Applications of an Inequality in Information Theory to Matrices

R. B. Bapat
Indian Statistical Institute
7, S.J.S. Sansanwal Marg
New Delhi-110016, India

Submitted by Richard A. Brualdi

ABSTRACT

If x and y are nonnegative vectors of order n, and if $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, then a well-known inequality asserts that $\prod_{i=1}^n x_i^{x_i} \geqslant \prod_{i=1}^n y_i^{x_i}$, with equality if and only if $x = y$. In this paper various situations are considered where this inequality can be applied to obtain inequalities concerning nonnegative matrices. In particular, inequalities are proved concerning nonnegative matrices which are diagonally equivalent, permanents and functions more general than the permanent, and diagonal products and circuit products of nonnegative matrices.

1. INTRODUCTION

Let

$$P^n = \left\{ x \in R^n : x \geqslant 0,\ \sum_{i=1}^n x_i = 1 \right\}.$$

Throughout this paper we take $0^0 = 1$ and $0 \log 0 = 0$. The following inequality is well known and has applications in information theory (see, for example [11, p. 58]).

THEOREM 1. *If $x \in P^n$, $y \in P^n$, then*

$$\prod_{i=1}^n x_i^{x_i} \geqslant \prod_{i=1}^n y_i^{x_i},$$

with equality if and only if $x = y$.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/86/$3.50

The purpose of this paper is to describe various situations where the inequality of Theorem 1 can be applied to obtain inequalities concerning nonnegative matrices. In Section 2 we prove an inequality concerning A and D_1AD_2, where A is a nonnegative matrix and D_1 and D_2 are diagonal matrices with positive diagonal entries. In a digression, it is also shown that a result due to London [9] can be strengthened by using majorization. In Section 3 we consider a function similar to but more general than the permanent. A generalized form of an inequality due to Bregman [7] is proved in a simple way. The last two sections are devoted to diagonal products and circuit products of nonnegative matrices. The contents of Sections 3 and 4 are partly based on [3].

2. DIAGONALLY EQUIVALENT MATRICES

Let $A=((a_{ij}))$ be a positive $m\times n$ matrix, and define a map $f\colon P^m\times P^n \to P^m\times P^n$ as follows:

$$f(x,y)=(\tilde{x},\tilde{y}),$$

where

$$\tilde{x}_i=\frac{x_i\sum_{j=1}^{n}a_{ij}y_j}{\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j},\qquad i=1,2,\dots,m,$$

$$\tilde{y}_j=\frac{y_j\sum_{i=1}^{n}a_{ij}x_i}{\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j},\qquad j=1,2,\dots,n. \tag{1}$$

It was shown in [4] that f maps $P^m\times P^n$ onto itself. Also, it is known (see, for example, [4]) that f is one-one. It will be seen that this fact can also be deduced from a result in Section 4 (see Corollary 12).

The following inequality was proved by Atkinson, Watterson, and Moran [1] (also see [11, p. 71]). For any $(x,y)\in P^m\times P^n$,

$$\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\tilde{x}_i\tilde{y}_j\geqslant\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j. \tag{2}$$

In fact, the result in [1] is the following, which may be easily obtained from (2) by a continuity argument.

LEMMA 2. *Let A be a nonnegative n × n matrix and let* $(x, y) \in P^m \times P^n$. *Then*

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\left(x_i \sum_{s=1}^{n} a_{is} y_s\right)\left(y_j \sum_{r=1}^{m} a_{rj} x_r\right) \geqslant \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j\right)^3.$$

We now prove a result which is stronger than Lemma 2.

THEOREM 3. *Let A be a positive m × n matrix and let f be defined as in* (1). *Then for any* $(x, y) \in P^m \times P^n$ *and for any vectors* $\lambda \geqslant 0$, $\mu \geqslant 0$,

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \lambda_i \mu_j \geqslant \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j\right) \prod_{i=1}^{m} \left(\frac{\lambda_i}{x_i}\right)^{\bar{x}_i} \prod_{j=1}^{n} \left(\frac{\mu_j}{y_j}\right)^{\bar{y}_j}.$$

Proof. We will assume that x, y, λ, μ are positive, and the general case will follow by a continuity argument. First note that the inequality of the theorem remains unchanged if each λ_i and each μ_j is multiplied by the same positive constant. So we assume without loss of generality that

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \lambda_i \mu_j = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j.$$

But now the result follows, after a trivial simplification, from the following inequality, which is true in view of Theorem 1:

$$\prod_{i,j} (a_{ij} x_i y_j)^{a_{ij} x_i y_j} \geqslant \prod_{i,j} (a_{ij} \lambda_i \mu_j)^{a_{ij} x_i y_j}.$$

The next result is an immediate consequence of Theorems 1,3 and the following fact, which is easily deduced from Theorem 1 and the arithmetic mean-geometric mean inequality: if $x \in P^n$, $y \in P^n$, then

$$\prod_{i=1}^{n} \{\theta x_i + (1-\theta) y_i\}^{x_i} \geqslant \prod_{i=1}^{n} y_i^{x_i} \qquad \text{for} \quad 0 \leqslant \theta \leqslant 1.$$

COROLLARY 4. *Let* $(x, y) \in P^m \times P^n$, *and let* $\lambda = \theta x + (1-\theta)\bar{x}$, $\mu = \theta y + (1-\theta)\bar{y}$, $0 \leqslant \theta \leqslant 1$, *where* $\bar{x}$ *and* $\bar{y}$ *are as defined in* (1). *Then*

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\lambda_i\mu_j \geqslant \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_i y_j.$$

COROLLARY 5. *Let A be a nonnegative* $n \times n$ *matrix, and suppose* u, v *are positive vectors such that the matrix* $((a_{ij}u_i v_j))$ *is doubly stochastic. Then for any* $\lambda \geqslant 0$, $\mu \geqslant 0$,

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\lambda_i\mu_j\right)^n \prod_{i=1}^{n} u_i v_i \geqslant n^n \prod_{i=1}^{n} \lambda_i\mu_i.$$

Proof. Define x, y as

$$x_i = \frac{u_i}{\sum_i u_i}, \qquad i = 1, 2, \dots, m,$$

and

$$y_j = \frac{v_j}{\sum_j v_j}, \qquad j = 1, 2, \dots, n.$$

Now apply Theorem 3.

If A is a nonnegative $n \times n$ matrix and if u, v are positive vectors such that $((a_{ij}u_i v_j))$ is doubly stochastic, the following lower bound was obtained by London [9]: For any $x > 0$

$$\prod_{i=1}^{n} u_i v_i \geqslant \frac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}x_j\right)}. \tag{3}$$

At this point we will digress from our main topic for a moment and show that the bound in (3) can be strengthened and put in a better perspective by using the concept of majorization.

Keeping the same notation as above, let $D_1 = \operatorname{diag}(u_1, \dots, u_n)$, $D_2 = \operatorname{diag}(v_1, \dots, v_n)$, let x be a nonnegative vector, and let $y = Ax$. Then

$$D_1 y = D_1 A D_2 (D_2^{-1} x).$$

Since $D_1 A D_2$ is doubly stochastic, it follows by a well-known theorem of Hardy, Littlewood, and Polya that

$$\begin{pmatrix} u_1 y_1 \\ \vdots \\ u_n y_n \end{pmatrix} \text{ is majorized by } \begin{pmatrix} x_1/v_1 \\ \vdots \\ x_n/v_n \end{pmatrix} \tag{4}$$

We refer to Marshall and Olkin [10] for definitions and standard results concerning majorization. It follows from (4) that

$$\prod_{i=1}^{n} u_i y_i \geqslant \prod_{i=1}^{n} \frac{x_i}{v_i},$$

which is the same as (3).

3. PERMANENTS AND GENERALIZED FUNCTIONS

Let S_n denote the set of permutations of $1, 2, \dots, n$. If A is an $n \times n$ matrix, the permanent of A, denoted by $\operatorname{per} A$, is defined as

$$\operatorname{per} A = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}.$$

We will denote by $A(i, j)$ the matrix obtained by deleting the ith row and the jth column of A. Let Ω_n denote the set of $n \times n$ doubly stochastic matrices. If $A \in \Omega_n$, define $g(A) = B = ((b_{ij}))$, where

$$b_{ij} = \frac{a_{ij} \operatorname{per} A(i, j)}{\operatorname{per} A} \qquad \text{for all} \quad i, j.$$

It was shown by Bregman [7] that the map $g: \Omega_n \to \Omega_n$ is one-one and onto, and he also proved the following interesting inequality. The inequality may also be deduced from Lemma 8 of Rothaus [12].

THEOREM 6. *For any nonnegative $n \times n$ matrix Z and for any $A \in \Omega_n$,*

$$\operatorname{per} Z \geqslant \operatorname{per} A \prod_{i,j=1}^{n} \left(\frac{z_{ij}}{a_{ij}} \right)^{b_{ij}},$$

where $B = g(A)$.

We will extend the results of Bregman to functions more general than the permanent.

Fix a nonempty set $K \subset S_n$, and for any $n \times n$ matrix A, define

$$\operatorname{per}_K A = \sum_{\sigma \in K} \prod_{i=1}^{n} a_{i\sigma(i)}.$$

If A is an $n \times n$ matrix and if $\sigma \in S_n$, we define $d_\sigma(A) = \prod_{i=1}^{n} a_{i\sigma(i)}$.

First we will prove the following.

LEMMA 7. *Suppose A, B are nonnegative $n \times n$ matrices such that*

$$\sum_{\sigma(i)=j,\, \sigma \in K} d_\sigma(A) = \sum_{\sigma(i)=j,\, \sigma \in K} d_\sigma(B) \qquad \text{for all} \quad i, j. \tag{5}$$

Then $d_\sigma(A) = d_\sigma(B)$ for all $\sigma \in K$.

Proof. Note that if, for some i, j, $\sigma(i) \neq j$ for any $\sigma \in K$, we take the sums in (5) to be zero.

Summing both sides of (5) with respect to j, we have

$$\operatorname{per}_K A = \operatorname{per}_K B.$$

If $\operatorname{per}_K A = 0$, the result is trivial, so suppose $\operatorname{per}_K A > 0$. Now

$$\begin{aligned} \prod_{\sigma \in K} d_\sigma(A)^{d_\sigma(B)} &= \prod_{i,j} a_{ij}^{\sum_{\sigma(i)=j,\, \sigma \in K} d_\sigma(B)} \\ &= \prod_{i,j} a_{ij}^{\sum_{\sigma(i)=j,\, \sigma \in K} d_\sigma(A)} \\ &= \prod_{\sigma \in K} d_\sigma(A)^{d_\sigma(A)}. \end{aligned}$$

Similarly

$$\prod_{\sigma \in K} d_\sigma(B)^{d_\sigma(A)} = \prod_{\sigma \in K} d_\sigma(B)^{d_\sigma(B)}$$

So

$$\prod_{\sigma \in K} d_\sigma(A)^{d_\sigma(B)} d_\sigma(B)^{d_\sigma(A)} = \prod_{\sigma \in K} d_\sigma(A)^{d_\sigma(A)} d_\sigma(B)^{d_\sigma(B)} \tag{6}$$

Since

$$\sum_{\sigma \in K} d_\sigma(A) = \sum_{\sigma \in K} d_\sigma(B),$$

it follows by Theorem 1 that

$$\prod_{\sigma \in K} d_\sigma(A)^{d_\sigma(A)} \geqslant \prod_{\sigma \in K} d_\sigma(B)^{d_\sigma(A)}$$

and

$$\prod_{\sigma \in K} d_\sigma(B)^{d_\sigma(B)} \geqslant \prod_{\sigma \in K} d_\sigma(A)^{d_\sigma(B)}$$

In view of (6), equality must hold in the above inequalities, and so $d_\sigma(A) = d_\sigma(B)$ for all $\sigma \in K$.

Note that if we take $K = S_n$ in Lemma 7, we get the known result that $g: \Omega_n \to \Omega_n$ is one-one. For, if $A, B \in \Omega_n$ and $g(A) = g(B)$, by Lemma 7 it follows that $d_\sigma(A) = d_\sigma(B)$ for all $\sigma \in S_n$. But then $A = B$, using, for example, Corollary 11.

The next result generalizes Theorem 6. The proof is based only on Theorem 1.

THEOREM 8. *Let A be a nonnegative $n \times n$ matrix and suppose* $\operatorname{per}_K A > 0$, *where K is a fixed nonempty subset of* S_n. *Then for any* $Z \geqslant 0$,

$$\operatorname{per}_K Z \geqslant \operatorname{per}_K A \prod_{i,j} \left(\frac{z_{ij}}{a_{ij}} \right)^{b_{ij}},$$

where $B = g_K(A)$.

Proof. We will assume that $A > 0$, $Z > 0$, and the general case will follow by a continuity argument.

Note that if each entry of Z is multiplied by the same positive constant, the inequality is unaffected, and so we may assume that $\mathrm{per}_K Z = \mathrm{per}_K A$. But now the inequality reduces to

$$\prod_{i,j} a_{ij}^{b_{ij}} \geqslant \prod_{i,j} z_{ij}^{b_{ij}}.$$

This is the same as the following inequality, which is true by Theorem 1:

$$\prod_{\sigma \in K} d_\sigma(A)^{d_\sigma(A)} \geqslant \prod_{\sigma \in K} d_\sigma(Z)^{d_\sigma(A)}.$$

COROLLARY 9. *Let A be a nonnegative $n \times n$ matrix with $\sum_{i=1}^n \sum_{j=1}^n a_{ij} = n$, and suppose $\mathrm{per}_K A > 0$ for some $K \subset S_n$. Then*

(i) $\mathrm{per}_K B \geqslant \mathrm{per}_K A \prod_{i,j}(b_{ij}/a_{ij})^{b_{ij}}$, *where $B = g_K(A)$;*
(ii) $\mathrm{per}_K B \geqslant \mathrm{per}_K A$, *with equality if and only if $A = B$.*

Proof. Inequality (i) is obtained by setting $Z = B$ in Theorem 8. Now (ii) follows from (i) by Theorem 1.

It must be pointed out that (ii) of Corollary 9 has been obtained by several authors in various forms. As noted earlier, when $K = S_n$, it was obtained by Bregman [7]. It may also be deduced from more general results of Rothaus [12] and of Baum and Eagon [5]. The proof given here seems much simpler than the known proofs.

4. DIAGONAL PRODUCTS

It was proved in [2] that if A and B are distinct $n \times n$ doubly stochastic matrices, then there exists $\sigma \in S_n$ such that $d_\sigma(A) > d_\sigma(B)$. It was also pointed out in that paper that the result does not hold if we only require that A and B have corresponding row and column sums equal. However, a proper reformulation of the result is true in this case, as we will see in this section.

Fix positive integers m, n, and let r and c be positive vectors of order m and n respectively, such that $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j$. We will denote by $U(r,c)$ the set of all nonnegative $m \times n$ matrices A with row sums $r_1, \dots, r_m$ and column sums $c_1, \dots, c_n$. Then $U(r,c)$ is a compact, convex set, and the extreme points of $U(r,c)$ can be described fairly easily (see, for example, [8]).

We now have the following.

THEOREM 10. *Let* $A \in U(r,c)$, *let* B *be a nonnegative* $m \times n$ *matrix such that* $\sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij} = \sum_{i=1}^{m} r_i$, *and suppose* $A \neq B$. *Then there exists an extreme point* X *of* $U(r,c)$ *such that*

$$\prod_{i,j} a_{ij}^{x_{ij}} > \prod_{i,j} b_{ij}^{x_{ij}}.$$

Proof. For any $Z \in U(r,c)$, let

$$\phi(Z) = \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij}(\log a_{ij} - \log b_{ij}).$$

Then the maximum of ϕ in $U(r,c)$ is attained at an extreme point, say X, of $U(r,c)$.

By Theorem 1,

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}(\log a_{ij} - \log b_{ij}) > 0,$$

and so $\phi(X) > 0$, which completes the proof.

COROLLARY 11. *Let* $A \in \Omega_n$, *and let* B *be a distinct nonnegative* $n \times n$ *matrix with* $\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} = n$. *Then there exists* $\sigma \in S_n$ *such that* $d_\sigma(A) > d_\sigma(B)$.

Proof. The result follows from Theorem 10 and the well-known fact that the extreme points of Ω_n are precisely the $n \times n$ permutation matrices.

COROLLARY 12. *Let* A *be an* $m \times n$ *positive matrix, let* $(x,y) \in P^m \times P^n$, $(u,v) \in P^m \times P^n$, *and suppose* $f(x,y) = f(u,v)$, *where* f *is the map defined in Section 2. Then* $(x,y) = (u,v)$.

Proof. Let B, C be the $n \times n$ matrices where

$$b_{ij} = \frac{a_{ij}x_i y_j}{\sum_i \sum_j a_{ij}x_i y_j}, \quad c_{ij} = \frac{a_{ij}u_i v_j}{\sum_i \sum_j a_{ij}u_i v_j} \qquad \text{for all} \quad i, j.$$

Then B and C are both in $U(\tilde{x}, \tilde{y})$. If $B \neq C$, then by Theorem 10, there exist extreme points W and Z of $U(\tilde{x}, \tilde{y})$ such that

$$\prod_{i,j} b_{ij}^{w_{ij}} > \prod_{i,j} c_{ij}^{w_{ij}} \quad \text{and} \quad \prod_{i,j} b_{ij}^{z_{ij}} < \prod_{i,j} c_{ij}^{z_{ij}}.$$

But then

$$\prod_{i,j} \left(\frac{b_{ij}}{c_{ij}} \right)^{w_{ij}} = \prod_{i,j} \left(\frac{b_{ij}}{c_{ij}} \right)^{z_{ij}},$$

which is a contradiction. Therefore $B = C$, and it follows that $(x, y) = (u, v)$.

5. CIRCUIT PRODUCTS

If C is an $n \times n$ 0-1 matrix, it will be called a circuit matrix if there is a sequence of distinct integers $i_1, i_2, \ldots, i_k$ such that C has 1's at positions $(i_1, i_2), (i_2, i_3), \ldots, (i_{k-1}, i_k), (i_k, i_1)$ and zeros elsewhere.

If A is an $n \times n$ nonnegative matrix and if C is an $n \times n$ circuit matrix, then $\prod_{i,j} a_{ij}^{c_{ij}}$ will be called a circuit product of A corresponding to C. We will state two lemmas which are mere reformulations of known results in graph theory. The proofs may be given along similar lines to Berge [6].

Lemma 13. *If $A \in U(r, r)$ for some $r > 0$, then A has a positive circuit product.*

Lemma 14. *Let A be an $n \times n$ nonnegative matrix with positive row sums $r_1, \ldots, r_n$ and column sums $c_1, \ldots, c_n$. Then A can be written as a linear combination of circuit matrices with positive coefficients if and only if $r_i = c_i$, $i = 1, 2, \ldots, n$.*

The next result is analogous to Corollary 11.

Theorem 15. *Let A, B be distinct nonnegative $n \times n$ matrices such that $A \in U(r, r)$, $r > 0$, and $\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} = \sum_{i=1}^{n} r_i$. Then a circuit product of A exceeds the corresponding circuit product of B.*

Proof. By Theorem 1,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(\log a_{ij} - \log b_{ij}) > 0,$$

while by Lemma 14,

$$A = \sum_{i=1}^{k} \alpha_i C^i,$$

where $\alpha_i > 0$ and C^i is a circuit matrix, $i = 1, 2, \ldots, k$. Now the result follows.

REFERENCES

1 F. A. Atkinson, G. A. Watterson, and P. A. P. Moran, A matrix inequality, *Quart. J. Math. Oxford Ser. (2)*, 11:137–140 (1960).

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

3 R. Bapat, On permanents and diagonal products of doubly stochastic matrices, Ph.D. thesis, Univ. of Illinois at Chicago, 1981.

4 R. Bapat, $D_1 A D_2$ theorems for multidimensional matrices, *Linear Algebra Appl.* 48:437–442 (1982).

5 L. E. Baum and J. A. Eagon, An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology, *Bull. Amer. Math. Soc.* 73:360–363 (1967).

6 C. Berge, *Graphs and Hypergraphs*, North Holland, Amsterdam, 1973.

7 L. I. Bregman, Certain properties of nonnegative matrices and their permanents, *Dokl. Akad. Nauk SSSR* 211:27–30 (1973).

8 R. A. Brualdi, Convex sets of nonnegative matrices, *Proc. Amer. Math. Soc.* 15:671–674 (1968).

9 D. London, On matrices with doubly stochastic pattern, *J. Math. Anal. Appl.* 34:648–652 (1971).

10 A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic, 1979.

11 C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, 1973.

12 O. J. Rothaus, Study of the permanent conjecture and some of its generalizations, *Israel J. Math* 18:75–96 (1974).

Received 10 September 1984; revised 15 May 1985