* 92

Linear and Multilinear Algebra, 1988, Vol. 23, pp. 63–78
Reprints available directly from the publisher.
Photocopying permitted by license only

Printed in the United States of America

Additive Decomposition of Nonnegative Matrices with Applications to Permanents and Scaling†

SHMUEL FRIEDLAND*
Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Chicago, Illinois 60680

CHI-KWONG LI**
Department of Mathematics, University of Wisconsin, Madison, Wisconsin 53706

and

HANS SCHNEIDER**
Department of Mathematics, University of Wisconsin, Madison, Wisconsin 53706

(*Received April 22, 1987*)

Let U_1 and U_2 be compact subsets of $m \times n$ nonnegative matrices with prescribed row sums and column sums. Given A in U_2, we study the quantity

$$\mu(U_1; A) = \max\{b : A - bB \text{ is nonnegative for some } B \text{ in } U_1\}$$

and the matrices B in U_1 that satisfy $A - \mu(U_1; A)B$ is nonnegative. The quantity

$$\mu^*(U_1, U_2) = \min\{\mu(U_1; A) : A \in U_2\}$$

is determined. Using the results obtained, we give a lower bound for the permanent of nonnegative matrices. Moreover, we study the scaling parameters of nonnegative matrices. An upper bound and an extremal characterization for their product are given.

1. INTRODUCTION

Let A be an $m \times n$ matrix with row i of sum r_i $(i = 1, \ldots, m)$ and

† Presented at the Third Haifa Matrix Theory Conference, January 1987.
* The work of this author was partially supported by NSF grant DMS-8520434.
** The work of these authors was partially supported by NSF grant DMS-8521521.

column j of sum c_j ($j=1,\ldots,n$). We call $r=(r_1,\ldots,r_m)$ the *row sum vector* and $c=(c_1,\ldots,c_n)$ the *column sum vector* of A. We denote by $U(r,c)$ (resp. $U(R,C)$) the class of all $m\times n$ nonnegative matrices with row sum vector r (resp. R) and column sum vector c (resp. C). Several authors [3, 6, 8] have studied the following problems.

(I) Given $U(r,c)$ and a matrix A in $U(R,C)$, determine the quantity

$$\mu(U(r,c);A)=\max\{b: A-bB\geqslant 0 \text{ for some } B\in U(r,c)\},$$

where a matrix $D\geqslant 0$ means that D is nonnegative.

(II) Given $U(r,c)$ and $U(R,C)$, determine the quantity

$$\mu^*(U(r,c),U(R,C))=\min\{\mu(U(r,c);A): A\in U(R,C)\}.$$

In particular, Cruse [6] (see also [3]) has given a complete answer to question (I), and Fulkerson [8] has solved problem (II) for certain sets of 0–1 matrices. The purposes of this note are as follows. For problem (I) we study the matrices B in $U(r,c)$ such that $A-\mu(U(r,c);A)B$ is nonnegative. We give a complete answer to question (II) and consider similar problems for sets of integral matrices. The results are then applied to obtain a lower bound for the permanent of nonnegative matrices. Bregman [4] (see also [1]) has also obtained a lower bound for permanent, we compare our result with his.

Given an $n\times n$ nonnegative matrix $A=(a_{ij})$, the positive numbers x_j and y_j ($1\leqslant j\leqslant n$) are called the *scaling parameters* of A if the matrix $(a_{ij}x_iy_j)$ is doubly stochastic. As a second application of our results we study the scaling parameters of nonnegative matrices. An upper bound and an extremal characterization for their product are obtained.

Clearly $U(r,c)$ is nonempty for $r,c\geqslant 0$ if and only if $\sum_{i=1}^m r_i=\sum_{j=1}^n c_j$. In the rest of the paper we always assume $r=(r_1,\ldots,r_m)$ and $c=(c_1,\ldots,c_n)$, where r_i and c_j are positive numbers satisfying $\sum_{i=1}^m r_i=\sum_{j=1}^n c_j$. When no confusion should arise, we write $\mu(A)$ for $\mu(U(r,c);A)$, and μ^* for $\mu^*(U(r,c),U(R,C))$.

2. DECOMPOSITION OF MATRICES

Let $\langle n\rangle$ denote the set $\{1,\ldots,n\}$. Suppose $I\subseteq\langle m\rangle$, $J\subseteq\langle n\rangle$ and $A=(a_{ij})\in U(R,C)$. We use the following notation:

$A[I,J]$: the submatrix of A lying in all rows i and all columns j, with $i\in I$ and $j\in J$,

$A_{I,J}$: the sum of the entries of $A[I,J]$,
R_I: the sum of those R_i with $i \in I$,
C_J: the sum of those C_j with $j \in J$,
I': the complement of I in $\langle m \rangle$,
J': the complement of J in $\langle n \rangle$.

Cruse [6] (see also [3]) has proved

THEOREM 2.1 *Given $U(r,c)$ and $A \in U(R,C)$, we have*

$$\mu(U(r,c);A) = \min\{A_{I,J}/(r_I - c_{J'}) : I \subseteq \langle m \rangle,\ J \subset \langle n \rangle,\ r_I - c_{J'} > 0\}.$$

From Theorem 2.1, one easily deduces the following corollary.

COROLLARY 2.2 *Given $U(r,c)$ and $A \in U(R,C)$, the following conditions are equivalent.*

(a) $\mu(U(r,c);A) = \mu(A) > 0$.
(b) *If $I \subseteq \langle m \rangle$ and $J \subseteq \langle n \rangle$ satisfy $r_I - c_{J'} > 0$, then $A_{I,J} > 0$.*
(c) *If $I \subseteq \langle m \rangle$ and $J \subseteq \langle n \rangle$ satisfy $A_{I,J} = 0$, then $r_I - c_{J'} \leqslant 0$.*

By Corollary 2.2 we see that the more nonzero entries A has, the more probable that $\mu(A)$ is positive. For the most extreme case, if every entry of A is positive, then $\mu(A)$ must be positive. This can also be deduced from the following observations.

(I) The matrix $A = (a_{ij})$ satisfies $\mu(A) > 0$ if and only if there exists $B = (b_{ij})$ in $U(r,c)$ such that $a_{ij} = 0$ implies $b_{ij} = 0$.
(II) The matrix $t^{-1}(r_i c_j)$, where $t = r_{\langle m \rangle} = c_{\langle n \rangle}$, is always in $U(r,c)$.

In fact, applying these observations in the extreme case, we easily get

$$\mu(A) \geqslant \min\{ta_{ij}/(r_i c_j) : i \in \langle m \rangle, j \in \langle n \rangle\} > 0.$$

Besides knowing $\mu(A)$, one might want to have more information about the matrix B such that the matrix $A - \mu(A)B \geqslant 0$, we have

THEOREM 2.3 *Suppose $U(r,c)$ and A are given such that $\mu(A) > 0$. Let $B \in U(r,c)$ satisfy $A = \mu(A)B + D$ with $D \geqslant 0$. Then for any $I \subseteq \langle m \rangle$ and $J \subseteq \langle n \rangle$ with $A_{I,J} = \mu(A)(r_I - c_{J'})$, we have $B[I,J] = A[I,J]/\mu(A)$ and $B[I',J'] = 0$. Consequently, $D[I,J] = 0$ and $D[I',J'] = A[I',J']$.*

Proof Suppose A and B satisfy the hypotheses of the theorem. In general, for any $I \subseteq \langle m \rangle$ and $J \subseteq \langle n \rangle$ we have

$$A_{I,J} \geqslant \mu(A)B_{I,J} \geqslant \mu(A)(r_I - c_{J'}).$$

If $A_{IJ} = \mu(A)(r_I - c_{J'})$, then $A_{IJ} = \mu(A)B_{IJ}$. It follows that

$$B[I,J] = A[I,J]/\mu(A) \quad \text{as} \quad A[I,J] - \mu(A)B[I,J] \geqslant 0.$$

Moreover, since $B_{IJ} = r_I - c_{J'} = B_{IJ} - B_{I'J'}$, we have $B[I',J'] = 0$. ■

Given $U(r,c)$ and A, there is no doubt that we can find B in $U(r,c)$ such that $A - \mu(A)B \geqslant 0$. Since the set of all those B that satisfy $A = \mu(A)B + D$ with $D \geqslant 0$ is convex, either there is only one element or there are infinitely many elements in $U(r,c)$ which satisfy the equation. We determine the conditions for the solution to be unique in the following theorem. Associate with the matrices B and D a (directed) bipartite graph $G(B,D)$ with nodes $\{x_1, \ldots, x_m, y_1, \ldots, y_n\}$. There is an arc (x_i, y_j) if the (i,j) entry of D is positive, and an arc (y_j, x_i) if the (i,j) entry of B is positive. We have

THEOREM 2.4 *Suppose $U(r,c)$ and A are given such that $\mu(A) > 0$. Let $A = \mu(A)B + D$ with $D \geqslant 0$. Then the matrix B is the unique element in $U(r,c)$ such that $A - \mu(A)B \geqslant 0$ if and only if the graph $G(B,D)$ does not contain cycles of length greater than two.*

Proof Let A and $U(r,c)$ satisfy the assumptions of the theorem. Suppose

$$A = \mu(A)B + D$$

with $D \geqslant 0$. If there is a cycle (x_{i_1}, y_{j_1}), (y_{j_1}, x_{i_2}), (x_{i_2}, y_{j_2}), $\ldots$, (x_{i_l}, y_{j_l}), (y_{j_l}, x_{i_1}) in $G(B,D)$ with $l > 1$, then we can construct a corresponding cycle matrix Q with the same dimension as B such that the (i_1, j_1) entry equals 1, (i_2, j_1) entry equals -1. (i_2, j_2) entry equals $1, \ldots, (i_1, j_l)$ entry equals -1. Then for a sufficiently small $\theta > 0$, we obtain another matrix $B' = B + \theta Q \in U(r,c)$ and $D' = D - \mu(A)\theta Q \geqslant 0$ satisfying

$$A = \mu(A)B' + D'.$$

Conversely, suppose there exist another matrix B' in $U(r,c)$ and $D' \geqslant 0$ such that the above equality holds. Then $B_0 = B - B' \neq 0$ is a matrix with zero row sums and zero column sums. Moreover, if the (i,j) entry of B_0 is positive. Then the (i,j) entry of B is positive; if the (i,j) entry of B_0 is negative, then the (i,j) entry of D is positive. Let the (i_1, j_1) entry of B_0 be positive. Then there must be a negative (i_2, j_1) entry, a positive (i_2, j_2) entry, a negative (i_3, j_2) entry.... We must eventually return to some previous position in a certain number of steps. This will give rise to a cycle in $G(B, D)$. Since the cycle in $G(B, D)$ is obtained by alternate vertical and horizontal movements in B_0, its length cannot be less than four. ∎

By Theorem 2.4 one can check whether there are any elements X in $U(r, c)$ other than B that satisfy $A - \mu(A)X \geqslant 0$ as follows. Start from a positive entry of D, move vertically to a positive entry of B, then move horizontally to a positive entry of D, then move vertically to a positive entry of $B, \ldots,$ until we revisit an entry in B or D. If we can find such a path, then there are other elements in $U(r, c)$ that satisfy the equation. We illustrate these procedures in the following examples.

Example 2.5 Let $U(r, c)$ be such that $r = c = (1, 1, 1)$. Suppose

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

$$\text{and} \qquad D = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then $\mu(A) = 1$ and $A = \mu(A)B + D$.

We can start from the $(1, 2)$ entry of D, move to the $(2, 2)$ entry of B, move to the $(2, 1)$ entry of D, move to the $(1, 1)$ entry of B, then move back to the $(1, 2)$ entry of D. So B is not the only element in $U(r, c)$ such that $A - \mu(A)B \geqslant 0$. In fact, interchanging the first two rows of B, we get another matrix B' in $U(r, c)$ such that $A - \mu(A)B' \geqslant 0$.

Example 2.6 Let $U(r, c)$ be such that $r = c = (1, 1, 1)$. Suppose

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad D = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then $\mu(A) = 1$ and $A = \mu(A)B + D$.

Clearly, whether we start from the $(1, 3)$ or the $(2, 3)$ entry of D, we cannot construct the required path. Therefore B is the only element in $U(r, c)$ such that $A - \mu(A)B \geqslant 0$.

3. DECOMPOSITION OF MATRIX SETS

THEOREM 3.1 *Given $U(r,c)$ and $U(R,C)$, we have*

$$\mu^*(U(r,c), U(R,C)) = \min\{[R_I - C_{J'}]^+/(r_I - c_{J'}) : I \subseteq \langle m\rangle, \\ J \subseteq \langle n\rangle, r_I - c_{J'} > 0\}.$$

where $[t]^+ = \max\{t, 0\}$.

Proof Let

$$\sigma = \min\{[R_I - C_{J'}]^+/(r_I - c_{J'}) : I \subseteq \langle m\rangle, J \subseteq \langle n\rangle, r_I - c_{J'} > 0\}.$$

If $A \in U(R,C)$, then for any $I \subseteq \langle m\rangle$ and $J \subseteq \langle n\rangle$

$$A_{I,J} \geqslant A_{I,J} - A_{I',J'} = R_I - C_{J'}.$$

It follows that $\mu(A) \geqslant \sigma$ for all A in $U(R,C)$. Thus $\mu^* \geqslant \sigma$. We shall show that there exists A in $U(R,C)$ satisfying $\mu(A) \leqslant \sigma$. As a result, $\mu^* \geqslant \sigma \geqslant \mu(A) \geqslant \mu^*$ implies $\mu^* = \sigma$. Now suppose $I \subseteq \langle m\rangle$ and $J \subseteq \langle n\rangle$ satisfy

$$\sigma = [R_I - C_{J'}]^+/(r_I - c_{J'}).$$

We consider two cases.

Case 1 $\sigma = 0$, i.e. $0 \geqslant R_I - C_{J'}$. Let $A \in U(R,C)$. If $A[I,J] = 0$, then $\mu(A) \leqslant A_{I,J}{}^*(r_I - c_{J'}) = \sigma$ and we are done. Suppose $A[I,J] \neq 0$ and $a_{ij} > 0$ with $i \in I$ and $j \in J$. Since $A_{I,J} - A_{I',J'} = R_I - C_{J'} \leqslant 0$, there exists $a_{pq} > 0$ with $p \in I'$ and $q \in J'$. Set $\delta = \min\{a_{ij}, a_{pq}\}$. Obtain A_1 from A by adding δ to its (i,q), (p,j) entries and subtracting δ from its (i,j), (p,q) entries. If $A_1[I,J] = 0$, then A_1 is the matrix in $U(R,C)$ with $\mu(A_1) = 0$; otherwise we repeat the above procedures until we get a matrix A_k in $U(R,C)$ such that $A_k[I,J] = 0$ and hence $\mu(A_k) = 0$.

Case 2 $\sigma > 0$, i.e., $0 < R_I - C_{J'} = C_J - R_{I'}$. In this case we can construct an A in $U(R,C)$ with $A[I',J'] = 0$ by similar method as in Case 1. Then we have $A_{I,J} = A_{I,J} - A_{I',J'} = R_I - C_{J'}$, and

$$\mu(A) \leqslant A_{I,J'}/(r_I - c_{J'}) = [R_I - C_{J'}]^+/(r_I - c_{J'}) = \sigma. \qquad \blacksquare$$

COROLLARY 3.2 *Given $U(r,c)$ and $U(R,C)$, the following are equivalent.*

(a) $\mu^*(U(r,c), U(R,C)) = \mu^* > 0$.

(b) *If $I \subseteq \langle m\rangle$ and $J \subseteq \langle n\rangle$ satisfy $r_I - c_J > 0$, then $R_I - C_J > 0$.*

(c) *If $I \subseteq \langle m\rangle$ and $J \subseteq \langle n\rangle$ satisfy $R_I - C_J \leqslant 0$, then $r_I - c_J \leqslant 0$.*

Note that $r_I - c_J = c_{J'} - r_{I'}$ and $R_I - C_J = C_{J'} - R_{I'}$. It follows that $r_I - c_J > 0$ if and only if $r_{I'} - c_{J'} < 0$, and $R_I - C_J > 0$ if and only if $R_{I'} - C_{J'} < 0$. Thus condition (b) of Corollary 3.2 is equivalent to

(b′) *If* $I \subseteq \langle m \rangle$ *and* $J \subseteq \langle n \rangle$ *satisfy* $r_I - c_J < 0$, *then* $R_I - C_J < 0$.

Similarly condition (c) is equivalent to

(c′) *If* $I \subseteq \langle m \rangle$ *and* $J \subseteq \langle n \rangle$ *satisfy* $R_I - C_J \geqslant 0$, *then* $r_I - c_J \geqslant 0$.

COROLLARY 3.3 *Let* Ω_n *be the set of all* $n \times n$ *doubly stochastic matrices. Suppose* $U(r, c) = \Omega_n$ *and* $U(R, C)$ *satisfies* $R_1 \geqslant \cdots \geqslant R_n$ *and* $C_1 \geqslant \cdots \geqslant C_n$. *Then*

$$\mu^*(\Omega_n, U(R, C)) = \min\left\{\left[\sum_{j=1}^{k+1} R_{n-j+1} - \sum_{j=1}^{k} C_j\right]^+ : 0 \leqslant k \leqslant n-1\right\}.$$

Proof Note that if $U(r, c) = \Omega_n$, then $m = n$ and $r_I - c_{J'} = |I| + |J| - n$. So by Theorem 3.1, we have

$$\mu^*(\Omega_n, U(R, C)) = \min\{[R_I - C_{J'}]^+/(|I| + |J| - n) : |I| + |J| - n > 0\}.$$

To get the conclusion, we prove that μ^* can always be attained at certain I_0, J_0 such that $I_0 = \{n - l, \ldots, n\}$ and $J_0' = \{1, \ldots, l\}$ for some $1 \leqslant l \leqslant n$. Suppose I, J are such that $|I| + |J| - n = k > 0$ and $\mu^* = [R_I - C_{J'}]^+/k$. Since $R_i \geqslant \mu^*$ for all $i = 1, \ldots, n$, and by our assumption on R_i and C_j, we have

$$R_I - C_{J'} \geqslant \sum_{i=1}^{|I|} R_{n-i+1} - \sum_{j=1}^{|J'|} C_j \geqslant (k-1)\mu^* + \sum_{i=|J|}^{n} R_i - \sum_{j=1}^{n-|J|} C_j.$$

Set I_0, J_0 so that $I_0 = \{|J|, \ldots, n\}$ and $J_0' = \{1, \ldots, n - |J|\}$. If $\mu^* = 0$, then $0 \geqslant R_I - C_{J'} \geqslant R_{I_0} - C_{J_0'}$. If $\mu^* > 0$,

$$k\mu^* = R_I - C_{J'} \geqslant (k-1)\mu^* + R_{I_0} - C_{J_0'} \geqslant k\mu^*.$$

So we have $[R_{I_0} - C_{J_0}]^+ = \mu^*$ as required. ■

By Theorem 3.1, it seems that we have to consider many expressions of the form $[R_I - C_{J'}]^+/(r_I - c_{J'})$ in order to find μ^*. By Corollary 3.3, we see that in actual computation, we usually may consider far fewer expressions.

In view of Corollary 3.3, one might expect to have

$$\mu(\Omega_n; A) = \min\{A_{I,J} : |I| + |J| = n + 1\}.$$

Unfortunately, it is not true in general as shown by

Example 3.4 Let

$$A = \begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 1 & 1 & 3 \\ 1 & 1 & 1 & 3 \\ 3 & 3 & 3 & 3 \end{bmatrix}.$$

Then

$$\mu(\Omega_4; A) = 4.5 < 6 = \min\{A_{I,J} : |I| + |J| = 5\}.$$

Now suppose that we have $|R_i - r_i| < \varepsilon$, $|C_i - c_i| < \varepsilon$ $(1 \leqslant i \leqslant n)$ for some positive ε. If $U(r, c) = \Omega_n$, it is clear from Corollary 3.3 that $\bar{\mu}^* = \mu^*(\Omega_n, U(R, C)) > 0$ if $\varepsilon < 1/(2n - 1)$. In fact, it is not hard to show that $\bar{\mu}^* > 0$ if $\varepsilon < 1/n$ (see Theorem 5.1 in Section 5). For general $U(r, c)$ this is false as the following example shows.

Example 3.5 Let $r = (1, 1)$, $c = (1 + \varepsilon, 1 - \varepsilon)$, where $\varepsilon > 0$. Let I be the 2×2 identity matrix. Since every matrix in $U(r, c)$ has positive $(2, 1)$ entry, we have $\mu(U(r, c); I) = 0$. It follows that $\mu^*(U(r, c), \Omega_2) = 0$.

4. SETS OF INTEGRAL MATRICES

In [8] Fulkerson studied the maximum number of disjoint permutation matrices a 0–1 matrix can contain. We consider similar problem for sets of integral matrices. Denote by $U'(r, c)$ (resp. $U'(R, C)$) the set of integral matrices with (integral) row sum vector r (resp. R) and column sum vector c (resp. C). Define

$$\pi(U'(r, c); A) = \max\{l : A - (B_1 + \cdots + B_l) \geqslant 0 \text{ with } B_1, \ldots, B_l \in U'(r, c)\}$$

for any A in $U'(R, C)$, and

$$\pi^*(U'(r, c), U'(R, C)) = \min\{\pi(U'(r, c); A) : A \in U'(R, C)\}.$$

As mentioned in [6], by network flow theory one easily proves an integral version of Theorem 2.1.

THEOREM 4.1 *Suppose $U'(r, c)$ is nonempty and $A \in U'(R, C)$. Then $\pi(U'(r, c); A)$ is the integral part of $\mu(U(r, c); A)$.*

By a slight modification of the proof of Theorem 3.1 we obtain

THEOREM 4.2 *Suppose $U'(R,C)$ and $U'(r,c)$ are nonempty. Then $\pi^*(U'(r,c), U'(R,C))$ is the integral part of the quantity $\mu^*(U(r,c), U(R,C))$.*

By Corollary 3.3 and Theorem 4.2, we get

COROLLARY 4.3 *Let P_n be the set of $n \times n$ permutation matrices. If $U'(R,C)$ is a set of integral matrices, then*

$$\pi^*(P_n, U'(R,C)) = \mu^*(\Omega_n, U(R,C)).$$

Note that if A is an $n \times n$ 0–1 matrix and $U'(r,c) = P_n$, then $\pi(P_n; A)$ is the maximum number of disjoint permutation matrices A contains. So Theorem 4.1 reduces to the result of Fulkerson [8]. However, if $U''(R,C)$ is a nonempty set of 0–1 matrices with row sum vector R and column sum vector C and

$$\tilde{\pi}(P_n, U''(R,C)) = \min\{\pi(P_n; A): A \in U''(R,C)\},$$

then the formula for $\tilde{\pi}(P_n, U''(R,C))$ is much more complicated than $\pi^*(P_n, U'(R,C))$ as shown in [8]. Nevertheless, in some particular cases, we may have better results as shown in the following theorem.

THEOREM 4.4 *Let k and s be positive integers. Assume that $n - 2k + 1 \geqslant \max\{k, s\}$. Let*

$$R_1 = \cdots = R_s = C_1 = \cdots = C_s = k - 1,$$

$$R_{s+1} = \cdots = R_n = C_{s+1} = \cdots = C_n = k.$$

Then $U''(R,C)$ is nonempty and

$$\tilde{\pi}(P_n, U''(R,C)) = \pi^*(P_n, U'(R,C)) = [k-s]^+.$$

Proof By Corollaries 3.3 and 4.3, one easily checks that

$$\pi^*(P_n, U'(R,C)) = \mu^*(\Omega_n, U(R,C)) = [k-s]^+.$$

In general, we have

$$\tilde{\pi}(P_n, U''(R,C)) \geqslant \pi^*(P_n, U'(R,C)).$$

So we only need to construct a matrix A in $U''(R,C)$ such that

$$\pi(P_n; A) \leqslant [k-s]^+.$$

Then the result of the theorem will follow.

Let T be the $k \times (k-1)$ matrix all filled with ones. Assume that $s \geqslant k$. Let Q be the $(2k-1) \times (2k-1)$ matrix such that

$$Q[I,I]=0, \qquad Q[I',I']=0, \qquad Q[I,I']=T, \qquad Q[I',I]=T',$$

where $I=\{1,\ldots,k\}$. Note that $\mathrm{per}(Q)=0$, k rows and columns of Q have $k-1$ ones and all other rows and columns contain k ones. Set $p=s-k$ and $q=n-s-k+1$. Then $p \geqslant 0$, $q \geqslant k$ and $m=p+q=n-2k+1$, By a result of Gale and Ryser (see [9, pp. 176–178]), we can construct an $m \times m$ 0–1 matrix D such that p rows and columns of D have $k-1$ ones and all other rows and columns contain k ones. (In fact, the row sum vector of D is majorized by any integral vectors with m entries whose sum equals $p(k-1)+qk$. In particular, it is majorized by the conjugate column sum vector of D. Hence the construction is possible.) Let A be the direct sum of Q and D. Then $\mathrm{per}(A)=0=[k-s]^{+}=\pi(p_n;A)$, s rows and columns of A have $k-1$ ones and all other rows and columns contain k ones. By a suitable permutation of the rows and columns of A, we get the required matrix in $U''(R,C)$.

Assume now that $k>s$. Let B be the $(2k-1)\times(2k-1)$ matrix obtained by adding ones to the $(1,1),\ldots,(k-s,k-s)$ positions of the matrix Q constructed above. Let $m=n-2k+1$. Then $m \geqslant k$, and we can consttuct an $m \times m$ 0–1 matrix C with k ones in every row and every column. Let A be the direct sum of B and C. Set $I=\{1,\ldots,k\}$ and $J=\{1,\ldots,k,2k,2k+1,\ldots,n\}$. Then by Theorems 4.1 and 2.1

$$\pi(P_n;A) \leqslant \mu(\Omega_n;A) \leqslant A_{IJ}/(|I|-|J'|)=k-s=[k-s]^{+}.$$

Clearly, s rows and columns of A have $k-1$ ones and all other rows and columns contain k ones. So by a suitable permutation of rows and columns of A, we get the required matrix in $U''(R,C)$. ∎

5. PERMANENTS AND SCALING PARAMETERS

In this section we concentrate on the relation between an $n \times n$ nonnegative matrix A and the set Ω_n. In particular, for an $n \times n$ nonnegative matrix A in $U(R,C)$, we define

$$\bar{\mu}(A)=\mu(\Omega_n;A)$$

and

$$\bar{\mu}^*(A)=\mu^*(\Omega_n,U(R,C)).$$

Furthermore, setting

$$S(A) = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} C_i$$

and

$$\omega(A) = \frac{1}{S(A)} \max_{1 \leqslant i \leqslant n} (|nR_i - S(A)|, |nC_i - S(A)|),$$

we have

THEOREM 5.1 *Let A be an $n \times n$ nonnegative matrix in $U(R, C)$. If $\bar{\mu}(A)$, $\bar{\mu}^*(A)$, $\omega(A)$ and $S(A)$ are defined as above, then*

$$\frac{n^n}{n!} \operatorname{per}(A) \geqslant \bar{\mu}(A)^n \geqslant \bar{\mu}^*(A)^n \geqslant \left(\frac{S(A)}{n}[1 - n\omega(A)]^+\right)^n.$$

Proof Since by Egorichev–Falikman Theorem

$$\operatorname{per}(B) \geqslant n!/n^n$$

for any B in Ω_n, we easily deduce

$$\operatorname{per}(A) \geqslant \bar{\mu}(A)^n n!/n^n.$$

To get the conclusion, we only need to prove

$$\bar{\mu}^*(A) \geqslant S(A)([1 - n\omega(A)]^+)/n.$$

Note that we may write

$$R_i = (1 + x_i\omega(A))S(A)/n$$

and

$$C_i = (1 + y_i\omega(A))S(A)/n,$$

where

$$1 \geqslant x_i, \qquad y_i \geqslant -1; \qquad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = 0.$$

Let $k = |J| = |I| - 1$. Suppose $1 \leqslant k < n/2$. Then

$$\begin{aligned}\sum_{i \in I} R_i - \sum_{j \in J} C_j &= \frac{S(A)}{n}\left[1 + \left(\sum_{i \in I} x_i - \sum_{j \in J} y_j\right)\omega(A)\right] \\ &\geqslant \frac{S(A)}{n}[1 + (-|I| - |J|)\omega(A)] \\ &\geqslant \frac{S(A)}{n}[1 - n\omega(A)].\end{aligned}$$

Suppose $n/2 \leqslant k < n$. Then

$$\begin{aligned}\sum_{i\in I} R_i - \sum_{j\in J} C_j &= \frac{S(A)}{n}\left[1 + \left(\sum_{i\in I} x_i - \sum_{j\in J} y_j\right)\omega(A)\right] \\ &= \frac{S(A)}{n}\left[1 + \left(-\sum_{i\in I'} x_i + \sum_{j\in J'} y_j\right)\omega(A)\right] \\ &\geqslant \frac{S(A)}{n}[1 + (-|I'| - |J'|)\omega(A)] \\ &\geqslant \frac{S(A)}{n}[1 - n\omega(A)].\end{aligned}$$

In both cases, we have

$$\left[\sum_{i\in I} R_i - \sum_{j\in J} C_j\right]^+ \geqslant \frac{S(A)}{n}[1 - n\omega(A)]^+.$$

In view of Corollary 3.3 we have

$$\bar{\mu}^*(A) \geqslant \frac{S(A)}{n}[1 - n\omega(A)]^+. \qquad \blacksquare$$

Bregman in [4] has obtained a bound for the permanent of nonnegative matrices as follows.

For any nonnegative $n \times n$ matrix $A = (a_{ij})$ and for any $B = (b_{ij}) \in \Omega_n$,

$$\operatorname{per}(A) \geqslant \operatorname{per}(B) \prod_{i,j=1}^{n} (a_{ij}/b_{ij})^{\lambda_{ij}},$$

where $\lambda_{ij} = b_{ij} \operatorname{per}(B[I, J])/\operatorname{per}(B)$ with $I = \langle n \rangle \backslash \{i\}$ and $J = \langle n \rangle \backslash \{j\}$.

We remark that in order to get the bound of Bregman, one has to find a B in Ω_n to make the comparison. Whether the bound of Bregman is better than ours depends on the choice of the matrix B. However, if $B \in \Omega_n$ satisfies $A = \bar{\mu}(A)B + D$ with $D \geqslant 0$, then

$$\begin{aligned}\operatorname{per}(A) &\geqslant \operatorname{per}(B) \prod_{i,j} (a_{ij}/b_{ij})^{\lambda_{ij}} \\ &\geqslant \operatorname{per}(B) \prod_{i,j} \bar{\mu}(A)^{\lambda_{ij}} \\ &\geqslant \bar{\mu}(A)^n n!/n^n.\end{aligned}$$

So we have

$$\bar{\mu}(A)^n n!/n^n \leqslant \max\left\{\operatorname{per}(B)\prod_{i,j}(a_{ij}/b_{ij})^{\lambda_{ij}} : B \in \Omega_n\right\}$$
$$\leqslant \operatorname{per}(A).$$

In any case, the bounds we obtain is simple and easy to compute.

Several authors [5, 12] have studied necessary and sufficient conditions on an $n \times n$ nonnegative matrix A such that there exists a doubly stochastic matrix B of the form XAY, where $X = \operatorname{diag}(x_1, \ldots, x_n)$ and $Y = \operatorname{diag}(y_1, \ldots, y_n)$ satisfy $x_i > 0$ and $y_i > 0$ for $i = 1, \ldots, n$, (for further exposition see [10]). The numbers x_i and y_i $(1 \leqslant i \leqslant n)$ are known as the scaling parameters of A. The following result was proved.

An $n \times n$ nonnegative matrix A has scaling parameters if and only if A has total support, i.e., every positive entry of A lies on a positive diagonal.

Using the results in the previous sections we give an upper bound for $\prod_{i=1}^n x_i y_i$.

THEOREM 5.2 *Let A be an $n \times n$ nonnegative matrix with scaling parameters x_j, y_j $(1 \leqslant j \leqslant n)$. Then*

$$\prod_{j=1}^n x_j^{-1} y_j^{-1} \geqslant \bar{\mu}(A)^n \geqslant \bar{\mu}^*(A)^n \geqslant \left(\frac{S(A)}{n}[1 - \omega(A)n]^+\right)^n.$$

Proof By Theorem 5.1, we only need to prove the first inequality. Let $XAY = B \in \Omega_n$, where $X = \operatorname{diag}(x_1, \ldots, x_n)$ and $Y = \operatorname{diag}(y_1, \ldots, y_n)$. By a result of Friedland [7]

$$\lim_{m \to \infty} [\operatorname{per}(B \otimes J_m)]^{1/m} = e^{-n}.$$

Thus

$$\lim_{m \to \infty} [\operatorname{per}(A \otimes J_m)]^{1/m} = e^{-n} \prod_{j=1}^n x_j^{-1} y_j^{-1}.$$

On the other hand, since $A - \bar{\mu}(A)D \geqslant 0$ for some $D \in \Omega_n$, we have

$$A \otimes J_m \geqslant \bar{\mu}(A)D \otimes J_m.$$

So

$$[\operatorname{per}(A \otimes J_m)]^{1/m} \geqslant \bar{\mu}(A)^n e^{-n}.$$

The result follows. ■

We remark that the first inequality in Theorem 5.2 also follows from the theorem in [2].

We give an extremal characterization for the product of scaling parameters in

THEOREM 5.3 *Let A be an $n \times n$ nonnegative matrix with scaling parameters x_j, y_j $(1 \leqslant j \leqslant n)$. Then*

$$\prod_{j=1}^{n} x_j^{-1} y_j^{-1} = \max\{\bar{\mu}(D_1 A D_2)^n : D_1 \text{ and } D_2 \text{ are } n \times n \text{ nonnegative diagonal matrices satisfying } \operatorname{per}(D_1 D_2) = 1\}.$$

Proof For an $n \times n$ nonnegative matrix B which can be scaled to a doubly stochastic matrix, let $\delta(B)$ denote the product of the reciprocals of the scaling parameters. Using this notation, we have $\delta(A) = \prod_{j=1}^{n} x_j^{-1} y_j^{-1}$. Moreover, if

$$S = \{D_1 A D_2 : D_1 \text{ and } D_2 \text{ are } n \times n \text{ nonnegative diagonal matrices satisfying } \operatorname{per}(D_1 D_2) = 1\}$$

then for any B in S

$$\delta(A) = \delta(B) \geqslant \bar{\mu}(B)^n$$

by Theorem 5.2. Thus

$$\delta(A) \geqslant \max\{\bar{\mu}(B)^n : B \in S\}.$$

Finally for $D_1 = \delta(A)^{1/n} X$ and $D_2 = Y$, we have $D_1 A D_2 \in S$ and

$$\bar{\mu}(D_1 A D_2)^n = \bar{\mu}(\delta(A)^{1/n} X A Y)^n = \delta(A).$$

The result follows. ■

COROLLARY 5.4 *Suppose $A \in \Omega_n$. Then for any nonnegative diagonal matrices D_1 and D_2 we have*

$$\bar{\mu}(D_1 A D_2)^n \leqslant \operatorname{per}(D_1 D_2).$$

Proof If D_1 or D_2 has zero diagonal entries, then both $\bar{\mu}(D_1 A D_2)$ and $\operatorname{per}(D_1 D_2)$ equal zero. If it is not the case, by Theorem 5.3,

$$\lambda^n \bar{\mu}(D_1 A D_2)^n = \bar{\mu}(\lambda D_1 A D_2)^n \leqslant 1,$$

where $\lambda = \operatorname{per}(D_1 D_2)^{-1/n}$. The result follows. ■

Corollary 5.4 is also a consequence of the theorem in [2].
In view of Corollary 3.3 and by Corollary 5.4, we have

COROLLARY 5.5 *Suppose $A \in \Omega_n$ and D_1, D_2 are nonnegative diagonal matrices. If D_1AD_2 has row sums $R_1 \geqslant \cdots \geqslant R_n$ and column sums $C_1 \geqslant \cdots \geqslant C_n$, then there exists k such that $1 \leqslant k \leqslant n-1$ and*

$$\left(\sum_{j=1}^{k+1} R_{n-j+1} - \sum_{j=1}^{k} C_j\right)^n \leqslant \operatorname{per}(D_1D_2).$$

By Corollary 5.4, we see the following interesting property of Ω_n. Suppose $A, B \in \Omega_n$, D_1 and D_2 are nonnegative diagonal matrices with $\operatorname{per}(D_1D_2) = 1$. If $D_1AD_2 - \lambda B \geqslant 0$, then $\lambda \leqslant 1$. In general one might want to know that for given $A = (a_{ij})$ and $B = (b_{ij})$, whether the set

$$\Gamma(A,B) = \{\sigma \geqslant 0 : D_1AD_2 - \sigma B \geqslant 0, \text{ where } D_1 \text{ and } D_2 \text{ are nonnegative diagonal matrices with } \operatorname{per}(D_1D_2) = 1\}$$

is bounded. If it is, what is the bound

$$\eta(A,B) = \sup\Gamma(A,B)?$$

Clearly in order that $\Gamma(A,B) \neq \{0\}$, we must have $b_{ij} = 0$ whenever $a_{ij} = 0$. If it is the case, then one may verify that

$$\sup\Gamma(A,B) = \sup_{\prod_{i=1}^n x_iy_i = 1} \min_{b_{ij}>0} \frac{x_ia_{ij}y_j}{b_{ij}} = \left(\inf_{\prod_{i=1}^n x_iy_i = 1} \max_{b_{ij}>0} \frac{x_ib_{ij}y_j}{a_{ij}}\right)^{-1}.$$

By a result of Saunders and Schneider [11], if

$$P(B) = \left\{\rho : \rho \text{ is a permutation such that } \prod_{i=1}^n b_{i\rho(i)} \neq 0\right\},$$

then

$$\inf_{\prod_{i=1}^n x_iy_i = 1} \max_{b_{ij}>0} \left(\frac{x_ib_{ij}y_j}{a_{ij}}\right) = \begin{cases} 0 & \text{if } P(B) = \varnothing, \\ \max_{\rho\in P(B)} \left(\prod_{i=1}^n \frac{b_{i\rho(i)}}{a_{i\rho(i)}}\right)^{1/n} & \text{otherwise.} \end{cases}$$

It follows that $\eta(A,B)$ is finite if and only if $\operatorname{per}(B) > 0$. It might be interesting to determine the quantity

$$\eta^*(U(r,c)) = \max\{\eta(A,B) : A, B \in U(r,c)\}$$

if all the elements in $U(r,c)$ have positive permanents.

References

[1] R. B. Bapat, Applications of an inequality in information theory to matrices, *Linear Algebra Appl.* **78** (1986), 107–117.

[2] R. B. Bapat and T. E. S. Raghavan, On diagonal products of doubly stochastic matrices, *Linear Algebra Appl.* **31** (1980), 71–75.

[3] S. K. Bhandari and S. D. Gupta, Two characterizations of doubly superstochastic matrices. *Sankhya: The Indian J. of Statistics* **47** (1985), 357–365.

[4] L. I. Bregman, Some properties of nonnegative matrices and their permanents, *Soviet Math. Dokl.* **14** (1973), No. 4, 945–949.

[5] R. A. Brualdi, S. V. Parter and H. Schneider, The diagonal equivalence of a nonnegative matrix to a stochastic matrix, *J. Math. Anal. and Appl.* **16** (1966), 31–50.

[6] A. B. Cruse, A proof of Fulkerson's characterization of permutation matrices, *Linear Algebra Appl.* **12** (1975), 21–28.

[7] S. Friedland, A lower bound for the permanent of a doubly stochastic matrix, *Annals of Math.* **110** (1979), 167–176.

[8] D. R. Fulkerson, The maximum number of disjoint permutations contained in a matrix of zeros and ones, *Canad. J. Math.* **16** (1964), 729–735.

[9] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, 1979.

[10] M. V. Menon and H. Schneider, The spectrum of a nonlinear operator associated with a matrix, *Linear Algebra Appl.* **2** (1969), 321–334.

[11] B. D. Saunders and H. Schneider, Cones, graphs and optimal scalings of matrices, *Linear and Multilinear Algebra* **8** (1979), 121–135.

[12] S. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, *Pacific J. Math.* **21** (1967), 343–348.