Distributions with Fixed Marginals and Related Topics
IMS Lecture Notes – Monograph Series Vol. 28, 1996

IN QUEST OF BIRKHOFF'S THEOREM IN HIGHER DIMENSIONS

BY XIN LI, P. MIKUSIŃSKI,
H. SHERWOOD, AND M. D. TAYLOR
University of Central Florida

A doubly stochastic matrix is a non-negative function defined on $\{1,\ldots,m\}\times\{1,\ldots,m\}$ such that all row and column sums are 1. A hypermatrix is a non-negative function defined on a set of the form $X = \{1,\ldots,m_1\}\times\ldots\times\{1,\ldots,m_n\}$. A hypermatrix is called multiply stochastic if it satisfies a suitably generalized version of the row and column condition for ordinary doubly stochastic matrices. Note that this is a double generalization of doubly stochastic matrices: we not only consider higher dimensions but allow "non-square matrices". Our goal is to describe extremal multiply stochastic hypermatrices in terms of their support, in terms of transfer vectors, and as local minima of the entropy function and to characterize the set of such extremals for a certain class of $3\times3\times3$ hypermatrices.

1. Introduction. An $n \times n$ matrix (m_{ij}) is called doubly stochastic if $m_{ij} \geq 0$ for all $i, j = 1,\ldots,n$ and

$$\sum_{i=1}^{n} m_{ij} = 1 \quad \text{for} \quad j = 1,\ldots,n$$

and

$$\sum_{j=1}^{n} m_{ij} = 1 \quad \text{for} \quad i = 1,\ldots,n.$$

A doubly stochastic matrix is called extremal if it is an extremal element in the set of all doubly stochastic matrices. In other words, a doubly stochastic matrix is extremal if it cannot be represented as a convex combination of other doubly stochastic matrices. Garrett Birkhoff (1946) proved the following characterization of extremal doubly stochastic matrices. It is known that the theorem was later rediscovered independently by John von Neumann, but he has not published his proof.

AMS 1991 Subject Classification: Primary 15A51; Secondary 15A48

Key words and phrases: Doubly stochastic matrix, extremal point, multiply stochastic hypermatrix.

A doubly stochastic matrix is extremal if and only if it is a permutation matrix.

In this paper we are concerned with possible generalizations of Birkhoff's theorem to higher dimensions. This direction has already been suggested by Gian-Carlo Rota and L. H. Harper (1971). We will investigate one particular situation in detail and give some general theorems which bear upon this question. The proofs of the general theorems can be found in Li, Mikusiński, Sherwood, and Taylor (1993).

2. Basic Definitions. If n is a natural number, then $\langle n\rangle$ stands for the set $\{1, 2, \ldots, n\}$. By an $m_1 \times \ldots \times m_n$ *hypermatrix* we mean a function

$$A : \langle m_1\rangle \times \ldots \times \langle m_n\rangle \longrightarrow \Re.$$

We call the elements of $\langle m_1\rangle \times \ldots \times \langle m_n\rangle$ *cells*. The space of all $m_1 \times \ldots \times m_n$ hypermatrices is denoted by $\mathcal{H}(\langle m_1\rangle \times \ldots \times \langle m_n\rangle)$. If $A \in \mathcal{H}(\langle m_1\rangle \times \ldots \times \langle m_n\rangle)$ then by supp A we mean the support of A, that is the set of all $i \in \langle m_1\rangle \times \ldots \times \langle m_n\rangle$ such that $A(i) \neq 0$.

Let $S = \{k_1, \ldots, k_r\}$ be a proper nonempty subset of $\langle n\rangle$. Then

$$P_S : \langle m_1\rangle \times \ldots \times \langle m_n\rangle \longrightarrow \langle m_{k_1}\rangle \times \ldots \times \langle m_{k_r}\rangle$$

is the projection map defined by $P_S(i_1, \ldots, i_n) = (i_{k_1}, \ldots, i_{k_r})$. By an S-*row* or an $(n-r)$-*row* of $\langle m_1\rangle \times \ldots \times \langle m_n\rangle$ we mean a set of the form $P_S^{-1}(i)$ where $i \in \langle m_{k_1}\rangle \times \ldots \times \langle m_{k_r}\rangle$. By the S-*marginal* of a hypermatrix A we mean the function

$$S(A) : \langle m_{k_1}\rangle \times \ldots \times \langle m_{k_r}\rangle \longrightarrow \Re$$

defined by

$$S(A)(i) = \sum_{j \in P_S^{-1}(i)} A(j)$$

We will also say that $S(A)$ is an r-*marginal* of A. Note that we can write

$$S : \mathcal{H}(\langle m_1\rangle \times \ldots \times \langle m_n\rangle) \longrightarrow \mathcal{H}(\langle m_{k_1}\rangle \times \ldots \times \langle m_{k_r}\rangle).$$

We say that a hypermatrix A has a *uniform* S-marginal provided $S(A)$ is a constant function.

Let $\mathcal{M} = \langle m_1\rangle \times \ldots \times \langle m_n\rangle$ and $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a collection of nonempty subsets of $\langle n\rangle$. A hypermatrix $A \in \mathcal{H}(\langle m_1\rangle \times \ldots \times \langle m_n\rangle)$ is called *multiply stochastic* if

1. $A(i) \geq 0$ for all cells i,
2. $\sum_i A(i) = \Delta$,

3. $S_k(A)$ is a constant function for each $k = 1, \ldots, m$,

where Δ is a fixed positive constant.

We can assume $\Delta = 1$, but it is often more convenient to choose some other constant.

The set of all multiply stochastic hypermatrices $A \in \mathcal{H}(\langle m_1 \rangle \times \ldots \times \langle m_n \rangle)$ with uniform S_1-$, \cdots, S_m$-marginals will be denoted by $\mathcal{H}(\mathcal{M}, \mathcal{S})$. Note that $\mathcal{H}(\mathcal{M}, \mathcal{S})$ is a convex set. A hypermatrix $A \in \mathcal{H}(\mathcal{M}, \mathcal{S})$ is called *extremal* in $\mathcal{H}(\mathcal{M}, \mathcal{S})$ if it cannot be represented as a convex combination of elements of $\mathcal{H}(\mathcal{M}, \mathcal{S})$ different from A.

THEOREM 1. *A hypermatrix A is extremal in $\mathcal{H}(\mathcal{M}, \mathcal{S})$ if and only if $B = A$ whenever $B \in \mathcal{H}(\mathcal{M}, \mathcal{S})$ and* $\operatorname{supp} B = \operatorname{supp} A$.

PROOF. Let $A \in \mathcal{H}(\mathcal{M}, \mathcal{S})$ be extremal. Suppose $B \in \mathcal{H}(\mathcal{M}, \mathcal{S})$ and $\operatorname{supp} B = \operatorname{supp} A$. Then there exists a negative number λ such that $C = (1-\lambda)A + \lambda B \in \mathcal{H}(\mathcal{M}, \mathcal{S})$. But then

$$A = \frac{1}{1-\lambda} C + \frac{-\lambda}{1-\lambda} B,$$

which is a contradiction, unless $B = A$.

Now assume that A is the only element of $\mathcal{H}(\mathcal{M}, \mathcal{S})$ with the given support. Suppose that $A = \lambda_0 B + (1-\lambda_0)C$, where $0 < \lambda_0 < 1$, and $B, C \in \mathcal{H}(\mathcal{M}, \mathcal{S})$. Note that for every λ in some neighborhood of λ_0 we have $\lambda B + (1-\lambda)C \in \mathcal{H}(\mathcal{M}, \mathcal{S})$ and $\operatorname{supp}(\lambda B + (1-\lambda)C) = \operatorname{supp} A$. If the support of B or C differs from the support of A, then $\lambda B + (1-\lambda)C$ is different for different λ. But this contradicts the assumption.

COROLLARY 2. *A hypermatrix A is not extremal in $\mathcal{H}(\mathcal{M}, \mathcal{S})$ if and only if it is possible to find a $B \in \mathcal{H}(\mathcal{M}, \mathcal{S})$ such that* $\operatorname{supp} B$ *is a proper subset of* $\operatorname{supp} A$.

3. Transfer Vectors. Birkhoff's original theorem can be easily proved by regarding each entry of the matrix as the amount of "mass" in that particular cell and then carefully shifting mass from one cell to another in such a way as to maintain the uniformity of the marginals. We now generalize this idea of possible "mass transfer" to higher dimensions.

Let $n \in \mathcal{N}$ and let S be a proper subset of $\langle n \rangle$. By a *transfer vector in the S-direction with domain $\langle m_1 \rangle \times \ldots \times \langle m_n \rangle$* we mean a hypermatrix

$$T : \langle m_1 \rangle \times \ldots \times \langle m_n \rangle \longrightarrow \Re$$

satisfying the following:

There exist distinct cells i_0 and i_1 belonging to the same S-row in $\langle m_1 \rangle \times \ldots \times \langle m_n \rangle$ such that

$$T(i) = \begin{cases} 1 & \text{if } i = i_0, \\ -1 & \text{if } i = i_1, \\ 0 & \text{otherwise.} \end{cases}$$

Let Q be a nonempty subset of $\mathcal{M} = \langle m_1 \rangle \times \ldots \times \langle m_n \rangle$ and let S be a proper subset of $\langle n \rangle$. Let i be a cell in the range of P_S. By $\mathcal{T}(Q, S, i)$ we mean the set of all transfer vectors T for which there exist $j_0, j_1 \in P_S^{-1}(i) \cap Q$ such that

$$T(j) = \begin{cases} 1 & \text{if } j = j_0, \\ -1 & \text{if } j = j_1, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $T \in \mathcal{T}(Q, S, i)$ if it transfers mass from one cell of Q to another cell of Q along the S-row $P_S^{-1}(i)$. Let

$$\mathcal{T}(Q, S) = \bigcup_{i \in P_S(\mathcal{M})} \mathcal{T}(Q, S, i).$$

If $T \in \mathcal{T}(Q, S)$, then we say that T is a *transfer vector in the direction S between cells of Q*. We say that $T_1, \ldots, T_r$ is a *sufficient* set of transfer vectors between the cells of Q in the direction S provided $T_1, \ldots, T_r \in \mathcal{T}(Q, S)$ and every element of $\mathcal{T}(Q, S)$ can be written as a linear combination of $T_1, \ldots, T_r$. We will call $T_1, \ldots, T_r$ a *minimal* sufficient set provided $T_1, \ldots, T_r$ are linearly independent.

Let $S_0, S_1, \ldots, S_q$ be distinct, proper, nonempty subsets of $\langle n \rangle$. Set $\mathcal{S} = (S_0, S_1, \ldots, S_q)$. For the transfer vector

$$T : \langle m_1 \rangle \times \ldots \times \langle m_n \rangle \longrightarrow \Re$$

we define the *marginal transfer vector determined by $\mathcal{S}$* to be

$$\mathcal{S}(T) = (S_0(T), S_1(T), \ldots, S_q(T)).$$

Note that the transformation $T \rightarrow \mathcal{S}(T)$ is a linear transformation.

The following theorem is a very general description of how the idea of transferring mass between cells is used in determining extremality of matrices and hypermatrices both in our work and that of others.

THEOREM 3. *A hypermatrix A is not extremal in $\mathcal{H}(\mathcal{M}, \mathcal{S})$ if and only if there exists a minimal sufficient set of transfer vectors $T_1, \ldots, T_p$ in the direction S_0 between cells of the support of A such that $\mathcal{S}(T_1), \ldots, \mathcal{S}(T_p)$ are* *linearly dependent. Furthermore, if there is one minimal sufficient set with this property, then every minimal set of transfer vectors between the cells of the support of A in any direction S_i whatever must have this property.*

4. Local Minima of the Entropy. The following result stands on its own and somewhat in isolation from the other results of this paper.

THEOREM 4. *A hypermatrix A is extremal in $\mathcal{H}(\mathcal{M}, \mathcal{S})$ if and only if the function*

$$H(A) = -\sum_i A(i)\ln(A(i))$$

has a local minimum on $\mathcal{H}(\mathcal{M}, \mathcal{S})$ at A.

5. Distribution Matrices. In this section we restrict our consideration to $\mathcal{M} = \langle n\rangle^3$, that is to hypermatrices of the form

$$A : \langle n\rangle^3 \longrightarrow \Re.$$

We also assume that $\mathcal{S} = \mathcal{S}_2 = (\{1,2\}, \{1,3\}, \{2,3\})$. In other words, $\mathcal{H}(\mathcal{M}, \mathcal{S})$ is the collection of all $\langle n\rangle^3$-hypermatrices with uniform 2-marginals. For convenience we take Δ, the sum of all entries in A, to be n^2. This set of hypermatrices will be denoted by $\mathcal{H}(\langle n\rangle^3, \mathcal{S}_2)$.

Let $n \in \mathcal{N}$ be fixed. Let $P_1, \ldots, P_{n!}$ be all the $n \times n$ permutation matrices. A vector $(v_1, \ldots, v_{n!})$ is called a *covering vector* if $0 \le v_i \le 1$ and $\sum_{i=1}^{n!} v_i P_i$ is the matrix all of whose entries are 1.

A $n \times n!$ matrix $M = (m_{i,j})$ is called a *distribution matrix* if

1. $0 \le m_{i,j} \le 1$,
2. $\mathcal{C}(M) = (\sum_{i=1}^{n} m_{i,1}, \ldots, \sum_{i=1}^{n} m_{i,n!})$ is a covering vector,
3. $\sum_{j=1}^{n!} m_{i,j} = 1$ for $i = 1, \ldots, n$.

Note that for every multiply stochastic matrix $A \in \mathcal{H}(\langle n\rangle^3, \mathcal{S}_2)$ there exists a distribution matrix M such that the matrix $\sum_{j=1}^{n!} m_{k,j} P_j$ represents the k-th "floor" of A, which allows us, by a slight abuse of notation, to write

$$A = \mathcal{A}(M) = \begin{bmatrix} \sum_{j=1}^{n!} m_{1,j} P_j \\ \sum_{j=1}^{n!} m_{2,j} P_j \\ \vdots \\ \sum_{j=1}^{n!} m_{n,j} P_j \end{bmatrix}.$$

The mapping $\mathcal{A}$ assigns a hypermatrix to a distribution matrix.

An $n \times n!$ distribution matrix is called extremal if it is extremal in the set of all $n \times n!$ distribution matrices. Note that this is different from the extremality of a distribution matrix with fixed marginals. For example, one can check that the matrix

$$M = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

is extremal in the set of all 3×6 distribution matrices with fixed covering vector $(1/4, 1/4, 1/4, 3/4, 3/4, 3/4)$. On the other hand it is not extremal in the set of all 3×6 distribution matrices, because

$$M = \frac{1}{2}\begin{pmatrix} \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{7}{16} & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{8} & \frac{5}{8} & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{16} & \frac{13}{16} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{5}{16} & \frac{5}{16} & \frac{5}{16} & \frac{1}{16} & 0 & 0 \\ 0 & 0 & 0 & \frac{5}{8} & \frac{3}{8} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{16} & \frac{11}{16} \end{pmatrix}.$$

For this reason Lemma 7 in the next section requires a separate proof.

Notice that the set of all $n \times n!$ distribution matrices is a bounded convex set, the same is true for the corresponding set of hypermatrices, and $\mathcal{A}$ is a linear mapping from the first set onto the second. It follows that the inverse image of any extremal point must be a convex subset of the set of distribution matrices and contain an extremal point. Thus we have the following:

THEOREM 5. *Every hypermatrix extremal in $\mathcal{H}(\langle n\rangle^3, \mathcal{S}_2)$ has an extremal distribution matrix.*

6. Extremal Elements of $\mathcal{H}(\langle 3\rangle^3, \mathcal{S}_2)$. In this section we are going to use the developed tools to give a complete characterization of extremal elements of $\mathcal{H}(\langle 3\rangle^3, \mathcal{S}_2)$. First we name all 3×3 permutation matrices:

$$P_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} P_2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} P_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$P_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} P_5 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} P_6 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

LEMMA 6. *Let M be a distribution matrix for an element of $\mathcal{H}(\langle 3\rangle^3, \mathcal{S}_2)$. Then the covering vector of M has the form $(\alpha, \alpha, \alpha, 1-\alpha, 1-\alpha, 1-\alpha)$ for some $\alpha \in [0, 1]$.*

PROOF. Let $(v_1, \ldots, v_6)$ be a covering vector. Then

$$\begin{aligned} v_1 + v_4 &= 1 \\ v_4 + v_2 &= 1 \\ v_2 + v_5 &= 1 \\ v_5 + v_3 &= 1 \\ v_3 + v_6 &= 1 \\ v_6 + v_1 &= 1 \end{aligned}$$

Thus every solution is of the form $(\alpha, \alpha, \alpha, 1-\alpha, 1-\alpha, 1-\alpha)$. Since $0 \le v_i \le 1$ we must have $\alpha \in [0, 1]$.

LEMMA 7. *A distribution matrix is extremal if and only if it is uniquely determined by its support.*

PROOF. Let M be an extremal distribution matrix. Suppose M_1 is another distribution matrix such that supp M_1 =supp M. If $\mathcal{C}(M) = (1,1,1,0,0,0)$ or $\mathcal{C}(M) = (0,0,0,1,1,1)$, then the assertion follows from Theorem 1. Now assume that $\mathcal{C}(M) = (\alpha, \alpha, \alpha, 1-\alpha, 1-\alpha, 1-\alpha)$ with $0 < \alpha < 1$. As in the proof of Theorem 1, there exists a negative number λ such that every entry of $M_2 = (1-\lambda)M + \lambda M_1$ is nonnegative. Moreover, since $\mathcal{C}(M_2) = (1-\lambda)\mathcal{C}(M) + \lambda\mathcal{C}(M_1)$, we can make sure that $|\lambda|$ is small enough so that $\mathcal{C}(M_2)$ is still a covering vector. Now the proof can be finished as the proof of Theorem 1.

COROLLARY 8. *A distribution matrix A is not extremal if and only if it is possible to find a distribution matrix B such that* supp B *is a proper subset of* supp A.

Observe that a permutation among the first three columns, or the last three columns, or any permutation of the rows of a distribution matrix produces another distribution matrix. In what follows, when we say a *permutation of a distribution matrix* we mean a permutation in this restricted sense. Note that permutations of a distribution matrix induce permutations of the corresponding hypermatrix.

THEOREM 9. *Every extremal element of $\mathcal{H}(\langle 3\rangle^3, \mathcal{S}_2)$ is a permutation of a hypermatrix generated by a permutation of one of the following two distribution matrices:*

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

or

$$B = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

PROOF. Let M be an extremal element of $\mathcal{H}(\langle 3\rangle^3, \mathcal{S}_2)$. We will show that the assumption that M cannot be generated by a permutation of A or a permutation of B leads to a contradiction.

Let $C = (c_{i,j})$, $i = 1,2,3, j = 1,2,3,4,5,6$, be an extremal distribution matrix generating M. Since its covering vector is $(\alpha, \alpha, \alpha, 1-\alpha, 1-\alpha, 1-\alpha)$ with $0 < \alpha < 1$, every column of C must contain at least one nonzero element. Denote $C_1 = (c_{ij})$, $i = 1,2,3$, $j = 1,2,3$ and $C_2 = (c_{ij})$, $i = 1,2,3$, $j = 4,5,6$. Note that C_1 cannot contain a permutation matrix, because M is extremal and not generated by a permutation of A. Consequently, C has to be of the form:

$$C = \begin{pmatrix} \bullet & \bullet & ? & ? & ? & ? \\ ? & ? & \bullet & ? & ? & ? \\ 0 & 0 & ? & ? & ? & ? \end{pmatrix}$$

(a $\bullet$ indicates a cell we know is in the support of C, a 0 indicates a cell we know is not in the support of C, and ? indicates a cell we are not sure about).

Now suppose that

$$C = \begin{pmatrix} \bullet & \bullet & ? & ? & ? & ? \\ ? & ? & \bullet & ? & ? & ? \\ 0 & 0 & \bullet & ? & ? & ? \end{pmatrix}.$$

Then, again because C_1 cannot contain a permutation matrix, we would have

$$C = \begin{pmatrix} \bullet & \bullet & ? & ? & ? & ? \\ 0 & 0 & \bullet & ? & ? & ? \\ 0 & 0 & \bullet & ? & ? & ? \end{pmatrix}.$$

But then, since the sum of each row of C is 1 and the entries in rows 2 and 3 of C_1 are less than α, rows 2 and 3 in C_2 would have to have at least two nonzero entries each, which would violate extremality of C (transfer vectors possible).

Therefore we can assume that

$$C = \begin{pmatrix} \bullet & \bullet & ? & ? & ? & ? \\ ? & ? & \bullet & ? & ? & ? \\ 0 & 0 & 0 & ? & \bullet & \bullet \end{pmatrix}$$

and also

$$C = \begin{pmatrix} \bullet & \bullet & ? & ? & ? & ? \\ ? & ? & \bullet & ? & ? & ? \\ 0 & 0 & 0 & 0 & \bullet & \bullet \end{pmatrix}$$

because if

$$C = \begin{pmatrix} \bullet & \bullet & ? & ? & ? & ? \\ ? & ? & \bullet & ? & ? & ? \\ 0 & 0 & 0 & \bullet & \bullet & \bullet \end{pmatrix}$$

then M could be also generated by a distribution matrix of the form

$$\begin{pmatrix} \bullet & \bullet & ? & ? & ? & ? \\ ? & ? & \bullet & ? & ? & ? \\ \bullet & \bullet & \bullet & ? & ? & ? \end{pmatrix}$$

which violates extremality of M (the support of M would contain a multiply stochastic matrix generated by a permutation of A).

Since M is not generated by B we must have

$$C = \begin{pmatrix} \bullet & \bullet & ? & ? & ? & ? \\ ? & ? & \bullet & 0 & ? & ? \\ 0 & 0 & 0 & 0 & \bullet & \bullet \end{pmatrix}$$

and then

$$C = \begin{pmatrix} \bullet & \bullet & ? & \bullet & ? & ? \\ ? & ? & \bullet & 0 & ? & ? \\ 0 & 0 & 0 & 0 & \bullet & \bullet \end{pmatrix}$$

because C must have a nonzero element in the fourth column. But this implies that

$$C = \begin{pmatrix} \bullet & \bullet & ? & \bullet & ? & ? \\ ? & ? & \bullet & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bullet & \bullet \end{pmatrix},$$

since C_2 cannot contain a permutation matrix. Now note that if

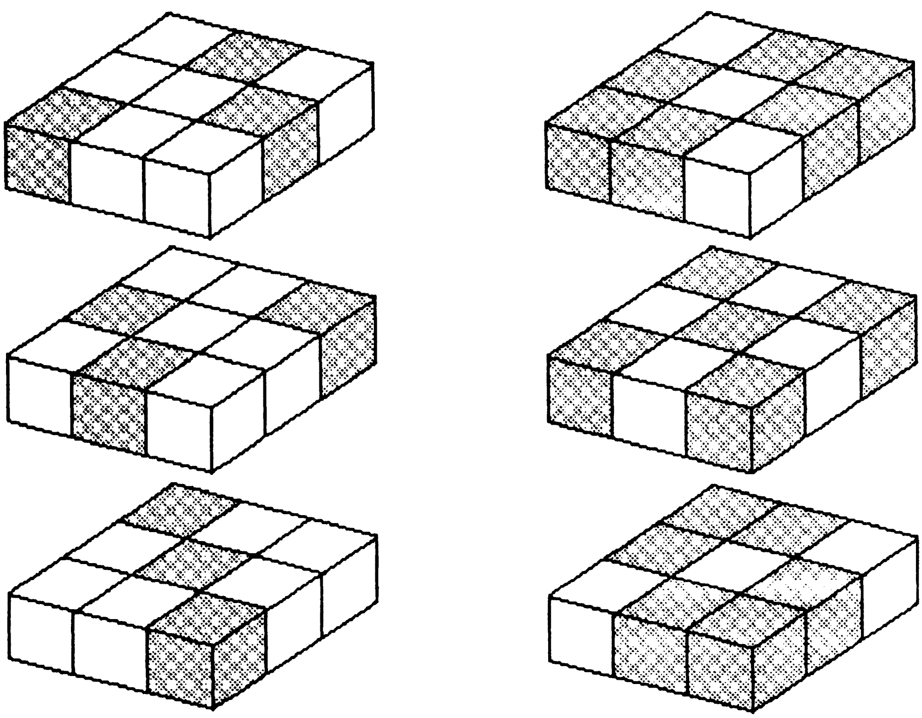
$$C = \begin{pmatrix} \bullet & \bullet & ? & \bullet & ? & ? \\ \bullet & ? & \bullet & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bullet & \bullet \end{pmatrix} \quad \text{or} \quad C = \begin{pmatrix} \bullet & \bullet & ? & \bullet & ? & ? \\ ? & \bullet & \bullet & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bullet & \bullet \end{pmatrix}$$

then C would contain a permutation of B. Thus

$$C = \begin{pmatrix} \bullet & \bullet & ? & \bullet & ? & ? \\ 0 & 0 & \bullet & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bullet & \bullet \end{pmatrix}.$$

But this is impossible, because the mass in the middle row has to be 1. Consequently there does not exist an extremal multiply stochastic matrix which cannot be generated by a permutation of either A or B.

THEOREM 10. *Every extremal element of* $\mathcal{H}(\langle 3\rangle^3, \mathcal{S}_2)$ *is a permutation of one of the following two hypermatrices:*



Note: the 3×3 *matrices represent the "floors" of the cube.*

References

BIRKHOFF, G. (1946). Tres observaciones sobre el algebra lineal, *Univ. Nac. Tucuman, Rev. Ser. A*, **5**, 147–151.

LI, XIN, MIKUSIŃSKI, P., SHERWOOD, H. and TAYLOR, M. D. (1993). Characterizations of extremality of multiply stochastic hypermatrices, (unpublished manuscript).

ROTA, GIAN-CARLO and HARPER, L. (1971). *Matching Theory, an Introduction, Advances in Probability and Related Topics*, Vol. 1, Marcel Dekker, Inc., New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CENTRAL FLORIDA
ORLANDO, FL 32816-1364
xli@pegasus.cc.ucf.edu
piotrm@pegasus.cc.ucf.edu
sherwood@pegasus.cc.ucf.edu
mtaylor@pegasus.cc.ucf.edu