An Extension of a Theorem of Darroch and Ratcliff in Loglinear Models and Its Application to Scaling Multidimensional Matrices

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

and

T. E. S. Raghavan
Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago
Chicago, Illinois 60680

Dedicated to Alan J. Hoffman on the occasion of his 65th birthday.

Submitted by Uriel G. Rothblum

ABSTRACT

Let $C=(c_{ij})$ be an $m\times n$ matrix with real entries. Let b be any nonzero m-vector. Let $K=\{\pi: C\pi=b,\ \pi\geqslant 0\}$ be bounded. Let $x=(x_1,x_2,\dots,x_n)$, $y=(y_1,y_2,\dots,y_n)$ be two nonnegative vectors with $y\in K$ and $x_j=0 \Leftrightarrow y_j=0$ for any coordinate j. Then it is shown that there exists a $\pi\in K$ and positive numbers $z_1,z_2,\dots,z_m$ such that $\pi_j=x_j\prod_{i=1}^m z_i^{c_{ij}}$ for all j. This theorem slightly generalizes a theorem of Darroch and Ratcliff in loglinear models with a completely different proof technique. The proof relies on an extension of a topological theorem of Kronecker to set valued maps and the duality theorem of linear programming. Many theorems in scaling of matrices and multidimensional matrices are direct consequences of this theorem. The main idea is to associate a suitable zero-one matrix of transportation with any multidimensional matrix. Some motivations for scaling applications are also discussed.

INTRODUCTION

Given a matrix $A=(a_{ij})_{m\times n}$ and given positive numbers $x_1,x_2,\dots,x_m$, $y_1,y_2,\dots,y_n$, the matrix $B=(a_{ij}x_iy_j)_{m\times n}$ is called a scaling of A. Under scaling the zero pattern of the entries remains unchanged. Often, the

655 Avenue of the Americas, New York, NY 10010 0024-3795/89/$3.50

problem is to select scales such that the matrix B lies in a convex set K. In many practical problems, the convex set is a polyhedron, such as $K = \{B: B = (b_{ij}),\ b_{ij} \geqslant 0,\ \Sigma_j b_{ij} = r_i,\ \Sigma_i b_{ij} = c_j\}$. Such a situation occurs in budget allocation problems [19]. Here A represents the past budget and B represents the unknown future budget with certain restrictions on the row sums and column sums. We have a similar situation in computed tomography and image reconstruction [12, 14]. Suppose the three dimensional matrix $A = (a_{ijk})$ is a fine grid discretization of the tumor intensity at location (i, j, k). After unit time, the growth of tumor results in a new matrix $B = (b_{ijk})$. The matrix A could have been observed during an operation. However, the matrix B could only be partially observed after the operation. For example, an X-ray mechanism might measure the intensity $\Sigma_j \Sigma_k b_{ijk} = r_i, \dots, \Sigma_i \Sigma_k b_{ijk} = s_j$, etc. In many situations it could also measure along diagonal rays such as $\Sigma\Sigma_{j+k=c} b_{ijk}$ for each given i and c, etc. Sometimes the linear constraints might themselves fix a unique matrix B. Otherwise one may use an extraneous criterion to select a B.

In the statistical literature the least squares principle is used as an extraneous criterion [8]. In engineering and other literature one uses the so-called entropy maximization principle [4, 11]. Sometimes a characteristic principle very specific to the problem at hand is also used to select a B. See [7, 8, 10, 13].

Scaling the data matrix before using an algorithm is a standard practice in numerical analysis [25]. For example, when using Gaussian elimination to solve the matrix equation $Ax = b$, the numerical algorithm might accumulate roundoff errors, and the final solution might be far removed from the solution to any matrix equation $\tilde{A}x = b$, where $\tilde{A}$ is near A. If the entries are too scattered, it has been observed that roundoff errors become significant. One possible solution is to consider a new matrix equation $Cy = d$, where $C = (a_{ij}\xi_i\eta_j)$ for some scales $\xi_1, \xi_2, \dots, \xi_n$ and $\eta_1, \eta_2, \dots, \eta_n$, so that the entries of C are less scattered. Here y and d are scaled versions of x and b.

Scaling problems were first considered by Sinkhorn [21] (also see [22]) to estimate the transition probabilities of a finite Markov chain. Bacharach [1] used them to correct for changing input-output coefficients of a Leontief system. Nonlinear operator approaches were initiated by Brualdi, Parter, and Schneider [6]. Recently, efficient algorithms have been proposed by Parlet and Landis [17] and Broyden and Raghavan [5].

Bapat [3] and independently Raghavan [18] considered the problem of scaling multidimensional matrices to achieve specific one dimensional marginals. Their proofs are quite different from each other, one relying on Kronecker's topological theorem and the other on the duality theorem of linear programming. Neither proof could handle the general case allowing for zero entries. The following proof combines the two methods to furnish a proof which takes care of the general case with zero entries. The proof actually extends the scaling of matrices to a more general setting. In fact we will slightly generalize a theorem of Darroch and Ratcliff in loglinear models. Its proof technique relies on extending the topological theorem of Kronecker to set valued maps and also applying the duality theorem of linear programming to generalized transportion matrices.[1] The following is the main

THEOREM 1. *Let $C=(c_{ij})$ be a real $m\times n$ matrix. Let b be any nonzero m-vector. Let $K=\{\pi : C\pi=b,\ \pi\geqslant 0\}$ be bounded. Let $x=(x_1,\dots,x_n)$, $y=(y_1,\dots,y_n)$ be two nonnegative vectors with the same zero pattern, that is, $x_j=0 \Leftrightarrow y_j=0$ for any coordinate j. If $y\in K$, then there exists a $\pi\in K$ where for some $z_1>0,\ z_2>0,\dots,z_m>0$*

$$\pi_j = x_j \prod_{i=1}^{m} z_i^{c_{ij}}, \qquad j=1,2,\dots,n.$$

Before we prove Theorem 1 we will briefly indicate how one can derive the following theorems of Brown [4], Sinkhorn [21], Bacharach [1], Menon [15], and many others [6, 3, 18].

THEOREM 2. *Let $x=(x_{ij})$, $Y=(y_{ij})$ be two $r\times s$ matrices with nonnegative entries. Let $x_{ij}=0 \Leftrightarrow y_{ij}=0$ for any ij. Let the row sums and column sums of Y be positive. Then there exist $u_1,u_2,\dots,u_r,v_1,v_2,\dots,v_s$, all positive, such that $D=(d_{ij})=(x_{ij}u_iv_j)$ has the same row sums and column sums as Y.*

The proof becomes free of notational confusion when we look at a simple situation, say for $r=2$, $s=3$. Let r_1, r_2 be the row totals and c_1, c_2, c_3 the column totals of Y. We will associate with the problem the $(0,1)$ matrix C with rs $(=6)$ columns and $r+s$ $(=5)$ rows. Here

$$C = \begin{array}{c} \\ u_1 \\ u_2 \\ v_1 \\ v_2 \\ v_3 \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} y_{11} & y_{12} & y_{13} & y_{21} & y_{22} & y_{23} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array} \quad \begin{array}{c} b \\ \begin{pmatrix} r_1 \\ r_2 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} \end{array}.$$

[1] Recently Franklin and Lorenz [9] and independently Rothblum [20] have given alternative proofs of this theorem.

Now the set $K = \{\pi : C\pi = b,\ \pi \geqslant 0\}$ is nonempty, as it contains $Y = (y_{ij})$. Since K is bounded, Theorem 1 is applicable. For example if the positive variables u_1, u_2, v_1, v_2, v_3, are associated with the rows of C, then we have a $\pi \in K$ such that, say,

$$\pi_{12} = x_{12} u_1^1 u_2^0 v_1^0 v_2^1 v_3^0 = x_{12} u_1 v_2.$$

In general we can see $(\pi_{ij}) = (x_{ij} u_i v_j)$. This completes the proof of the theorem.

THEOREM 3. *Let $X = (x_{ijk})$, $Y = (y_{ijk})$ be two $r \times s \times t$ matrices with nonnegative entries and with the same zero pattern (i.e. $x_{ijk} = 0 \Leftrightarrow y_{ijk} = 0$). If the one dimensional marginals $\sum_j \sum_k y_{ijk} = r_i$, $\sum_i \sum_k y_{ijk} = c_j$, $\sum_i \sum_j y_{ijk} = t_k$ are all positive, then there exist positive scalars $u_1, u_2, \dots, v_r$, $v_1, v_2, \dots, v_s, w_1, \dots, w_t$ such that $(\pi_{ijk}) = (x_{ijk} u_i v_j w_k)$ has the same one dimensional marginals as Y.*

The proof is more or less identical to the previous proof. Here we extend the transportation $(0,1)$ matrix to satisfy the one-dimensional marginal requirements. For example, in a case (say) with $r = 2$, $s = 3$, $t = 2$, the column corresponding to (say) y_{231} in C will be $(0,1,0,0,1,1,0)^T$. Thus by Theorem 1

$$\pi_{231} = x_{231} u_1^0 u_2^1 v_1^0 v_2^0 v_3^1 w_1^1 w_2^0 = x_{231} u_2 v_3 w_1.$$

In general $(\pi_{ijk}) = (x_{ijk} u_i v_j w_k)$. This shows that Theorem 3 is a direct consequence of Theorem 1.

REMARK. Bapat [3] and independently Raghavan [18] proved the above theorem, only for the special case when all the entries x_{ijk} and y_{ijk} are positive in X and Y respectively.

We will need the following lemmas for the proof of Theorem 1.

LEMMA 1. *Let $K = \{x \in \mathbf{R}^n : x \geqslant 0,\ Ax = b\}$, where A is a real $m \times n$ matrix and b a real m-vector. Let K be bounded. Let $g: K \to K$ be a continuous map such that for any $x \in K$, $x_i = 0 \Rightarrow g_i(x) = 0$. Then g maps K onto K.*

The following proof is reproduced from Bapat's thesis [2] for the sake of convenience.

Proof. We will assume that for any $i \in \{1,2,\ldots,n\}$, there exist x and y in K such that $x_i \neq y_i$. For otherwise we may ignore the ith coordinate and work with $\mathbf{R}^{n-1}$. Let ∂K be the relative boundary of K. That is, $x \in \partial K$ iff for any open ball B in $\mathbf{R}^n$ containing x, the set $B \cap (x: Ax = b\}$ has points in K as well as points not in K. We will show that ∂K is precisely the set of points in K with at least one zero coordinate. Suppose $x \in \partial K$. If K has just one point, the theorem is trivial. Otherwise let x^0 be a relative interior point of K. Points of the type $x^0 + \lambda(x - x^0)$ are in K for $0 \leqslant \lambda \leqslant 1$. If for some $\tilde{\lambda} > 1$ we have $\tilde{x} = x^0 + \tilde{\lambda}(x - x^0) \in K$, then x belongs to the line segment joining $\tilde{x}$ and the interior point x^0, and therefore it is a relative interior point, a contradiction. Thus if $\lambda > 1$ then $x^0 + \lambda(x - x^0) \notin K$. Now $A[x^0 + \lambda(x - x^0)] = b$, and $x^0 + \lambda(x - x^0)$ has a negative coordinate for all $\lambda > 1$. Thus x has a zero coordinate.

Conversely suppose $x \in K$ and $x_i = 0$ for some i. If the columns of A are independent, then $K = \{x\}$. Otherwise $Ay = 0$ for some $y \in R^n$. Further, we may assume that $y_i > 0$. For if $y_i = 0$ whenever $Ay = 0$, then for any two points x, z in K, we have $x_i = z_i$, contradicting the assumption made in the beginning. Now let B be an open ball in $\mathbf{R}^n$ containing x. For any small $\epsilon > 0$, the vector $x - \epsilon y$ belongs to $B \cap \{x: Ax = b\}$ and it has its ith coordinate negative. Thus $x \in \partial K$.

Now for any $t \in [0,1]$, define $\phi_t: K \to K$ by $\phi_t(x) = (1-t)g(x) + tx$. Then ϕ_t maps ∂K into itself. Also $\phi_0(x) = g(x)$ and $\phi_1(x) = x$ for all $x \in \partial K$. Thus the map $g: \partial K \to \partial K$ is homotopic to the identity map. Now by a theorem of Kronecker [16] it follows that g maps K onto K. This completes the proof. ∎

The next lemma, which generalizes the above lemma to set valued maps, is needed in the sequel.

LEMMA 2. *Let $K = \{x: x \geqslant 0,\ Ax = b,\ x \in \mathbf{R}^n\}$ be a convex polyhedra as in Lemma 1. Let $\Phi: K \to 2^K$ be a point to set map satisfying the following conditions*:

(i) *For each $x \in K$, $\Phi(x)$ is a nonempty closed bounded convex subset of K.*

(ii) *The map Φ is upper semicontinuous, i.e., if $x^{(n)} \to x$, $y^{(n)} \in \Phi(x^n)$, and $y^{(n)} \to y$, then $y \in \Phi(x)$.*

(iii) *If the jth coordinate x_j of x is zero, then for any $y \in \Phi(x)$ the jth coordinate $y_j = 0$.*

Then for any $u \in K$, there exists a $z^* \in K$ such that $u \in \Phi(z^*)$.

Proof. Let $P_k, k = 1, 2, \dots$ be a fine mesh of simplicial partitions of the polyhedron K with the following properties:

(i) If $T \in \mathbf{P}_k$, T is a simplex.

(ii) For each k, the number of elements (simplexes) in $\mathbf{P}_k$ is finite.

(iii) If $T \in \mathbf{P}_k$ and if F is a proper face of T, then $F \in \mathbf{P}_k$.

(iv) If $T_1 \in \mathbf{P}_k$, $T_2 \in \mathbf{P}_k$, then $T_1 \cap T_2$ is a face of T_1 and T_2 or $T_1 \cap T_2 = \varnothing$.

(v) $\bigcup_{T \in \mathbf{P}_k} T = K$.

(vi) If $T \in \mathbf{P}_k$ and x, y are vertices of T, then $\|x - y\| < 1/k$. (Here $\|\cdot\|$ is the Euclidean norm.)

(vii) If $X^{(k)}$ is the collection of all vertices of the simplexes in $\mathbf{P}_k$, then $X^{(k)} \subset X^{(k+1)} \subset \cdots$, $k = 1, 2, \dots$.

(viii) $X^{(1)} \supseteq$ the set of all extreme points of K.

We know that such a sequence of finer and finer simplicial partitions exists for any polyhedron K [21]. For each $x \in X^{(k)}$ choose any $y \in \Phi(x)$, and define a single valued mapping as follows:

$$f_k(x) = y \qquad \text{if} \quad x \in X^{(k)}$$

$$f_k\left(\sum_{i=1}^{r} \lambda_i x_i^{(k)} \right) = \sum_{i=1}^{r} \lambda_i f_k\left(x_i^{(k)}\right) \qquad \text{if} \quad x_1^{(k)}, x_2^{(k)}, \dots, x_r^{(k)}$$

are the vertices of a smallest dimensional simplex T containing $x = \sum_{k=1}^{r} \lambda_i x_i^{(k)}$, where $\lambda_i > 0$, $\sum_i \lambda_i = 1$. By assumption (iv), the smallest dimensional simplex in $\mathbf{P}^{(k)}$ containing x is unique. Clearly $f_k \colon K \to K$ is continuous. We claim $f_k \colon \partial K \to \partial K$. From Lemma 1 it is sufficient to prove that when $u = (u_1, \dots, u_n) \in K$ and $u_j = 0$, one has $(f_k(u))_j = 0$. Let $u = \sum_{i=1}^{r} \lambda_i x_i^{(k)}$, where $x_1^{(k)}, x_2^{(k)}, \dots, x_r^{(k)}$ are the vertices of the smallest dimensional simplex containing u. Since $u_j = 0$ and $x_i^{(k)} \geqslant 0$, the jth coordinate of each vertex $x_i^{(k)}$ is zero. By the conditions imposed on the upper semicontinuous map $\Phi(x)$, the jth coordinate of any $y \in \Phi(x_i^k)$, and in particular of $f_k(x_i^{(k)})$, is zero. Thus for $u = \sum \lambda_i x_i^{(k)}$,

$$\left(f_k\left(\sum_{i=1}^{r} \lambda_i x_i^{(k)} \right) \right)_j = \sum \lambda_i \left(f_k\left(x_i^{(k)}\right) \right)_j = 0.$$

This proves our claim that $f_k \colon \partial K \to \partial K$. As in Lemma 1, f_k is homotopic to the identity map. Now given any $u \in K$, we have, by Lemma 1, some $z^{(k)} \in K$ such that $u = f_k(z^{(k)})$. Without loss of generality, $z^{(k)} \to z^*$. Since

$u = f_k(z^k) \in \Phi(z^{(k)})$ for all k and $z^{(k)} \to z^*$, we have $u \in \Phi(z^*)$, by the upper semicontinuity of Φ. Hence the lemma. ∎

We are now ready to prove Theorem 1.

Proof of Theorem 1. Let $S = \{\pi : C\pi = b, \pi \geqslant 0, \pi_j = 0$ whenever $y_j = 0$ for any $j\}$. Since $y \in S$, the closed bounded convex set S is nonempty. For any $h \in S$ define

$$p_j(h) = \begin{cases} \log(x_j/h_j) & \text{if} \quad x_j > 0, \quad h_j > 0, \\ 0 & \text{if} \quad x_j = 0 \quad (\text{in this case } h_j = 0 \text{ when } h \in S), \\ \log\infty & \text{if} \quad x_j > 0, \quad h_j = 0. \end{cases}$$

Define $0\log 0 = 0$, $0 \log\infty = 0$, $0 + \infty = \infty$. Let

$$\varphi(h) = \left\{\pi^* : (p, \pi^*) = \min_{\pi \in S} (p, \pi)\right\}.$$

We claim that $\varphi(h)$ is well defined and nonempty. Since $b \neq 0$ and $y \in S$, we have $y \neq 0$ and therefore $x \neq 0$. Let $\delta = \min\{x_j : x_j > 0\}$. Since S is bounded, for any $\pi \in S$ we have $0 \leqslant \pi_j < L$ for all coordinates j for some L. Therefore $p_j \geqslant \log(\delta/L)$. However, $\log(\delta/L) < 0$. If $p_j \geqslant 0$, then $p_j\pi_j \geqslant 0 \geqslant L\log(\delta/L)$. If $p_j < 0$ then $0 < -p_j < -\log(\delta/L)$ and $0 \leqslant \pi_j < L$, which implies $-p_j\pi_j < -L\log(\delta/L)$. Thus in all cases, $p_j\pi_j > L\log(\delta/L)$ and in particular $(p, \pi) > nL\log(\delta/L)$. Let $\alpha = \inf_{\pi \in S}(p, \pi)$. Clearly $\alpha \leqslant \sum_j h_j \log(x_j/h_j) < N$ for some N. [Here we notice that if $h_j = 0$, $h_j\log(x_j/h_j) = 0$. If $x_j = 0$, $p_j = 0$ by assumption.] Thus the infinum α is finite. Let $(p, \pi^r) \to \alpha$. Without loss of generality, $\pi^r \to \pi^*$. If $p_j = \infty$, then $\pi_j^r \equiv 0$ for all r, and $\pi_j^* = 0$. If p_j is finite, $p_j\pi_j^r \to p_j\pi_j^*$. Thus $(p, \pi^*) = \inf(p, \pi)$. This shows that $\varphi(h) \neq \varnothing$. More importantly, if $h_j = 0$ and $\xi \in \varphi(h)$, then $\xi_j = 0$; for otherwise if $\xi_j > 0$, then $x_j > 0$ and $p_j = \infty$ with $(p, \xi) = \infty$, a contradiction to the assumption that $\xi \in \varphi(h)$. Next, we will check that φ is upper semicontinuous. Let $h^r \to h^*$ and $\pi^r \in \varphi(h^r)$, $\pi^r \to \pi^*$. Suppose $h_j^* = 0$. Then $\pi_j^* = 0$, for otherwise $\pi_j^* > 2\epsilon > 0$. Therefore $\pi_j^r \geqslant \epsilon$ for large r. When $\pi_j^* > 0$, we have $x_j > 0$ and $p_j^r = \log(x_j/h_j^r) \to \infty$. Thus $p_j^r\pi_j^r \to \infty$. For any other i we know that $p_i^r\pi_i^r \geqslant L\log(\delta/L)$ and so $(p^r, \pi^r) \to \infty$. However, $\pi^r \in \varphi(h^r)$ and $(p^r, \pi^r) \leqslant (p^r, h^r) \leqslant N$ for any r. This contradiction shows that when $h_j^* = 0$, π_j^* is also zero. If $h_j^* > 0$, then $p_j^r\pi_j^r \to p_j^*\pi_j^*$ [here $p_j^* = \log(x_j/h_j^*)$]. Thus in all cases $(p^r, \pi^r) \to (p^*, \pi^*)$. Also, for any $\pi \in S$, $(p^r, \pi) \to (p^*, \pi)$ as an extended real valued function. Now

$(p^r, \pi^r) \leqslant (p^r, \pi)$ for all $\pi \in S$. Thus $(p^*, \pi^*) = \lim_r (p^r, \pi^r) \leqslant \lim_r (p^r, \pi) = (p^*, \pi)$ for all $\pi \in S$. This shows that $\pi^* \in \varphi(h^*)$ and $\varphi(h^*)$ is upper semicontinuous. Next we will check that $\varphi(h)$ is closed. Let $\pi^r \in \varphi(h)$ and $\pi^r \to \pi^0$. If $p_j = 0$ then $p_j \pi_j^r \to p_j \pi_j^0$. If $p_j = \log \infty$, then $\pi_j^r \equiv 0$ in r and $p_j \pi_j^r \equiv 0$ in r by our definition $(0 \log \infty = 0)$. Thus $(p, \pi^r) \to (p, \pi^0)$ and $\pi^0 \in \varphi(h)$. Convexity is trivially satisfied.

Continuing with our proof, the map $\varphi : S \to 2^S$ satisfies the conditions of Lemma 2. Therefore, there exists an $h^* \in S$ such that $y \in \varphi(h^*)$. Let $p^* = p(h^*)$ be the associated vector. We have y as an optimal solution to the linear programming problem

$$\min \sum_j p_j^* \pi_j$$

subject to

$$\sum_{j=1}^{m} c_{ij} \pi_j = b_i, \qquad i = 1, 2, \ldots, m,$$

$$\sum_{j=1}^{n} c_{m+1, j} \pi_j = 0 \qquad \left(\text{here } c_{m+1, j} = 0 \text{ if } y_j > 0 \text{ and } c_{m+1, j} = 1 \text{ if } y_j = 0\right),$$

$$\pi_j \geqslant 0, \qquad j = 1, 2, \ldots, n.$$

By the duality theorem of linear programming [24], we have an optimal solution to the dual satisfying complementary slackness conditions. Namely, there is an optimal solution to

$$\max \sum_{i=1}^{m} b_i w_i$$

subject to

$$\sum_{i=1}^{m} c_{ij} w_i + c_{m+1, j} w_{m+1} \leqslant p_j^*, \qquad j = 1, 2, \ldots, n. \qquad (**)$$

If $(w_1^*, \dots, w_{m+1}^*)$ is optimal for the dual, then the dual inequalities of the above step are exact equalities whenever $y_j > 0$. When $y_j > 0$, $c_{m+1,j} = 0$ and we have

$$\sum_{i=1}^{m} c_{ij} w_i = p_j^* = \log \frac{x_j}{h_j^*}.$$

Clearly since $y \in \varphi(h^*)$, when $h_j^* = 0$ we have $y_j = 0$, and therefore when $x_j > 0$, we have $y_j > 0$ and also $h_j^* > 0$. This shows that p is finite when $x_j > 0$. If $\exp(-w_i) = z_i$, then taking exponents on both sides of the equation $\sum_i c_{ij} w_i = \log(x_j / h_j^*)$, we get $h_j^* = x_j \prod_i z_i^{c_{ij}}$ for all $x_j > 0$. When $x_j = 0$, the equation trivially holds. This completes the proof of the theorem. ∎

Remark 2. The condition that x and y have the same zero pattern cannot be dispensed with. The following is a counterexample. Let x be the matrix $x = (x_{ij})$ given by

$$x = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}.$$

Let Y be any doubly stochastic matrix. For any scale factors $u_i > 0$, $v_j > 0$, $i, j = 1,2,3$, the matrix $H = (x_{ij} u_i v_j)$ has the same zero pattern and hence the permanent of H is zero. However, any doubly stochastic matrix has positive permanent. Thus H can never be doubly stochastic.

Remark 3. Our main theorem is a sharpening of a theorem of Darroch and Ratcliff [7]. In their theorem they assume $x_i > 0$ for all i. (In their notation $\pi_i > 0$ for all i.) Our theorem allows for zero entries as well. In the problem of scaling matrices, zero entries have always been somewhat thorny to handle. While for the positive case the proof of the theorem by Darroch and Ratcliff is constructive, for the general case our proof is nonconstructive. For additional computationally efficient scaling algorithms for multidimensional matrices see Broyden and Raghavan [5].

Remark 4. In image reconstruction and computed tomography one measures, via an X-ray machine, the total intensity along rays. Thus for the three dimensional nonnegative matrices $A = (a_{ijk})$ and $B = (b_{ijk})$ with the same zero pattern, if all the two dimensional marginals such as $\sum_k b_{ijk} = f_{ij}$, $\sum_j b_{ijk} = g_{ik}$, $\sum_i b_{ijk} = b_{jk}$ are specified for each pair (i, j), (j, k), (i, k), then the same proof of Theorem 1 gives positive matrices (u_{ij}), (v_{jk}), (w_{ik}) such that the three dimensional matrix $h_{ijk} = a_{ijk}u_{ij}v_{jk}w_{ik}$ satisfies the same two dimensional marginals as B. In particular H also has the same one dimensional marginals as B. The proof involves the same technique as in Theorem 2. One associates a $(0,1)$ matrix C with as many columns as the number of entries in B. The rows correspond to possible two dimensional marginals. With indices (i, j) a variable u_{ij} is associated. Similarly v_{jk}, w_{ik} are introduced. The extension to higher dimensions with lower dimensional marginals is automatic.

We wish to thank Professor R. A. Brualdi for some useful conversations we had with him on this problem.

REFERENCES

1 M. Bacharach, Estimating nonnegative matrices from marginal data, *Internat. Econom. Rev.* 6:294–310 (1965).

2 R. B. Bapat, On Permanents and Diagonal Products of Doubly Stochastic Matrices, Ph.D. Thesis, Univ. of Illinois at Chicago, 1981.

3 R. B. Bapat, $D_1 {}_A D_2$ theorems for multidimensional matrices, *Linear Algebra Appl.* 48:437–442 (1982).

4 D. T. Brown, A note on approximations of discrete probability distributions, *Inform. and Control* 2:386–392 (1959).

5 C. Broyden and T. E. S. Raghavan, Scaling algorithms for multidimensional matrices, manuscript, 1987.

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

7 J. N. Darroch and D. Ratcliff, Generalized iterative scaling for log-linear models, *Ann. Math. Statist.* 43:1470–1480 (1972).

8 S. E. Fienberg, *The Analysis of Cross Classified Data*, M.I.T. Press, Cambridge, Mass., 1977.

9 J. Franklin and J. Lorenz, On the scaling of multidimensional matrices, preprint.

10 D. V. Gokhale, An iterative procedure for analyzing log-linear models, *Biometrics* 27:681–687 (1971).

11 I. J. Good, Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, *Ann. Math. Statist.* 34:911–934 (1963).

12 G. T. Herman and A. Lint, Iterative reconstruction algorithms, *Comput. Biol. Med.* 6:276 (1976).

13 C. T. Ireland and S. Kullback, Contingency tables with given marginals, *Biometrika* 55:179–188 (1968).

14 A. V. Lakshminarayanan and A. Lent, Methods of least squares and SIRT in reconstruction, *J. Theoret. Biol.* 76:267–295 (1979).

15 M. V. Menon, Matrix links, an extremization problem and the reduction of a nonnegative matrix to one with prescribed row and column sums, *Canad. J. Math.* 20:225–232 (1968).
16 J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic, New York, 1970.
17 B. N. Parlet and T. L. Landis, Methods for scaling to doubly stochastic form, *Linear Algebra Appl.* 48:53–70 (1982).
18 T. E. S. Raghavan, On pairs of multidimensional matrices, *Linear Algebra Appl.* 62:263–268 (1984).
19 T. E. S. Raghavan, On pairs of multidimensional matrices and their applications, *Contemp. Math.* 46:339–354 (1985).
20 U. Rothblum, Generalized scalings satisfying linear constraints, preprint.
21 R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.* 35:876–879 (1964).
22 R. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, *Pacific J. Math.* 21:343–348 (1967).
23 J. Stillwell, *Classical Topology and Combinatorial Group Theory*, Springer Verlag, New York, 1980.
24 H. M. Wagner, *Principles of Operations Research*, Prentice-Hall, Englewood Cliffs, N.J., 1975.
25 J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Her Majesty's Stationery Office, England, 1963.

Received 10 October 1987; final manuscript accepted 6 October 1988