincreasing in r, and that for fixed r, $n^{-1/s}\|A\|_{r,s}$ is nondecreasing and $\|A\|_{r,s}$ is nonincreasing in s.

The trivial observation that for fixed (p, q) the only dependence on (r, s) in Proposition 2 is through the functions $\mathrm{sgn}(p - r)$ and $\mathrm{sgn}(q - s)$ immediately yields the following statement.

Proposition 3. *If equality holds in (1.5) , if sgn($p - r'$)=sgn($p - r$), and if sgn($q - s'$)=sgn($q - s$), then equality also holds in (1.5) when the pair (r, s) is replaced by (r', s').*

Remark. *By using the inequality (1.5) with $r = p'$ and $s = q'$, one sees that Proposition 2 also shows that equality in (1.5) implies that the same equality when (p, q, r, s) is replaced by (p', q', r', s'), provided* $\mathrm{sgn}(p' - r') = \mathrm{sgn}(p - p') = \mathrm{sgn}(p - r)$ *and* $\mathrm{sgn}(q' - s') = \mathrm{sgn}(q - q') = \mathrm{sgn}(q - s)$.

When $p = q = 2$, Proposition 2 enables us to give a characterization of all matrices for which equality holds in the bound (1.5). As usual, we denote the Hermitian transpose of a matrix A by A^*.

Theorem 2. *If $r, s \in [1, \infty]$, the equality*

$$\|A\|_{r,s} = m^{[(1/2)-(1/r)]_+}\, n^{[(1/s)-(1/2)]_+}\, \|A\|_{2,2}.$$

is valid if and only if A has a singular value decomposition

$$A = U\Sigma V^*$$

in which

(i) the first column of the unitary matrix U is in $K_{\mathrm{sgn}(2-s)}$,
(ii) the first column of the unitary matrix V is in $K_{-\mathrm{sgn}(2-r)}$, and
(iii) the (11) entry of the nonnegative diagonal matrix Σ is its maximal entry.

The first two theorems will be proved in Section 2.

When p and q are not both 2, Proposition 2 will still help to characterize those matrices for which equality holds in the bound (1.5). Because of Proposition 3, the results will only depend on the relative sizes of p and r and of q and s. Our most complete characterization is for the case in which $r < p$ and $s > q$, which is treated in Section 3.

Theorem 3. *Let ρ denote the largest absolute value of the entries of A, so that $\rho = \|A\|_{1,\infty}$.*

If $1 < p \le q < \infty$, then equality holds in (1.5) for some (and hence every) $r < p$ and $s > q$, if and only if A has the properties

(i) every entry of A which has the absolute value ρ is the only nonzero element of its row and of its column, and
(ii) if C is the matrix obtained from A by replacing all elements of absolute value ρ by zero, then $\|C\|_{p,q} \le \rho$.

If $p > q$, then equality holds in (1.5) for $r < p$ and $s > q$ if and only if A has at most one nonzero entry.

Theorem 3′ in Section 3 shows that the Property (i) is sufficient for the existence of a $p > 1$ and a $q < \infty$ such that equality holds in (1.5) for all $r < p$ and $s > q$ when $p > 1$ is sufficiently small and $q < \infty$ is sufficiently large.

Section 4 deals with the cases in which $r < p$ and $s < q$ or $r > p$ and $s > q$. We shall establish the following results.

Theorem 4. *Let σ denote the largest ℓ_1 norm of the columns of A, so that $\sigma = \|A\|_{1,1}$.*

If equality holds in (1.5) for some $r < p$ and $s < q$, then A has the properties

(i) the entries of any column whose ℓ_1 norm is equal to σ all have the same absolute value $n^{-1}\sigma$,
(ii) every column with this property is orthogonal to all the other columns of A, and
(iii) $\sigma = n^{1-(1/q)}\|A\|_{p,q}$.

Conversely, if the matrix A has a column all of whose entries have the absolute values $n^{-1/q}\|A\|_{p,q}$, then equality holds in (1.5) for all $r < p$ and $s < q$.

If $p > 2$, then equality holds in (1.5) for $r < p$ and $s < q$ if and only if A has only one nonzero column, and all the entries of this column have the same absolute value.

Theorem 5. *Let σ denote the largest ℓ_1 norm of the rows of A, so that $\sigma = \|A\|_{\infty,\infty}$. If equality holds in (1.5) for some $r > p$ and $s > q$, then A has the properties* *(i)* *the entries of any row whose* ℓ_1 *norm is equal to* σ *all have the same absolute value* $m^{-1}\sigma$,

(ii) *every row with this property is orthogonal to all the other rows of* A, *and*

(iii) $\sigma = m^{1/p}\|A\|_{p,q}$.

Conversely, if the matrix A *has a rows all of whose entries have the absolute values* $m^{(1/p)-1}\|A\|_{p,q}$, *then equality holds in (1.5) for all* $r > p$ *and* $s > q$.

If $q < 2$, *then equality holds in (1.5) for all* $r > p$ *and* $s > q$ *if and only if* A *has only one nonzero row, and all the entries of this row have the same absolute value.*

Theorem 4′ in Section 4 shows that the Properties (i) and (ii) of Theorem 4 are sufficient for the existence of $p > 1$ and $q > 1$ such that equality holds in (1.5) whenever $r < p$ and $s < q$. Theorem 5' gives the analogous result for $r > p$, $s > q$.

Section 5 considers the case where $r > p$ and $s < q$. The following result is obtained.

Theorem 6. *Equality holds in (1.5) for* $r > p$ *and* $s < q$ *if and only if there is a vector* $\mathbf{v}$ *with the properties*

(i) $\mathbf{v}$ *is an eigenvector of the matrix* A^*A,

(ii) *all the entries of* $\mathbf{v}$ *have the absolute value 1,*

(iii) *all the entries of* $A\mathbf{v}$ *have the same absolute value* τ, *and*

(iv) $\tau = m^{1/p}n^{-1/q}\|A\|_{p,q}$.

In particular, equality holds in (1.5) when $p = q = 2$, $r > 2$, *and* $s < 2$ *if and only if* A^*A *has an eigenvector* $\mathbf{v}$ *with the properties (ii) and (iii) which corresponds to its largest eigenvalue.*

We observe that when the matrix A is real, one has a choice of defining the induced norm $\|A\|_{p,q}$ with respect to either the real or the complex Hölder spaces, and that these two norms may differ for some (p, q). Our results are valid for either choice.

Consider, for instance, the matrix $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. The last statement of Theorem 6 with the complex eigenvector $(1,\ i)$ of $A^*A = 2I$, shows that when $r \geq 2 \geq s$ the norm $\|A\|_{r,s}$ on the complex vector space C^2 is equal to $2^{(1/s)-(1/r)+(1/2)}$. On the other hand, a simple computation shows that on the real vector spaces, $\|A\|_{\infty,1} = 2$ while $\|A\|_{2,2}$ is still $2^{1/2}$. Thus in the real norm, equality does not hold in (1.5) when $p = q = 2$, $r > 2$, and $s < 2$. Therefore the real norm $\|A\|_{r,s}$ is strictly less than $2^{(1/s)-(1/r)+(1/2)}$, and hence less than the complex norm, when $r > 2$ and $s < 2$.

2. PROOFS OF THEOREMS 1 AND 2.

We begin by proving Theorem 1.

Proof of Theorem 1. We recall the derivation in [SS] of the inequality (1.1). For any $\mathbf{x} \neq \mathbf{0}$ with $A\mathbf{x} \neq 0$ we have

$$\frac{\nu_2(A\mathbf{x})}{\mu_2(\mathbf{x})} = \frac{\mu_1(\mathbf{x})}{\mu_2(\mathbf{x})}\frac{\nu_2(A\mathbf{x})}{\nu_1(A\mathbf{x})}\frac{\nu_1(A\mathbf{x})}{\mu_1(\mathbf{x})}. \tag{2.1}$$

Because the maximum of a product is bounded by the product of the maxima, we obtain the inequality (1.1).

Suppose there is a maximizer of the left-hand side of (2.1), which is not a maximizer of one of the factors on the right. Since all the factors are bounded by their maxima and one of them is strictly less than its maximum, the right-hand side of (1.1) is strictly greater than the left-hand side. Therefore the condition of Proposition 1 is necessary for equality.

If there is a maximizer $\mathbf{v}$ of all three quotients on the right of (2.1), then the maximum of the left-hand side is bounded below by the right-hand side of (1.1). Since we already know that it is bounded above by the same quantity, we conclude that equality holds in (1.1). This establishes Theorem 1.

Proof of Theorem 2. We observe that a maximizer of the ratio $\|A\mathbf{v}\|_2/\|\mathbf{v}\|_2$ is an eigenvector of the matrix A^*A which corresponds to its largest eigenvalue. By Proposition 2, equality in (1.5) with $p = q = 2$ implies that a maximizer $\mathbf{v}$ of $\|A\mathbf{x}\|_r/\|\mathbf{x}\|_q$ is such an eigenvector, that it is in $K_{-\mathrm{sgn}(2-r)}$, and that the eigenvector $A\mathbf{v}$ of AA^* is in $K_{\mathrm{sgn}(2-s)}$.

Thus we can construct (see. e.g., the proof of Theorem 2.3-1 in [GvL]) a singular value decomposition $A = U\Sigma V^*$ in which the first column of the unitary matrix U is the vector $\|A\mathbf{v}\|_2^{-1}A\mathbf{v} \in K_{\mathrm{sgn}(2-s)}$ and the first row of the unitary matrix V is $\|\mathbf{v}\|_2^{-1}\mathbf{v} \in K_{-\mathrm{sgn}(2-r)}$. The (11) element of the nonnegative diagonal matrix Σ is the square root of the largest eigenvalue of A^*A, which is the maximal element of Σ.

The converse follows from the fact that the first column of V is a maximizer of $\|A\mathbf{x}\|_2/\|\mathbf{x}\|_2$ and the converse statement of Proposition 2, so that Theorem 2 is proved.

Remark. *If the matrix A is a scalar multiple of a unitary matrix and the absolute values of all its entries are equal to a number ρ, then A has a singular value decomposition with $U = n^{-1/2}\rho^{-1}A$, $\Sigma = n^{1/2}\rho I$, and $V = I$, and another singular value decomposition with $U = I$, $\Sigma = n^{1/2}\rho I$, and $V = n^{-1/2}\rho^{-1}A^*$. Hence Theorem 2 shows that equality holds in (1.5) whenever $p = q = 2$ and r and s are either both bounded above by 2 or bounded below by 2.*

Examples of such matrices include the Hadamard matrices, which are orthogonal matrices whose entries have the values ± 1 (see [H, p. 128, §6.13]), and the matrices which represent the finite Fourier transforms.

3. THE CASE $r < p$, $s > q$.

The following lemma will be used in the proofs of Theorems 3, 4, and 6. We recall the definition of the conjugate index $p^* = p/(p-1)$ of an index p, and the fact that A^* denotes the Hermitian transpose of the matrix A.

We also recall the identity

$$\|A^*\|_{q^*,p^*} = \|A\|_{p,q}, \tag{3.1}$$

which simply states that the norm of the adjoint of a transformation is equal to the norm of the transformation.

Lemma 3.1. *Suppose that a maximizer $\mathbf{v}$ of $\|A\mathbf{x}\|_q/\|\mathbf{x}\|p$ has the properties that*

(i) all its nonzero components have the same absolute value, and

(ii) the same is true of $A\mathbf{v}$.

If $1 < p < \infty$, or $p = 1$ and $\mathbf{v} \in K_1$, or $p = \infty$ and $\mathbf{v} \in K_{-1}$, then $\mathbf{v}$ is an eigenvector of the matrix A^*A.

Proof. Because of the duality relation (3.1), we have

$$\begin{aligned}\|A^*A\mathbf{v}\|_{p^*} &\le \|A\|_{p,q}\|A\mathbf{v}\|_{q^*}\\ &= (\|A\mathbf{v}\|_q/\|\mathbf{v}\|_p)\|A\mathbf{v}\|_{q^*}.\end{aligned}$$

It is easily seen from the property (ii) that

$$\|A\mathbf{v}\|_2^2 = \|A\mathbf{v}\|_q\|A\mathbf{v}\|_{q^*}.$$

Therefore

$$\mathbf{v}\cdot A^*A\mathbf{v} = \|A\mathbf{v}\|_q\|A\mathbf{v}\|_{q^*} \ge \|\mathbf{v}\|_p\|A^*A\mathbf{v}\|_{p^*}.$$

This shows that equality holds in the Hölder inequality for the bilinear form $\mathbf{v}\cdot A^*A\mathbf{v}$ in $\ell_p \times \ell_{p^*}$. If $1 < p < \infty$, this implies that the vector $A^*A\mathbf{v}$ must be a multiple of the vector with components $\|\mathbf{v}\|_p^{p-2}v_j$. (See, e.g., [HLP p. 26#14].) By Property (i) this vector is a multiple of $\mathbf{v}$, which proves the result for this case.

If $p = 1$ so that $p*^= \infty$, and if $\mathbf{v}$ has no zero component, it is easily seen that equality in Hölder's inequality implies that $A^*A\mathbf{v}$ is proportional to the vector with components $|v_j|^{-1}v_j$, and we reach the same conclusion. This is the case when $p = 1$ and $\mathbf{v} \in K_1$.

If $p = \infty$ so that $p^* = 1$, one easily sees that equality in the Hölder inequality implies that $A^*A\mathbf{v}$ has zero components where $\mathbf{v}$ does. Therefore,if $\mathbf{v} \in K_{-1}$ so that it has only one nonzero component, $A^*A\mathbf{v}$ is again proportional to $\mathbf{v}$.

Thus the Lemma is proved in all cases.

Proof of Theorem 3. Suppose that equality holds in (1.5) for some $r < p$ and $s > q$. By Proposition 3, equality holds for all r and s which satisfy this inequality, and in particular for $r = 1$ and $s = \infty$. It is easily verified that $\|A\mathbf{x}\|_\infty/\|\mathbf{x}\|_1 \le \rho$, the largest absolute value of any entry of A, and that that this bound is attained when $\mathbf{x}$ is in the direction of a coordinate which corresponds to a column in which an element of magnitudeρ occurs. Thus a unit vector $\mathbf{v}$ in such a coordinate direction is a maximizer of the ratio.

Thus Proposition 2 shows that if $\mathbf{v}$ is a unit vector in the direction of such a column, the column $A\mathbf{v}$ has exactly one nonzero element, and $\mathbf{v}$ is a maximizer of the ratio $\|A\mathbf{x}\|_q/\|\mathbf{x}\|_p$. The first of these properties says that any column of A which contains an element of magnitude ρ has but one nonzero element, while the second property implies that the absolute value ρ of the nonzero element equals $\|A\|_{p,q} = \|A\|_{1,\infty}$. There may, of course, be several maximizers, and therefore several columns with singleton elements of magnitude ρ.

Since $\mathbf{v}$ and $A\mathbf{v}$ are both in coordinate directions and $p > 1$, we can apply Lemma 3.2 to show that $\mathbf{v}$ is an eigenvector of A^*A. Therefore if $\mathbf{x}$ is a coordinate vector orthogonal to $\mathbf{v}$, it is also orthogonal to $A^*A\mathbf{v}$, which implies that $A\mathbf{x}$ is orthogonal to $A\mathbf{v}$. This means that a column which contains a single nonzero element of magnitude ρ is orthogonal to all the other columns of A. In other words, an element of magnitude ρ is also the only nonzero element of its row as well as of its column, so that Property (i) is established.

If we choose a trial vector $\mathbf{x}$ whose components in the directions of the columns with elements of magnitude ρ are zero, then $A\mathbf{x} = C\mathbf{x}$ where C is defined in the statement of Theorem 1. Therefore $\|C\|_{p,q} \le \|A\|_{p,q} = \rho$. This is Property (ii).

To prove the converse statement for $p \le q$, we define $B = A - C$, and decompose any vector $\mathbf{x}$ into $\mathbf{y} + \mathbf{z}$, where the components of $\mathbf{z}$ are zero in the directions corresponding to columns which contain elements of magnitude ρ and the components of $\mathbf{y}$ in the remaining directions vanish. Then by Property (i) and two applications of Proposition 1

$$\begin{aligned}\|A\mathbf{x}\|_q &= \{(\rho\|\mathbf{y}\|_q)^q + \|C\mathbf{z}\|_q^q\}^{1/q} \\ &\le \{(\rho\|\mathbf{y}\|_p)^q + (\|C\|_{p,q}\|\mathbf{z}\|_p)^q\}^{1/q} \\ &\le \{(\rho\|\mathbf{y}\|_p)^p + (\|C\|_{p,q}\|\mathbf{z}\|_p)^p\}^{1/p} \\ &\le \max\{\rho, \|C\|_{p,q}\}\|\mathbf{x}\|_p.\end{aligned}$$

That is,

$$\|A\|_{p,q} = \max\{\rho, \|C\|_{p,q}\}. \tag{3.2}$$

Thus Property (ii) shows that $\|A\|_{p,q} = \rho = \|A\|_{1,\infty}$, and the proof of the converse statement is complete.

To prove the last assertion of Theorem 3 assume that $p > q$ and that equality holds in (1.5) for $r < p$ and $s > q$. Choose a trial vector $\mathbf{x}$ whose component in the direction a column with a singleton element of magnitude ρ is one and which has one other nonzero component α. Let b be any entry of A in the column which corresponds to α. Then because $p > q$,

$$\frac{\|A\mathbf{x}\|_q}{\|\mathbf{x}\|_p} \ge \frac{(\rho^q + |\alpha b|^q)^{1/q}}{(1 + |\alpha|^p)^{1/p}} = \rho + (1/p)\rho^{1-q}|b|^q|\alpha|^q + o(|\alpha|^q)$$

for small α. Because $\rho = \|A\|_{p,q}$, the right-hand side must be bounded by ρ, and we conclude that $b = 0$. Because b is an arbitrary element of any column other that that with the entry of magnitude ρ, we conclude that all other columns of A are zero, so that A has only one nonzero entry.

Finally, a simple computation show that if A has only one nonzero entry, and if the magnitude of this entry is ρ, then $\|A\|_{r,s} = \rho$ for all r and s, so that equality holds for all p, q, r, and s.

Thus all parts of Theorem 3 have been established.

Because it is difficult to compute the p, q norm for most p and q, it is difficult to verify Property (ii) of Theorem 3. We shall show that the easily verified Property (i) is sufficient to assure the existence of some $p > 1$ and $s < \infty$ such that equality holds in (1.5) when $r < p$ and $s > q$.

Theorem 3′. *Let A have Property (i) of Theorem 3. Let C be the matrix obtained from A by replacing all elements of absolute value $\rho = \|A\|_{1,\infty}$ by 0, so that $\|C\|_{1,\infty} < \rho$. If p and q satisfy the inequalities $p \le q$ and*

$$m^{1-(1/p)} n^{(1/q)} \|C\|_{1,\infty} \le \rho, \tag{3.2}$$

then equality holds in (1.5) for all $r < p$ and $s > q$. The inequality (3.2) is satisfied if p is sufficiently close to 1 and q is sufficiently large.

Proof. Since (1.5) shows that

$$\|C\|_{p,q} \leq m^{1-(1/p)} n^{1/p} \|C\|_{1,\infty},$$

the inequality (3.2) and the equation (3.2) imply that $\|A\|_{p,p} = \rho$. That is, Property (ii) of Theorem 3 holds, the the equality follows.

4. THE CASES $r < p,\ s < q$ AND $r > p,\ s > q$.

Proof of Theorem 4. Proposition 3 shows that if equality holds in (1.5) for some $r < p$ and $s < p$, it holds for all such r and s, and in particular for $r = s = 1$. The triangle inequality shows that $\|A\mathbf{x}\|_1/\|\mathbf{x}\|_1 \leq \sigma$, the largest ℓ_1 norm of the columns of A. Moreover, this bound is attained when $\mathbf{x}$ is in the direction of any coordinate whose corresponding column has the ℓ_1 norm σ. Thus if $\mathbf{v}$ is a coordinate vector in such a direction, it is a maximizer for the ratio.

Proposition 2 states that if $\mathbf{v}$ is a unit vector in one of these coordinate directions, the elements of the corresponding column $A\mathbf{v}$ must have equal absolute values, and $\mathbf{v}$ must also be a maximizer of $\|A\mathbf{x}\|_q/\|\mathbf{x}\|_p$. These two facts give the properties (i) and (iii) of Theorem 3.

Since $p > 1$, Lemma 3.1 shows that $\mathbf{v}$ is an eigenvector of A^*A. As in the proof of Theorem 3, this implies that if $\mathbf{x}$ is a coordinate vector perpendicular to $\mathbf{v}$, then it is also perpendicular to $A^*A\mathbf{v}$, so that the column $A\mathbf{x}$ is perpendicular to $A\mathbf{v}$. This is the property (ii)

To prove the converse statement, we observe that if A has a column whose elements have the absolute value $n^{-1/q}\|A\|_{p,q}$, then a unit vector $\mathbf{v}$ in the direction of this column is a maximizer of the ratio $\|A\mathbf{x}\|_q/\|bfx\|_p$. Therefore the converse statement of Proposition 2 implies that equality holds in (1.5).

To prove the last statement of Theorem 3, we suppose that equality holds in (1.5) for $r < p$ and $s < q$. Then there is at least one column $\mathbf{c}$ of A all of whose entries have the absolute value $n^{-1}\sigma = n^{-1/q}\|A\|_{p,q}$. Let $\mathbf{c}$ be one such column, let $\mathbf{b}$ be any other column of A, and let α be a real parameter. The adjoint relation (3.1) leads to the inequality

$$\|A^*(\mathbf{c} + \alpha\mathbf{b})\|_{p^*} \leq \|A\|_{p,q}\|\mathbf{c} + \alpha\mathbf{b}\|_{q^*} = n^{-1+(1/q)}\sigma\|\mathbf{c} + \alpha\mathbf{b}\|_{q^*}. \tag{4.1}$$

We observe that for small α

$$|c_j + \alpha b_j|^{q^*} = |c_j|^{q^*} + q^* \operatorname{Re}(\alpha|c_j|^{q^*-2}\bar{c}_j b_j) + O(\alpha^2).$$

We sum on j and use the properties that the entries of $\mathbf{c}$ all have the absolute value $n^{-1}\sigma$ and that $\mathbf{b}$ is orthogonal to $\mathbf{c}$ to find that

$$\|\mathbf{c} + \alpha\mathbf{b}\|_{q^*}^{q^*} = n^{-1+(1/q^*)}\sigma + O(\alpha^2) = n^{-1/q}\sigma + O(\alpha^2). \tag{4.2}$$

Since $\mathbf{c}$ and $\mathbf{b}$ are orthogonal, the entry of $A^*(\mathbf{c}+\alpha\mathbf{b})$ which corresponds to the column $\mathbf{c}$ is $n^{-1}\sigma^2$, while the entry which corresponds to the column $\mathbf{b}$ is $\alpha\|\mathbf{b}\|_2^2$. We obtain a lower bound for the left-hand side of (4.1) by replacing all the other entries by zero. For small α this lower bound takes the form

$$\|A^*(\mathbf{c}+\alpha\mathbf{b})\|_{p^*} \geq n^{-1}\sigma^2 + (p^*)^{-1}(n^{-1}\sigma^2)^{1-p^*}\|\mathbf{b}\|_2^{2p^*}\alpha^{p^*} + O(\alpha^{2p^*}).$$

where $K > 0$.

By putting this and (4.2) into (4.1), we find the inequality

$$n^{-1}\sigma^2 + (p^*)^{-1}(n^{-1}\sigma^2)^{1-p^*}\|\mathbf{b}\|_2^{2p^*}\alpha^{p^*} + O(\alpha^{2p^*}) \leq n^{-1}\sigma^2 + O(\alpha^2).$$

We observe that $p^* < 2$ because $p > 2$. We cancel the first terms from the two sides, divide by α^{p^*}, and let α approach zero to see that $\|\mathbf{b}\|_2 = 0$. That is, every column other than $\mathbf{c}$ is zero. This establishes the last statement of Theorem 4, and the Theorem is proved.

Theorem 5 will follow easily from Theorem 4 and the following lemma.

Lemma 4.1. *If equality holds in (1.5), then equality also holds when A is replaced by A^*, the pair (r, s) is replaced by (s^*, r^*) and the pair (p, q) is replaced by (q^*, p^*).*

Proof. We recall the adjoint equation (3.1), namely $\|A^*\|_{q^*,p^*} = \|A\|_{p,q}$ We also note that in going from A to A^* the dimensions m and n are interchanged, and that by definition $(1/q^*) - (1/s^*) = (1/s) - (1/q)$ and $(1/r^*) - (1/p^*) = (1/p) - (1/r)$. Therefore, the replacements indicated in the Lemma leave both sides of (1.5) unchanged, which proves the Lemma.

Proof of Theorem 5. By Lemma 4.1, equality in (1.5) is equivalent to the equality when A is replaced by A^*, (p, q) is replaced by (q^*, p^*), and (r, s) is replaced by (s^*, r^*). Since $s > q$ implies $s^* < q^*$ and $r > p$ implies $r^* < p^*$, the application of Theorem 4 to A^* with the above index replacements gives the statement of Theorem 5.

As in the case of Theorem 3, it is difficult to verify the last hypothesis of Theorems 4 and 5. We shall prove that the easily verified Properties (i) and (ii) are sufficient to assure the existence of $p, q \in (1, \infty]$ such that equality holds in (1.5) when $r < p$ and $s < q$.

Theorem 4′. *Let A have the Properties (i) and (ii) of Theorem 3. Let C be the matrix obtained from A by replacing all columns with the ℓ_1 norm $\sigma = \|A\|_{1,1}$ by zero, so that $\|C\|_{1,1} < \sigma$. If p satisfies the inequality*

$$\begin{aligned}(4.3)\qquad & (2mn)^{1-(1/p)}\sigma^{-1}\|C\|_1^0 + [2^{1-(1/p)} - 1] \\ & \cdot [(p/2)^{1/(2-p)} n^{(-3p^2+2p+4)/[2p(2-p)]} m^{-2(p-1)/(2-p)} (\|C\|_{1,1}/\sigma)^{2/(2-p)}] \leq 1,\end{aligned}$$

and $q \leq p$, then equality holds in (1.5) for $r < p$ and $s < q$. The inequality (5.3) is satisfied when p is sufficiently near 1.

Proof of Theorem 4′. We recall that C is the matrix obtained from A by replacing those columns whose ℓ_1 norm is σ by 0. Thus $\|C\|_{1,1} < \sigma$. Let $B = A - C$, so that all the nonzero elements of B have the magnitude $n^{-1}\sigma$, and every column of B is orthogonal to all other columns of A. To establish the Theorem, we only need to show that the inequality (4.3) implies that $\|A\|_{p,p} = n^{-1+(1/p)}\sigma = n^{-1+(1/p)}\|A\|_{1,1}$.

Decompose an arbitrary vector $\mathbf{x} \neq 0$ into $\mathbf{x} = \mathbf{y} + \mathbf{z}$, where $\mathbf{z}$ is obtained from $\mathbf{x}$ by replacing those elements which correspond to the nonzero columns of B by zero, and $\mathbf{y} = \mathbf{x} - \mathbf{z}$.

We see from the conditions (i) and (ii) of Theorem 4 that for the above decomposition $\mathbf{x} = \mathbf{y} + \mathbf{z}$,

$$\|B\mathbf{x}\|_2^2 = n^{-1}\sigma^2 \|\mathbf{y}\|_2^2. \tag{4.4}$$

In particular, $\|B\|_{2,2} = n^{-1/2}\sigma$, so that B satisfies the conditions of Theorem 3 with $p = q = 2$. Therefore,

$$\|B\|_{r,s} = n^{-1+(1/s)}\sigma \tag{4.5}$$

for all r and s in the interval [1,2]. On the other hand, the inequality (1.5) shows that

$$\|C\|_{r,s} \le m^{1-(1/r)} \|C\|_{1,1}. \tag{4.6}$$

for $r, s \ge 1$. Therefore by the triangle inequality

$$\|A\mathbf{x}\|_p \le n^{-1+(1/p)}\sigma \|\mathbf{y}\|_p + m^{1-(1/p)} \|C\|_{1,1} \|\mathbf{z}\|_p. \tag{4.7}$$

Proposition 2 shows that

$$\|\mathbf{x}\|_p = (\|\mathbf{y}\|_p^p + \|\mathbf{z}\|_p^p)^{1/p} \ge 2^{-1+(1/p)} (\|\mathbf{y}\|_p + \|\mathbf{z}\|_p). \tag{4.8}$$

We see from (4.7) and (4.8) that

$$\frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \le n^{-1+(1/p)}\sigma \tag{4.9}$$

whenever

$$\|\mathbf{y}\|_p \le \frac{1 - (2mn)^{1-(1/p)}\sigma^{-1}\|C\|_1^0}{2^{1-(1/p)} - 1} \|\mathbf{z}\|_p. \tag{4.10}$$

Thus the bound (4.4) is valid when the ratio $\|\mathbf{y}\|_p / \|\mathbf{z}\|_p$ is not too large. To obtain this bound for larger values of this ratio, we note that

$$\|A\mathbf{x}\|_p^p = \sum_{j=1}^{n} (|(B\mathbf{y})_j|^2 + 2\mathrm{Re}[\overline{(B\mathbf{y})_j} (C\mathbf{z})_j] + |(C\mathbf{z})_j|^2)^{p/2}. \tag{4.11}$$

Because $p \le 2$, the function $w^{p/2}$ is concave, so that for any positive d and w

$$w^{p/2} \le d^{p/2} + (p/2) d^{(p/2)-1} (w - d).$$

We apply this inequality with $d = n^{-1}\|B\mathbf{y}\|_2^2$ to each term of the sum on the right of (4.11) and use the fact that the range of C is orthogonal to the range of B to see that

$$\|A\mathbf{x}\|_p^p \le n^{1-(p/2)} \|B\mathbf{y}\|_2^p + (p/2) n^{1-(p/2)} \|B\mathbf{y}\|_2^{p-2} \|C\mathbf{z}\|_2^2. \tag{4.12}$$

The equation (4.5) shows that the first term on the right is bounded by $n^{-1+(1/p)}(\sigma\|\mathbf{y}\|_p)^p$. Therefore we see that the inequality (4.9) is valid when

$$(p/2)n^{1-(p/2)}\|B\mathbf{y}\|_2^{p-2}\|C\mathbf{z}\|_2^2 \le n^{p-1}\sigma^p\|z\|_p^p. \tag{4.13}$$

We see from (4.6) that $\|C\mathbf{z}\|_2 \le m^{1-(1/p)}\|C\|_{1,1}\|\mathbf{z}\|_p$, and from (4.4) and (1.4) that

$$\|B\mathbf{y}\|_2 = n^{-1/2}\sigma\|\mathbf{y}\|_2 \ge n^{-1/p}\sigma\|\mathbf{y}\|_p.$$

Therefore the inequality (4.13), and hence also (4.9), is implied by

$$\|\mathbf{y}\|_p \ge (p/2)^{1/(2-p)}n^{(-3p^2+2p+4)/[2p(2-p)]}m^{-2(p-1)/(2-p)}(\|C\|_{1,1}/\sigma)^{2/(2-p)}\|\mathbf{z}\|_p. \tag{4.14}$$

We now observe that the inequality (4.3) states that the coefficient on the right of (4.14) is no larger than that in (4.10). Therefore at least one of these inequalities inequalities is satisfied for every $\mathbf{y}$ and $\mathbf{z}$. That is, the inequality (4.9) holds for all $\mathbf{x}$, so that $\|A\|_{p,p} \le n^{-1+1/p}\sigma = n^{-1+1/p}\|A\|_{1,1}$. Because (1.5) gives the inequality in the opposite direction, we conclude that equality holds in (1.5) for $r < p$ and $s < q$. Thus Theorem $4'$ is established.

By using Lemma 4.1 and applying Theorem $4'$ to A^*, we obtain the analogous result.

Theorem $5'$. *Let A have the Properties (i) and (ii) of Theorem 5. Let C be the matrix obtained from A by replacing all rows with the ℓ_1 norm $\sigma = \|A\|_{1,1}$ by zero, so that $\|C\|_{1,1} < \sigma$. If q satisfies the inequality*

$$\begin{aligned}(2mn)^{1/q}\sigma^{-1}\|C\|_1 + [2^{1/q} - 1]\\ \cdot [(q^*/2)^{1/(2-q^*)}n^{(-3(q^*)^2+2q^*+4)/[2q^*(2-q^*)]}m^{-2(q^*-1)/(2q^*p)}(\|C\|_{1,1}/\sigma)^{2/(2-q^*)}] \le 1,\end{aligned} \tag{4.15}$$

and $p \le q$, then equality holds in (1.5). The inequality (4.15) is satisfied when p is sufficiently near 1.

5. THE CASE $r > p,\ s < q$.

Proof of Theorem 6. Suppose that Equality holds in (1.5) with $r > p$ and $s < q$. Proposition 2 shows that every maximizing vector $\mathbf{v}$ of the ratio $\|A\mathbf{x}\|_1/\|\mathbf{x}\|_\infty$ has the properties (ii) its components have equal absolute values, which we normalize to 1; (iii) the components of $A\mathbf{v}$ have equal absolute values, which we call τ; and (iv) $\|A\mathbf{v}\|_q/\|\mathbf{v}\|_p = \|A\|_{p,q}$. Because $p < \infty$, Lemma 3.1 shows that $\mathbf{v}$ is an eigenvector of A^*A, which is Property (i). Thus the first part of Theorem 6 is proved.

On the other hand, a vector $\mathbf{v}$ with the properties (ii), (iii), and (iv) is a maximizer of the ratio $\|A\mathbf{x}\|_q/\|\mathbf{x}\|_p$, so that Proposition 2 also establishes the converse statement.

The last statement of Theorem 6 clearly follows from the rest when $p = q = 2$, so the Theorem is proved.

We are unable to find an analog of Theorems $3'$, $4'$, and $5'$ for this case. We confine ourselves to the following simple observations.

1. If we define V to be the diagonal unitary matrix whose diagonal entries are the components of $\mathbf{v}$ and D to be the diagonal, unitary matrix whose diagonal entries are the components of the vector $\tau^{-1}\overline{A\mathbf{v}}$, the conditions of Theorem 4 imply that all the row sums of the matrix DAV are τ and that all its column sums are $m^{-1}n\tau$. Conversely, if one can find two matrices D and V with these properties, then the vector $\mathbf{v}$ whose components are the diagonal entries of V has the properties (i), (ii), and (iii) of Theorem 4. Thus equality holds in (1.5) for $r > p$ and $s < q$ if and only if there are matrices D and V with these properties and $\tau = m^{1/p}n^{-1/q}|A\|_{p,q}$.

2. A sufficient condition for equality to hold in (1.5) when $p = q = 2$, $r > 2$, and $s < 2$ is that there exist diagonal unitary matrices D and V such that the matrix DAV has nonnegative entries, equal row sums, and equal column sums. When $m = n$, DAV is a multiple of a doubly stochastic matrix.

3. The matrices with a single nonzero element which occur in the last statement of Theorem 3 can be thought of as the tensor product of two vectors in K_{-1}. Similarly, the matrices in the last statements of Theorems 4 and 5 are tensor products. It is easily verified that if $A = \mathbf{c} \otimes \mathbf{b}$ so that its entries have the form $c_i b_j$, then $\|A\|_{r,s} = \|\mathbf{b}\|_{r^*}\|\mathbf{c}\|_s$. Then Proposition 1 shows that when $A = \mathbf{c} \otimes \mathbf{b}$, equality holds in (1.5) if and only if $\mathbf{a} \in K_{-\mathrm{sgn}(p-r)}$ and $\mathbf{c} \in K_{\mathrm{sgn}(q-s)}$.

Theorem 6 and the fact that $\|A\|_{1\infty} = \rho$ show that equality holds in (1.5) for all p, q, r, s with $1 \leq p < r \leq \infty$ and $1 \leq s < q \leq \infty$ if and only if A is the tensor product of two vectors in K_1.

REFERENCES

[GvL] Gene H. Golub and Charles F. van Loan, Matrix Computations, Johns Hopkins U. Press, 1983.

[H] N. J. Higham, Accuracy and Stability of Algorithms, SIAM 1996.

[HJ] R. A. Horn and Charles R. Johnson, Matrix Analysis, Cambridge, 1985.

[HLP] G. H. Hardy, J. E. Littlewood, and G. Polya, Inequalities, Cambridge, 1952.

[SS] H. Schneider and G. A. Strang, Comparison theorems for supremum norms, Numer. Math. 4:15–20, 1962.