Matrices of Zeros and Ones with Fixed Row and Column Sum Vectors*

Richard A. Brualdi
Department of Mathematics
University of Wisconsin
Madison, Wisconsin 53706

ABSTRACT

Let m and n be positive integers, and let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ be nonnegative integral vectors. We survey the combinatorial properties of the set of all $m\times n$ matrices of 0's and 1's having r_i 1's in row i and s_j 1's in column j. A number of new results are proved. The results can also be formulated in terms of the set of bipartite graphs with a bipartition into m and n vertices having degree sequence R and S, respectively. They can also be formulated in terms of the set of hypergraphs with m vertices having degree sequence R and n edges whose cardinalities are given by S.

1. INTRODUCTION

Let m and n be positive integers, and let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ be nonnegative integral vectors. Denote by $\mathfrak{A}(R,S)$ the set of all $m\times n$ matrices $A=[a_{ij}]$ satisfying

$$a_{ij}=0 \text{ or } 1 \qquad \text{for} \quad i=1,\dots,m \text{ and } j=1,\dots,n; \tag{1.1}$$

$$\sum_{j=1}^{n} a_{ij}=r_i \qquad \text{for} \quad i=1,\dots,m; \tag{1.2}$$

$$\sum_{i=1}^{m} a_{ij}=s_j \qquad \text{for} \quad j=1,\dots,n. \tag{1.3}$$

Thus a matrix of 0's and 1's belongs to $\mathfrak{A}(R,S)$ provided its row sum vector is R and its column sum vector is S. This set $\mathfrak{A}(R,S)$ was the subject of intensive study during the late 1950s and early 1960s by H. J. Ryser, D. R. Fulkerson, R. M. Haber, and D. Gale, and many remarkable theorems were proved. The purpose of this article is to survey the combinatorial properties

*Research partially supported by National Science Foundation Grant No. MCS 76-06374 A01.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/80/050159+73$01.75

of $\mathfrak{A}(R,S)$ including some recent results. In doing so, we present in some cases new proofs of theorems which may be more transparent than those in the literature. Also there appear here for the first time a number of new results, notably the solution (Theorem 6.8) of a problem posed by Ryser [56, p. 76] in 1963. Other new results include Theorems 3.10, 4.2, 4.4, 5.8, 5.9, 6.8, 6.10, 7.3, 8.3, and 8.13, and Corollaries 5.6 and 8.6. Over twenty problems are proposed.

Before proceeding we give two alternative interpretations of $\mathfrak{A}(R,S)$. Let $X=\{x_1,\ldots,x_m\}$ and $Y=\{y_1,\ldots,y_n\}$ be disjoint sets of m and n elements, respectively. Let $\mathrm{BG}(R,S)$ denote the collection of all bipartite graphs G with the following properties:

(BG1) The vertices of G are $x_1,\ldots,x_m,y_1,\ldots,y_n$.
(BG2) Each edge of G joins a vertex in X to a vertex in Y.
(BG3) The degree (or valency) of x_i is r_i for $i=1,\ldots,m$, and the degree of y_j is s_j for $j=1,\ldots,n$.

Then there is a one-to-one correspondence between the matrices in $\mathfrak{A}(R,S)$ and the bipartite graphs in $\mathrm{BG}(R,S)$, determined as follows. If $A=[a_{ij}]\in\mathfrak{A}(R,S)$, there corresponds the bipartite graph $G\in\mathrm{BG}(R,S)$ wherein there is an edge joining x_i to y_j if and only if $a_{ij}=1$ $(1\leqslant i\leqslant m,\ 1\leqslant j\leqslant n)$. To $G\in\mathrm{BG}(R,S)$ there corresponds the matrix $A=[a_{ij}]\in\mathfrak{A}(R,S)$ wherein $a_{ij}=1$ if and only if there is an edge in G joining x_i and y_j $(1\leqslant i\leqslant m,\ 1\leqslant j\leqslant n)$. Thus A is the *incidence matrix* of G. It follows that each of our theorems concerning $\mathfrak{A}(R,S)$ can be formulated as a theorem about the set $\mathrm{BG}(R,S)$ of bipartite graphs.

Now let $E=\{e_1,\ldots,e_m\}$ be a set of m elements, and let $\mathfrak{F}(R,S)$ denote the collection of families $H=(F_1,\ldots,F_n)$ of subsets of E satisfying

$$|F_j|=s_j \qquad \text{for} \quad j=1,\ldots,n, \tag{1.4}$$

$$|\{\,j: e_i\in F_j,\ 1\leqslant j\leqslant n\}|=r_i \qquad \text{for} \quad i=1,\ldots,m. \tag{1.5}$$

In words, the jth set contains s_j elements and the ith element is contained in r_i sets. There is a one-to-one correspondence between the matrices in $\mathfrak{A}(R,S)$ and the families in $\mathfrak{F}(R,S)$, determined in the following way. To $A=[a_{ij}]\in\mathfrak{A}(R,S)$ there corresponds the family $H=(F_1,\ldots,F_n)$ in $\mathfrak{F}(R,S)$ wherein $e_i\in F_j$ if and only if $a_{ij}=1$ $(i=1,\ldots,m,\ j=1,\ldots,n)$. To the family $H=(F_1,\ldots,F_n)$ in $\mathfrak{F}(R,S)$ there corresponds the matrix $A=[a_{ij}]\in\mathfrak{A}(R,S)$ where $a_{ij}=1$ if and only if $e_i\in F_j$ $(i=1,\ldots,m,\ j=1,\ldots,n)$. Thus the matrix A is the *incidence matrix* of the family H of subsets of E. A family in $\mathfrak{F}(R,S)$ can be regarded as a hypergraph [4] which has m vertices and n edges, where the degrees of vertices are $r_1,\dots,r_m$ and the cardinalities of the edges are $s_1,\dots,s_n$. Thus each of our theorems about $\mathfrak{A}(R,S)$ can be regarded as a theorem about a collection of hypergraphs. When S is the n-tuple $(2,\dots,2)$, then a hypergraph in $\mathcal{F}(R,S)$ is just an ordinary graph. Hence each of our theorems about $\mathfrak{A}(R,S)$ can be interpreted as a theorem about the collection of graphs with n edges and with the degree sequence of its vertices equal to R.

2. EXISTENCE

In 1957 Gale [26] and Ryser [52] independently obtained necessary and sufficient conditions for the set $\mathfrak{A}(R,S)$ to be nonempty. Gale's derivation was based on his supply-demand theorem for network flows, while Ryser used an inductive construction to produce a matrix in $\mathfrak{A}(R,S)$. Fulkerson and Ryser [24] gave a construction for a matrix in a nonempty $\mathfrak{A}(R,S)$ with some special properties, and this matrix was later used by Ryser [56, pp. 63–64] to streamline his proof of the criterion for $\mathfrak{A}(R,S)$ to be nonempty. Using an alternative form of Gale's network flow theorem, Ford and Fulkerson [19, p. 82] obtained a criterion for the nonemptiness of $\mathfrak{A}(R,S)$ in terms of the so-called "structure matrix" introduced by Ryser [54] and shown to be so important in the evaluation of certain parameters associated with $\mathfrak{A}(R,S)$. In order to proceed we need to introduce some terminology.

Let $U=(u_1,\dots,u_n)$ be a vector with real coordinates. Then U is called *monotone* provided $u_1\geqslant\cdots\geqslant u_n$. In general, there exists a vector $U'=(u_1',\dots,u_n')$ obtained from U by a rearrangement of coordinates such that U' is monotone. We call U' the *monotone rearrangement* of U. Now let $V=(v_1,\dots,v_n)$ be another vector with n coordinates, and let $V'=(v_1',\dots,v_n')$ be the monotone rearrangement of V. Then U is said to be *majorized* by V (or V majorizes U), written $U\prec V$, provided

$$u_1'+\cdots+u_i'\leqslant v_1'+\cdots+v_i' \qquad \text{for} \quad 1,\dots,n, \tag{2.1}$$

with equality holding when $i=n$. This definition is due to Muirhead and is known to be important in the study of symmetric means [33, pp. 45–64]. The relation $\prec$ is a reflexive, transitive relation on the set of vectors with n real coordinates, but it is not a partial order unless one identifies all rearrangements of a vector (or restricts oneself to monotone vectors).

Now let $R=(r_1,\dots,r_m)$ be a vector whose components are integers satisfying $0\leqslant r_i\leqslant n$ for $i=1,\dots,m$. Let $\tau=r_1+\cdots+r_m$, so that R is an ordered partition of τ. Define $A(R;n)$ to be the $m\times n$ matrix of 0's and 1's where for $i=1,\dots,m$ row i consists of r_i 1's followed by $n-r_i$ 0's. Let the column sum vector of $A(R;n)$ be the monotone vector $R^*=(r_1^*,\dots,r_n^*)$. It follows that

$$r_j^*=|\{i:r_i\geqslant j,\ i=1,\dots,m\}| \qquad \text{for} \quad j=1,\dots,n, \tag{2.2}$$

$\tau=r_1^*+\cdots+r_n^*$, and R and R^* are conjugate partitions of τ.

Let m and n be positive integers, and let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ be vectors. Let $I\subseteq\{1,\dots,m\}$ and $J\subseteq\{1,\dots,n\}$. We define

$$t_{R,S}(I,J)=|I||J|+\sum_{i\notin I} r_i-\sum_{j\in J} s_j. \tag{2.3}$$

Since R and S are usually fixed, we shall write $t(I,J)$ in place of $t_{R,S}(I,J)$. Let A be an $m\times n$ matrix. Then $A[I,J]$ denotes the submatrix of A whose rows are indexed by I and whose columns are indexed by J. If $\bar{I}=\{1,\dots,m\}-I$ and $\bar{J}=\{1,\dots,n\}-J$, then $A(I,J)=A[\bar{I},\bar{J}]$. For a matrix X of 0's and 1's, we write $\sigma_0(X)$ for the number of 0's of X and $\sigma_1(X)$ for the number of 1's of X.

We now state and prove the main theorem for the existence of a matrix in $\mathfrak{A}(R,S)$. The matrix constructed in the proof will play an important role in subsequent sections.

THEOREM 2.1 (Gale [26], Ryser [52; 56, pp. 63–65], Ford and Fulkerson [19, pp. 79–82]). *Let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ be nonnegative integral vectors. Then the following three statements are equivalent*:

(2.1.1) *There exists a matrix in* $\mathfrak{A}(R,S)$;

(2.1.2) $t(I,J)\geqslant 0$ *for all* $I\subseteq\{1,\dots,m\}$ *and* $J\subseteq\{1,\dots,n\}$, *and* $r_1+\cdots+r_m=s_1+\dots+s_n$;

(2.1.3) $S\prec R^*$, *and* $r_k\leqslant n$ *for* $k=1,\dots,m$.

Proof. First suppose that (2.1.1) is satisfied, and let $A\in\mathfrak{A}(R,S)$. An easy calculation then shows that for $I\subseteq\{1,\dots,m\}$ and $J\subseteq\{1,\dots,n\}$,

$$t(I,J)=\sigma_0(A[I,J])+\sigma_1(A(I,J)).$$

It now follows that (2.1.2) holds.

Now suppose that (2.1.2) holds. Let $1 \leqslant k \leqslant n$, and let $J \subseteq \{1,\dots,n\}$ with $|J| = k$. Define $I = \{i : r_i > k,\ i = 1,\dots,m\}$. Then it follows from (2.1.2) that

$$\begin{aligned}\sum_{j\in J} s_j &\leqslant |I|\,|J| + \sum_{i\notin I} r_i \\ &= |I|k + \sum_{i\notin I} r_i \\ &= \sum_{i=1}^{m} \min\{r_i, k\} \\ &= \sum_{j=1}^{k} r_j^*.\end{aligned}$$

It now follows that $S \prec R^*$.

Let $k \in \{1,\dots,m\}$, and let $I = \{k\}$ and $J = \{1,\dots,n\}$. Then using (2.1.2) we obtain

$$0 \leqslant n + \sum_{i\notin I} r_i - \sum_{j=1}^{n} s_j = n - r_k.$$

Thus (2.1.3) holds.

Finally, we suppose that (2.1.3) is satisfied and prove by induction on n that $\mathfrak{A}(R,S)$ is nonempty. There is no loss of generality in assuming that R and S are monotone, and we make this assumption now. For $n=1$, $A(R;1) \in \mathfrak{A}(R,S)$, and so we suppose that $n > 1$. Consider the matrix $A(R;n) \in \mathfrak{A}(R,R^*)$, where $R^* = (r_1^*,\dots,r_n^*)$. From (2.1.3) it follows that

$$s_1 + \cdots + s_{n-1} \leqslant r_1^* + \cdots + r_{n-1}^*,$$

$$s_1 + \cdots + s_n = r_1^* + \cdots + r_n^*,$$

and hence that

$$r_n^* \leqslant s_n. \tag{2.4}$$

The equation above and the monotonicity of S and R^* now imply

$$r_1^* \geqslant s_n. \tag{2.5}$$

It now follows from (2.4) and (2.5) that we may shift the final 1 in some of the rows of $A(R;n)$ so that the last column of the resulting matrix B contains exactly s_n 1's. We may choose the 1's to shift in such a way that they occur in those rows corresponding to the largest s_n coordinates of R, giving preference to the lowest rows in case of ties. Thus the matrix obtained from B by deleting column n has a monotone row sum vector $C=(c_1,\dots,c_m)$ and is the matrix $A(C,n-1)$. Let the column sum vector of $A(C,n-1)$ be $C^*=(c_1^*,\dots,c_{n-1}^*)$ so that C and C^* are conjugate partitions. We now show that

$$(s_1,\dots,s_{n-1})\prec C^*. \tag{2.6}$$

Let $1\leqslant k\leqslant n-1$. Then

$$\begin{aligned}\sum_{j=1}^{k} c_j^* &= \sum_{i=1}^{m} \min\{k,c_i\}\\ &= \sum_{i=1}^{s_n} \min\{k,r_i-1\}+\sum_{i=s_n+1}^{m} \min\{k,r_i\}.\end{aligned} \tag{2.7}$$

Suppose first that $r_{k+1}^*\geqslant s_n$. Since R is monotone, we conclude that $r_i\geqslant k+1$ for $i=1,\dots,s_n$. It then follows that for $i=1,\dots,s_n$, $\min\{k,r_i-1\}=k=\min\{k,r_i\}$ and hence

$$\sum_{j=1}^{k} c_j^* = \sum_{j=1}^{k} r_j^* \geqslant \sum_{j=1}^{k} s_j.$$

Now suppose that $r_{k+1}^*<s_n$. Then

$$r_i\geqslant k+1,\quad \min\{k,r_i-1\}=k=\min\{k,r_i\}\qquad \text{for}\quad i=1,\dots,r_{k+1}^*;$$

$$r_i\leqslant k,\quad \min\{k,r_i-1\}=\min\{k,r_i\}-1\qquad \text{for}\quad i=r_{k+1}^*+1,\dots,s_n.$$

It now follows from (2.7) and (2.1.3) that

$$\begin{aligned}\sum_{j=1}^{k} c_j^* &= \sum_{i=1}^{m} \min\{k, r_i\} + r_{k+1}^* - s_n \\ &= \sum_{j=1}^{k+1} r_j^* - s_n \\ &\geqslant \sum_{j=1}^{k+1} s_j - s_n \\ &\geqslant \sum_{j=1}^{k} s_j.\end{aligned}$$

Since $s_1 + \cdots + s_{n-1} = c_1^* + \cdots + c_{n-1}^*$, we conclude that (2.6) holds. It now follows by induction on n that (2.1.1) holds, and the proof of the theorem is complete. ■

The inductive proof given above furnishes a simple column-by-column construction for a matrix in $\mathfrak{A}(R,S)$ when $S \prec R^*$ (when S is not majorized by R^*, the construction must fail at some stage). This same construction was used by both Ryser [52] and Gale [26] in their original proofs, except that both proceeded by constructing the columns in order of decreasing column sums. While Gale placed the 1's in each column in those positions corresponding to the largest row sums at each stage (giving preference to the topmost positions in case of ties), it was Fulkerson and Ryser [24] who recognized that giving preference to the lower positions in case of ties leads to the construction of a matrix in $\mathfrak{A}(R,S)$ with some special properties. As shown in [24], the columns can be constructed in any order. The matrix constructed by carrying out the steps in the inductive proof above is denoted by $\tilde{A}$ and will be further investigated in Sec. 7. We sometimes refer to it as the *canonical matrix* of $\mathfrak{A}(R,S)$. It follows from the construction that for each $j = 1, \ldots, n$,

$$\tilde{A}[\{1,\ldots,m\},\{1,\ldots,j\}]$$

has a monotone row sum vector (and, of course, a monotone column sum vector). An example of such a matrix $\tilde{A}$ is the following. Let $m=5$, $n=6$,

$R=(5, 4, 3, 3, 2)$, and $S=(4, 4, 3, 2, 2, 2)$. Then

$$\tilde{A}=\begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Suppose $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ are both monotone, nonnegative integral vectors with $r_1+\cdots+r_m=s_1+\cdots+s_n$. It then follows that for $k=0,1,\dots,m$ and $l=0,1,\dots,n$,

$$t_{kl}=\min\{t(I,J): I\subseteq\{1,\dots,m\}, |I|=k,$$
$$J\subseteq\{1,\dots,n\}, |J|=l\},$$

where

$$t_{kl}=kl+\sum_{i>k} r_i-\sum_{j<l} s_j.$$

Using Theorem 2.1, we conclude that $\mathfrak{A}(R,S)$ is nonempty if and only if the $(m+1)\times(n+1)$ matrix $T=[t_{ij}]$ is a nonnegative matrix (all entries are nonnegative). This matrix was introduced by Ryser [54] in his study of $\mathfrak{A}(R,S)$ and is called the *structure matrix* for $\mathfrak{A}(R,S)$. The structure matrix will be examined further in Sec. 4, and is important in the determination of several parameters associated with $\mathfrak{A}(R,S)$.

We conclude this section with some additional comments on Theorem 2.1 and some generalizations. A proof of the equivalence of (2.1.1) and (2.1.3) using results of transversal theory can be found in Mirsky [48; 49, pp. 76–78, 206–209] along with a direct proof of the equivalence of (2.1.2) and (2.1.3) [49, p. 207]. As noted by Anstee [1], the equivalence of (2.1.2) and (2.1.3) follows from the relation

$$t_{r_f^* f}=\sum_{i=1}^{f} r_i^*-\sum_{i=1}^{f} r_i.$$

Mirsky [48; 49, pp. 205–206] obtained the more general result concerning the existence of an integral matrix whose entries, row sums, and column sums all lie within prescribed bounds:

THEOREM 2.2. *Let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ be nonnegative integral vectors. Let $C=[c_{ij}]$ be an $m\times n$ matrix of 0's and 1's. Then there* *exists an integral matrix* $A=[a_{ij}]$ *with row sum vector* R *and column sum vector* S *such that* $0 \leqslant a_{ij} \leqslant c_{ij}$ $(i=1,\dots,m,\ j=1,\dots,n)$ *if and only if* $r_1 + \cdots + r_m = s_1 + \cdots + s_n$ *and*

$$\sum_{i \in I, j \in J} c_{ij} + \sum_{i \notin I} r_i - \sum_{j \in J} s_j \geqslant 0$$

for all $I \subseteq \{1,\dots,m\}$ *and* $J \subseteq \{1,\dots,n\}$.

Let R and S be monotone integral vectors. As noted by Mirsky [49, p. 213], Fulkerson's theorem [21; 19, pp. 85–87], which gives necessary and sufficient conditions for the existence of a matrix in $\mathfrak{A}(R,S)$ with zero trace, can be derived from Theorem 2.2. Recently, Anstee [1] has generalized Fulkerson's theorem (and the Gale-Ryser theorem) by obtaining necessary and sufficient conditions for the existence of a matrix in $\mathfrak{A}(R,S)$ which has zeros in a prescribed set of positions consisting of at most one position per column. Let P be a set of positions for an $m \times n$ matrix such that the number of positions of P in row i is c_i for $i=1,\dots,m$ and the number of positions of P in column j is d_j for $j=1,\dots,n$. Suppose that $d_j \leqslant 1$ and $s_j \leqslant m-d_j$ for $j=1,\dots,n$ and that $r_i \leqslant n-c_i$ for $i=1,\dots,m$. Let $B(R;n)$ be the $m \times n$ matrix of 0's and 1's with row sum vector R such that the 1's of $B(R;n)$ are as far to the left as possible subject to the restriction that they do not occupy any of the positions of P. Let R^{**} be the column sum vector of $B(R;n)$. Then assuming R and S are monotone, Anstee [1] proved that there is a matrix in $\mathfrak{A}(R,S)$ which has 0's in all positions in P if and only if $S \prec R^{**}$. Fulkerson's theorem is the special case which results when $m=n$ and $P=\{(1,1),\dots,(m,m)\}$. All of the theorems cited can be regarded as special cases of the integral version of the supply-demand theorem of Gale [26] (see also [19, pp. 38–39]) and consequently can be derived from it. A transfinite extension of the matrix results above is derived in [8].

3. INTERCHANGES

We first introduce some notions of graph theory. We shall not be complete; the uninitiated reader can find more details in [4] or [32]. Let D be a directed graph. An *elementary circuit* γ (of *length* p) of D is a sequence $(z_1,z_2,\dots,z_p,z_1)$ of vertices of D such that $z_1,z_2,\dots,z_p$ are distinct and $\alpha(\gamma)=\{(z_1,z_2),\dots,(z_{p-1},z_p),\ (z_p,z_1)\}$ is a set of arcs of D. We call $\alpha(\gamma)$ the (set of) *arcs* of γ. If the directed graph D has the property that for each vertex z the number of arcs entering z equals the number of arcs exiting z, then it follows by an easy induction that there exist elementary circuits $\gamma_1,\dots,\gamma_q$ of D such that $\alpha(\gamma_1),\alpha(\gamma_2),\dots,\alpha(\gamma_q)$ is a partition of the set of arcs of D. Neither the elementary circuits $\gamma_1,\dots,\gamma_q$ nor their number q is uniquely determined by D. An example will be given below. The largest such q will be called the *circuit packing number* of D and will be denoted by $\mathrm{cp}(D)$.

Now let $X=\{x_1,\dots,x_m\}$ and $Y=\{y_1,\dots,y_n\}$ be disjoint sets of m and n elements, respectively. To each $m\times n$ matrix $C=[c_{ij}]$ with entries from the set $\{-1,0,1\}$ we associate a *directed bipartite graph* $\Gamma(C)$ with vertices $x_1,\dots,x_m, y_1,\dots,y_n$ as follows. In $\Gamma(C)$ there is an arc (x_i,y_j) from x_i to y_j if $c_{ij}=1$, and an arc (y_j,x_i) from y_j to x_i if $c_{ij}=-1$. There are no other arcs in $\Gamma(C)$. Let $\mathcal{C}_{m,n}$ denote the set of all $m\times n$ matrices with entries from $\{-1,0,1\}$ with each row and column sum equal to 0. Let $C\in\mathcal{C}_{m,n}$. Then for each vertex z of $\Gamma(C)$ the number of arcs entering z equals the number of arcs exiting z. Thus there exist elementary circuits $\gamma_1,\dots,\gamma_q$ (necessarily of even length) such that $\alpha(\gamma_1),\dots,\alpha(\gamma_q)$ is a partition of the set of arcs of $\Gamma(C)$. If we regard γ_i as a directed graph with vertices $x_1,\dots,x_m,y_1,\dots,y_n$ and arcs $\alpha(\gamma_i)$, then there exists a matrix $C_i\in\mathcal{C}_{m,n}$ such that $\Gamma(C_i)=\gamma_i$ $(i=1,\dots,q)$. Since γ_i is an elementary circuit, each C_i has either no or two nonzero entries in each row and column; in particular C has an even number of nonzero entries. Since $\alpha(\gamma_1),\dots,\alpha(\gamma_q)$ is a partition of the arcs of $\Gamma(C)$, it follows that $C=C_1+\cdots+C_q$. Moreover, if $i\neq j$, then C_i and C_j do not have nonzero entries in the same position. We call the matrices $C_1,\dots,C_q$ *elementary circuit matrices* and refer to them as *pairwise position disjoint*. For an example, let

$$C=\begin{bmatrix} 1 & -1 & -1 & 1\\ 1 & 1 & -1 & -1\\ -1 & 1 & 1 & -1\\ -1 & -1 & 1 & 1\end{bmatrix}. \tag{3.1}$$

We can find 4 elementary circuits of length 4 in $\Gamma(C)$ whose arcs partition the set of arcs of $\Gamma(C)$, and 2 elementary circuits of length 8 with the same property. These lead to the following representations of C as sums of elementary matrices:

$$C=E_1+E_2+E_3+E_4,$$

where

$$E_1=\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad E_2=\begin{bmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$E_3=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \quad E_4=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}; \tag{3.2}$$

and

$$C=C_1+C_2,$$

where

$$C_1=\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \quad C_2=\begin{bmatrix} 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}. \tag{3.3}$$

Now let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ be nonnegative integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$. Let $B, A\in\mathfrak{A}(R,S)$, and define C to be the matrix $B-A$. Then $C\in\mathcal{C}_{m,n}$. Hence to the ordered pair of matrices $B, A\in \mathfrak{A}(R,S)$ we can associate a directed bipartite graph $\Gamma(B,A)=\Gamma(C)$. From our previous discussion it follows that there exist pairwise position disjoint, elementary circuit matrices $C_1,\dots,C_q$ such that

$$B=A+C_1+\cdots+C_q.$$

For example, let $R=S=(2,2,2,2)$, and let

$$B=\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad A=\begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}. \tag{3.4}$$

Then $C=B-A$ is the matrix defined by (3.1), and $B=A+E_1+E_2+E_3+E_4$, where E_1, E_2, E_3, E_4 are defined by (3.2). Also $B=A+C_1+C_2$, where C_1 and C_2 are defined by (3.3).

Ryser [52] has defined an *interchange* to be a transformation which replaces the 2×2 submatrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.5}$$

of a matrix A of 0's and 1's with the 2×2 submatrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{3.6}$$

or vice versa. If the submatrix (3.5) [or (3.6)] lies in rows k,l and columns u,v, then we call the interchange a *$(k,l;u,v)$-interchange*. Clearly an interchange (and hence any finite sequences of interchanges) does not alter the row and column sum vectors of a matrix, and therefore transforms a matrix in $\mathfrak{A}(R,S)$ into another matrix in $\mathfrak{A}(R,S)$. Ryser [52] has proved the converse of this result by inductively showing that given $A,B\in\mathfrak{A}(R,S)$ there is a sequence of interchanges which transforms A into B. In [56, p. 68] this same result is obtained by showing that given a matrix $A\in\mathfrak{A}(R,S)$, there exists a sequence of interchanges which transforms A into the canonical matrix $\tilde{A}$ of $\mathfrak{A}(R,S)$. We give a direct constructive proof of Ryser's result below.

Let A be an $m\times n$ matrix of 0's and 1's, and let B be obtained from A by an interchange. Let this interchange be a $(k,l;u,v)$-interchange, and let $E(k,l;u,v)=[e_{ij}]$ be the matrix in $\mathcal{C}_{m,n}$ defined by

$$\begin{aligned} &e_{ku}=e_{lv}=1,\\ &e_{kv}=e_{lu}=-1,\\ &e_{ij}=0 \qquad \text{otherwise.} \end{aligned}$$

Then it follows that $B=A+E(k,l;u,v)$. The matrix $E(k,l;u,v)$ is an elementary circuit matrix corresponding to an elementary circuit of length 4. We refer to such matrices as *interchange matrices*. It follows that given a sequence of t interchanges which transforms a matrix A into a matrix B, there exists a sequence of t interchange matrices (not necessarily pairwise disjoint) $E_1,\dots,E_t$ such that $B=A+E_1+\cdots+E_t$, where $A+E_1,\dots,A+E_1+\cdots+E_t$ are matrices of 0's and 1's. The converse also holds. For example, if A and B are the matrices of (3.4), then E_1,E_2,E_3,E_4, where these matrices are defined in (3.2), is a sequence of interchange matrices corresponding to a sequence of interchanges transforming A into B.

THEOREM 3.1. *Let $A=[a_{ij}]$ and $B=[b_{ij}]$ be matrices in $\mathfrak{A}(R,S)$, and let $B=A+C_1+\cdots+C_q$, where $C_1,\dots,C_q$ are pairwise disjoint, elementary circuit matrices. Suppose C_1 has $2k$ nonzero entries. Then there exists a sequence of $k-1$ interchanges which transforms A into $A+C_1$.*

Proof. Let $C_1=[c_{ij}]$. There is no loss of generality in assuming that $c_{11}=\cdots=c_{kk}=-1$, $c_{12}=\cdots=c_{k-1,k}=c_{k1}=1$, and $c_{ij}=0$ otherwise. This is because we may apply the same row and column permutations to A and B, and change the order of the coordinates of R and S accordingly. Thus $a_{11}=\cdots=a_{kk}=1=b_{12}=\cdots=b_{k-1,k}=b_{k1}$, while $a_{12}=\cdots=a_{k-1,k}=a_{k1}=0=b_{11}=\cdots=b_{kk}$. Define an integer p by $a_{11}=\cdots=a_{p-1,1}=1$, $a_{p1}=0$. Since $a_{k1}=0$, p exists and satisfies $2\leqslant p\leqslant k$. Then the sequence of $p-1$ interchanges with labels

$$(p-1,p;1,p),\quad (p-2,p-1;1,p-1),\dots,\quad (1,2;1,2)$$

transforms A into a matrix A' such that $B=A'+C_1'+C_2+\cdots+C_q$, where $C_1',C_2,\dots,C_q$ are pairwise disjoint, elementary circuit matrices and C_1' has length $2(k-p+1)$. We may replace C_1 by C_1' and A by A' in the above argument and inductively obtain a sequence of $k-p$ interchanges which transforms A' into $A'+C_1'$. Hence there exists a sequence of $k-1$ interchanges which transforms A into $A'+C_1'=B-(C_2+\dots+C_q)=A+C_1$. This completes the proof of the theorem. ∎

Informally, the theorem says that if A and B differ by an elementary circuit of length $2k$ plus possibly some other positions, then there is a sequence of $k-1$ interchanges which transforms A into a matrix A^* that agrees with B at the positions of the circuit and *with A at all other positions.*

THEOREM 3.2. *Let A and B be matrices in $\mathfrak{A}(R,S)$, and let $B-A=C_1+\cdots+C_q$, where $C_1,\dots,C_q$ are pairwise disjoint, elementary circuit matrices. Let the number of nonzero entries of C_i be $2k_i$ $(1\leqslant i\leqslant q)$. Then there exists a sequence of $k_1+\cdots+k_q-q$ interchanges which transforms A into B.*

Proof. The theorem follows from Theorem 3.1 by induction on q. ∎

Ryser's theorem [52] that a matrix in $\mathfrak{A}(R,S)$ can be transformed into any other by a sequence of interchanges follows from Theorem 3.2. The proof of Theorem 3.1 furnishes an algorithm for transforming $A\in\mathfrak{A}(R,S)$

into $B \in \mathfrak{A}(R,S)$. The elementary circuit matrices $C_1,\ldots,C_q$ are easily found by a row and column scanning procedure applied to $B-A$, and the interchanges were constructed in the proof of Theorem 3.1. As already pointed out, the matrices $C_1,\ldots,C_q$ in Theorem 3.2 are not uniquely determined; more importantly, the lengths of the resulting sequences of interchanges may differ. For example, let A and B be the matrices defined in (3.4). Using $B=A+E_1+E_2+E_3+E_4$, where E_1, E_2, E_3, E_4 are the pairwise disjoint, elementary circuit matrices defined in (3.2), we have a sequence of 4 interchanges which transforms A to B (since E_1, E_2, E_3, E_4 are actually interchange matrices). Using $B=A+C_1+C_2$, where C_1 and C_2 are defined in (3.3), we obtain a sequence of 6 interchanges which transforms A into B. Suppose A and B differ in $2d$ positions. Then if $B-A$ can be written as a sum of q pairwise disjoint, elementary circuit matrices, then by Theorem 3.2 there exists a sequence of $d-q$ interchanges which transforms A into B. It follows that larger values of q give smaller sequences of interchanges transforming A into B. Let $i(A,B)$ equal the minimum length of a sequence of interchanges which transforms A into B, let $d(A,B)$ equal the number of positions in which A and B differ, and let $q(A,B)$ equal the largest number of pairwise disjoint, elementary circuit matrices whose sum equals $B-A$. We then have the following.

COROLLARY 3.3.

$$i(A,B) \leqslant \frac{d(A,B)}{2} - q(A,B).$$

In particular, since $q(A,B) \geqslant 1$, we obtain $i(A,B) \leqslant d(A,B)/2-1$, an inequality also noted by Anstee [1]. We now prove that equality holds in Corollary 3.3. (We are indebted to H. J. Ryser for pointing out that this theorem had previously been obtained by Walkup [65].)

THEOREM 3.4. *Let $A, B \in \mathfrak{A}(R,S)$. Then*

$$i(A,B) = \frac{d(A,B)}{2} - q(A,B). \tag{3.7}$$

Proof. We prove the theorem by using induction on $t=i(A,B)$. If $t=1$, then $d(A,B)=4$, $q(A,B)=1$, and (3.7) holds. Let $t>1$. Then there exist matrices $A_0=A, A_1,\ldots,A_t=B$ in $\mathfrak{A}(R,S)$ such that for $i=1,\ldots,t$, A_i can be obtained from A_{i-1} by an interchange. Since $i(A,B)=t$, it follows that

$i(A_1, B)=t-1$. Thus by the inductive assumption,

$$t-1=i(A_1, B)=\frac{d(A_1, B)}{2}-q(A_1, B).$$

Let $q_1=q(A_1, B)$, and let $C_1,\ldots,C_{q_1}$ be pairwise disjoint, elementary circuit matrices such that $B-A_1=C_1+\ldots+C_{q_1}$. Let $E=A_1-A$, so that E is an interchange matrix. We now distinguish five cases according to the number k of matrices $C_1,\ldots,C_{q_1}$ which have a nonzero entry in the same position as a nonzero entry of E. We note that since $(B-A)=(B-A_1)+E$, where each of the three matrices has entries taken from $\{-1,0,1\}$, nonzero entries of $B-A_1$ and E in the same position have opposite signs.

Case $k=0$. We then have $d(A,B)=4+d(A_1,B)$ and $q(A,B)\geqslant q(A_1,B)+1$. Hence

$$\begin{aligned}\frac{d(A,B)}{2}-q(A,B)&\leqslant\frac{4+d(A_1,B)}{2}-q(A_1,B)-1\\&=2+(t-1)-1=t.\end{aligned}$$

Case $k=1$. It follows that $d(A,B)\leqslant d(A_1,B)+2$ and $q(A,B)\geqslant q(A_1,B)$, or $d(A,B)=d(A_1,B)-4$ and $q(A,B)\geqslant q(A_1,B)-1$. In the former case,

$$\begin{aligned}\frac{d(A,B)}{2}-q(A,B)&\leqslant\frac{d(A_1,B)+2}{2}-q(A_1,B)\\&=(t-1)+1=t;\end{aligned}$$

in the latter case,

$$\begin{aligned}\frac{d(A,B)}{2}-q(A,B)&\leqslant\frac{d(A_1,B)-4}{2}+q(A_1,B)+1\\&=t-1-2+1=t-2\leqslant t.\end{aligned}$$

Case $k=2$. We have $d(A,B)\leqslant d(A_1,B)$ and $q(A,B)\geqslant q(A_1,B)-1$. Thus

$$\begin{aligned}\frac{d(A,B)}{2}-q(A,B)&\leqslant\frac{d(A_1,B)}{2}-q(A_1,B)+1\\&=t-1+1=t.\end{aligned}$$

Case $k=3$. It follows that $d(A,B)\leqslant d(A_1,B)-2$ and $q(A,B)\geqslant q(A_1,B)-2$. Thus

$$\frac{d(A,B)}{2}-q(A,B)\leqslant\frac{d(A_1,B)-2}{2}-q(A_1,B)+2$$

$$=(t-1)-1+2=t.$$

Case $k=4$. We then have $d(A,B)=d(A_1,B)-4$ and $q(A,B)\geqslant q(A_1,B)-3$. Hence

$$\frac{d(A,B)}{2}-q(A,B)\leqslant\frac{d(A_1,B)-4}{2}-q(A_1,B)+3$$

$$=(t-1)-2+3=t.$$

In each case we have found that

$$i(A,B)\geqslant\frac{d(A,B)}{2}-q(A,B).$$

It now follows from Corollary 3.3 that (3.7) holds, and the theorem follows by induction. ∎

Theorem 3.4 provides a formula of sorts for the minimum number $i(A,B)$ of interchanges required to transform a given matrix $A\in\mathfrak{A}(R,S)$ to another matrix $B\in\mathfrak{A}(R,S)$. In [56, p. 68] $i(A,B)$ is described as "apparently a hopelessly complicated function" of A and B. As to whether (3.7) is a *formula* for $i(A,B)$ is a matter of interpretation, since it replaces the determination of a minimum (number of interchanges) by the determination of a maximum (number of pairwise disjoint elementary circuits in the graph $\Gamma(B,A)$).

We note the following corollary to Theorem 3.4.

COROLLARY 3.5. *Let* $A,B\in\mathfrak{A}(R,S)$, *and suppose* $B-A$ *is an elementary circuit matrix with* $2k$ *nonzero entries. Then* $i(A,B)=k-1$.

Let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ be nonnegative integral vectors such that $\mathfrak{A}(R,S)$ is nonempty. We then define a graph $G(R,S)$, called the *interchange graph* of $\mathfrak{A}(R,S)$, as follows. The vertices of $G(R,S)$ are the matrices in $\mathfrak{A}(R,S)$. For $A,B\in\mathfrak{A}(R,S)$, there is an edge joining A and B if and only if B can be obtained from A by a single interchange ($B-A$ is an interchange matrix). The fact that a matrix in $\mathfrak{A}(R,S)$ can be transformed into any other by a sequence of interchanges is equivalent to the statement that $G(R,S)$ *is a connected graph*. In a graph G an *elementary chain* (of length p) joining vertices x and y is a sequence $x=x_1,x_2,\ldots,x_{p+1}=y$ of $p+1$ distinct vertices such that x_i is joined to x_{i+1} by an edge for $i=1,\ldots,p$. The distance between two vertices in a connected graph is the length of the shortest elementary chain joining the two vertices; it follows that for $A,B\in\mathfrak{A}(R,S)$ the distance between A and B in $G(R,S)$ is given by $i(A,B)$. We know very little about this graph apart from its connectedness and its distance function given by Theorem (3.4).

For an example let $R=S=(2,2,1)$. Then $\mathfrak{A}(R,S)$ consists of the following 5 matrices:

$$A_1=\begin{bmatrix}1&1&0\\1&1&0\\0&0&1\end{bmatrix},\quad A_2=\begin{bmatrix}1&1&0\\1&0&1\\0&1&0\end{bmatrix},\quad A_3=\begin{bmatrix}1&1&0\\0&1&1\\1&0&0\end{bmatrix},$$

$$A_4=\begin{bmatrix}0&1&1\\1&1&0\\1&0&0\end{bmatrix},\quad A_5=\begin{bmatrix}1&0&1\\1&1&0\\0&1&0\end{bmatrix}.$$

It is then easy to verify that $G(R,S)$ is the graph drawn in Fig. 1.

We now pose some questions concerning interchange graphs which seem particularly interesting. The *diameter* of a graph is the greatest distance between a pair of vertices of the graph.

PROBLEM 3.6. *Investigate the diameter* $\max\{i(A,B): A, B\in\mathfrak{A}(R,S)\}$ *of* $G(R,S)$. In particular, find a tight upper bound on the diameter depending only on the size m of R and size n of S. It follows easily from (3.7) that the diameter cannot exceed $mn/2-1$. We *conjecture* that it cannot exceed

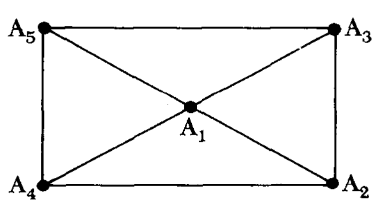

FIG. 1. An interchange graph.

$mn/4$ but are unable to prove this. If true, the diameter when $m=n=2k$ and $R=S=(k,\ldots,k)$ would equal $mn/4$, since there are matrices A and B in $\mathfrak{A}(R,S)$ with $d(A,B)=mn$. The diameter of the interchange graph of Fig. 1 is 2.

The *eccentricity* of a vertex x of a graph is the maximum of the distances from x to each vertex of the graph. The *radius* of a graph is the minimum of the eccentricities of its vertices. A vertex is called *central* if its eccentricity equals the radius of the graph.

PROBLEM 3.6. *Investigate the radius of* $G(R,S)$ *and determine a matrix in* $\mathfrak{A}(R,S)$ *which is central.* For the interchange graph of Fig. 1, the radius is 1 and the only central vertex is A_1. The canonical matrix $\tilde{A}$ for $\mathfrak{A}(R,S)$ when $R=S=(2,2,1)$ is the matrix A_2, and it follows that $\tilde{A}$ need not be central.

PROBLEM 3.7. *Does the interchange graph* $G(R,S)$ *have a hamiltonian cycle?* That is, is it possible to order the matrices in $\mathfrak{A}(R,S)$ as $A_1, A_2, \ldots, A_p$ [p equals the number of matrices in $\mathfrak{A}(R,S)$] so that A_{i+1} is obtainable from A_i by an interchange ($i=1,\ldots,p-1$) and A_1 is obtainable from A_p by an interchange? If there is no hamiltonian cycle, does there exist a hamiltonian chain (in the above we do not insist that A_1 be obtainable from A_p by an interchange)? If there is no hamiltonian chain, does there exist a matrix $A \in \mathfrak{A}(R,S)$ and elementary chains of $G(R,S)$ such that A is the initial vertex of each of these chains and every other matrix in $\mathfrak{A}(R,S)$ is a vertex of exactly one of these chains?

The above questions all relate to generating the matrices in $\mathfrak{A}(R,S)$ with as little redundancy as possible.

The *connectivity* of a graph G is the minimal number of vertices whose removal (with all incident edges) leaves a nonconnected graph or a graph with one vertex. The connectivity of the graph of Fig. 1 is readily seen to be 3. A graph with more than one vertex is connected if and only if its connectivity is at least 1. Thus it follows from Theorem 3.2 that the connectivity of the interchange graph is at least 1 if $\mathfrak{A}(R,S)$ contains at least two matrices. If $\mathfrak{A}(R,S)$ contains only two matrices, the connectivity of $G(R,S)$ is 1.

PROBLEM 3.8. *Determine the connectivity of the interchange graph* $G(R,S)$.

As a partial solution to the above problem we establish the following.

THEOREM 3.9. *Suppose $\mathfrak{A}(R,S)$ contains at least three matrices. Then the connectivity of $G(R,S)$ is at least 2.*

Proof. Let A,B,C be three distinct matrices in $\mathfrak{A}(R,S)$. We need to show that A can be transformed to B by a sequence of interchanges where none of the matrices intermediary to A and B equals C. Consider an elementary chain $A=A_1,A_2,\dots,A_k=B$ joining A and B, where $A_1,\dots,A_k$ are distinct matrices and A_{i+1} is obtainable from A_i by an interchange for $i=1,\dots,k-1$. If none of $A_2,\dots,A_{k-1}$ equals C, there is nothing more to prove. Suppose $A_t=C$ where $1<t<k$. We show that there is an elementary chain joining A_{t-1} and A_{t+1} of length 1 or 2 which avoids C, and hence an elementary chain joining A and B avoiding C.

Let A_t be obtained from A_{t-1} by the interchange $I(p,q;r,s)$ and let A_{t+1} be obtained from A_t by the interchange $I(a,b;u,v)$. Suppose first that $X=\{p,q\}\times\{r,s\}$ and $Y=\{a,b\}\times\{u,v\}$ are disjoint. Let A_t' be the matrix obtained from A_{t-1} by the interchange $I(a,b;u,v)$. Then A_{t+1} is obtained from A_t' by the interchange $I(p,q;r,s)$. Thus A_{t-1},A_t',A_{t+1} is an elementary chain avoiding C. If X and Y are not disjoint, then they have either 1 or 2 elements in common. First consider the case of 2 elements in common. The two interchanges then affect only the entries in a 2×3 or 3×2 matrix. It suffices to consider the first possibility. There is essentially only one possibility for the corresponding 2×3 submatrices of A_{t-1},A_t,A_{t+1} and these are indicated below:

$$A_{t-1}=\begin{bmatrix}1&0&0\\0&1&1\end{bmatrix},\qquad A_t=\begin{bmatrix}0&1&0\\1&0&1\end{bmatrix},\qquad A_{t+1}=\begin{bmatrix}0&0&1\\1&1&0\end{bmatrix}.$$

Since the third of these is obtainable directly from the first by an interchange, the conclusion follows. We now consider the case where X and Y have only one element in common. The two interchanges then effect only the entries in a 3×3 submatrix. There is essentially only one possibility for the corresponding 3×3 submatrices of A_{t-1}, A_t, A_{t+1} as indicated below:

$$A_{t-1}=\begin{bmatrix}1&0&f\\0&1&1\\e&1&0\end{bmatrix},\qquad A_t=\begin{bmatrix}0&1&f\\1&0&1\\e&1&0\end{bmatrix},\qquad A_{t+1}=\begin{bmatrix}0&1&f\\1&1&0\\e&0&1\end{bmatrix}.$$

Since e and f may be 0 or 1, there are four possible cases. Each leads to an elementary chain A_{t-1}, A_t', A_{t+1} where $A_t' \neq C$ as shown below:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

The theorem now follows. ■

While there are a number of other questions that one might ask about the interchange graph $G(R,S)$ (the maximum degree of a vertex, chromatic index, characterization of graphs that are isomorphic to interchange graphs), the above seem most intriguing, and while probably hard, some progress can perhaps be made. In addition, their resolution should shed some light on $\mathfrak{A}(R,S)$.

4. THE STRUCTURE MATRIX

Throughout this section we assume that $R=(r_1,\ldots,r_m)$ *and* $S=(s_1,\ldots,s_n)$ *are monotone, nonnegative integral vectors with* $r_1+\cdots+r_m=s_1+\cdots+s_n$, this common value being designated by τ. The *structure matrix* $T=[t_{kl}]$ for $\mathfrak{A}(R,S)$ was defined in Sec. 2 by

$$t_{kl}=kl+\sum_{i>k} r_i-\sum_{j<l} s_j \qquad (k=0,\ldots,m, \quad l=0,\ldots,n). \tag{4.1}$$

In particular, its entries are determined by R and S. It follows from Theorem 2.1 that $\mathfrak{A}(R,S)$ is nonempty if and only if T is a nonnegative matrix. We note that the entries of row 0 and column 0 of T are given by

$$\begin{aligned} t_{0l} &= \tau - \sum_{j<l} s_j \qquad (l=0,\ldots,n),\\ t_{k0} &= \sum_{i>k} r_i = \tau - \sum_{i\le k} r_i \qquad (k=0,\ldots,m). \end{aligned} \tag{4.2}$$

Thus $t_{00}=\tau$, $t_{0n}=t_{m0}=0$, $t_{0l}-t_{0,l+1}=s_{l+1}$ $(l=0,\ldots,n-1)$, and $t_{k0}-t_{k+1,0}=r_{k+1}(k=0,\ldots,m-1)$. So to prescribe R and S is tantamount to prescribing the $m+n+1$ entries of row 0 and column 0 of T from which the other entries of T are uniquely determined. The following recursive formulas of Ryser [54] follow easily from (4.1):

$$t_{k,l+1} = t_{kl} + k - s_{l+1} \qquad (k=0,\ldots,m;\, l=0,\ldots,n-1), \tag{4.3}$$

$$t_{k+1,l} = t_{kl} + l - r_{k+1} \qquad (k=0,\ldots,m-1;\, l=0,\ldots,n). \tag{4.4}$$

From (4.3) or (4.4) we obtain for $k=0,\ldots,m-1$ and $l=0,\ldots,n-1$ that

$$(t_{kl}-t_{k,l+1})-(t_{k+1,l}-t_{k+1,l+1})=1,$$

which can also be written as

$$(t_{kl}-t_{k+1,l})-(t_{k,l+1}-t_{k+1,l+1})=1. \tag{4.5}$$

Thus consecutive differences of corresponding entries in a pair of adjacent columns of the structure matrix T themselves differ by 1. A similar statement applies to rows. Using either of these two relations, we obtain a quick way to determine the entries of T when the entries of row 0 and column 0 are given.

A sequence $c_0, c_1, \ldots, c_p$ of real numbers is called *convex* provided

$$c_i + c_{i+2} \geqslant 2c_{i+1} \qquad (i=0,\ldots,p-2).$$

Since this inequality can be rewritten as

$$c_i - c_{i+1} \geqslant c_{i+1} - c_{i+2} \qquad (i=0,\ldots,p-2),$$

a sequence is convex if and only if its difference sequence is nonincreasing. It follows that a convex sequence $c_0, c_1, \ldots, c_p$ is *unimodal*, that is, there exists q with $0 \leqslant q \leqslant p$ such that $c_0 \geqslant \cdots \geqslant c_q$, $c_q \leqslant \cdots \leqslant c_p$.

LEMMA 4.1. *The entries of each row and of each column of the structure matrix* $T=[t_{kl}]$ *for* $\mathfrak{A}(R,S)$ *form a convex sequence.*

Proof. Let $0\leqslant e\leqslant m$ and $0\leqslant f\leqslant n-2$. It follows from (4.3) that

$$\begin{aligned} t_{ef}+t_{e,f+2} &= (t_{e,f+1}-e+s_{f+1})+(t_{e,f+1}+e-s_{f+2}) \\ &= 2t_{e,f+1}+(s_{f+1}-s_{f+2}) \\ &\geqslant 2t_{e,f+1}. \end{aligned}$$

Thus the entries in row e of T form a convex sequence. Similarly, by using (4.4) we find that $t_{ef}+t_{e+2,f}\geqslant 2t_{e+1,f}$ for $0\leqslant e\leqslant m-2$ and $0\leqslant f\leqslant n$, and the entries in column f also form a convex sequence. ■

THEOREM 4.2. *Let* f *be an integer with* $1\leqslant f\leqslant n$. *Then the smallest entry in column* f *of the structure matrix* $T=[t_{ij}]$ *for* $\mathfrak{A}(R,S)$ *occurs in row* r_f^* *of* T, *where* $R^*=(r_1^*,\dots,r_n^*)$ *is conjugate to* R.

Proof. Let $q=r_f^*$. By Lemma 4.1 the entries of column f of T form a convex and hence unimodal sequence. Thus it suffices to show that

$$t_{q-1,f}-t_{q,f}\geqslant 0, \qquad t_{qf}-t_{q+1,f}<0. \tag{4.6}$$

It follows from (4.4) that

$$t_{q-1,f}-t_{qf}=r_q-f, \qquad t_{qf}-t_{q+1,f}=r_{q+1}-f. \tag{4.7}$$

But q, being r_f^*, equals the number of terms of $r_1, r_2,\dots r_n$ which are at least f. Since R is monotone, it follows that $r_q\geqslant f$ and $r_{q+1}<f$. Hence (4.6) follows from (4.7). ■

Note that in the proof above we have shown that the last occurrence of the smallest entry in column f of T is in row $q=r_f^*$. Using (4.1), we calculate that

$$\begin{aligned} t_{qf} &= r_f^* f+(r_{q+1}+\cdots+r_m)-(s_1+\cdots+s_f) \\ &= (r_1^*+\cdots+r_f^*)-(s_1+\cdots+s_f), \end{aligned}$$

a relation noted by Anstee [1]. Hence it follows from Theorem 4.2 that T is a nonnegative matrix if and only if $s_1+\cdots+s_f \leqslant r_1^*+\cdots+r_f^*$ for $f=1,\ldots,n$. Since $r_1+\cdots+r_n=s_1+\cdots+s_n$, the latter is equivalent to $S \prec R^*$, thus yielding a direct proof of the equivalence of (2.1.2) and (2.1.3).

Let $U=\{u_0,u_1,\ldots,u_m\}$ and $V=\{v_0,v_1,\ldots,v_n\}$ be disjoint sets of $m+1$ and $n+1$ elements, respectively. Let $K_{m+1,n+1}$ be the *complete bipartite graph* whose set of vertices is $U \cup V$ and whose set of edges consists of all edges joining a vertex in U and a vertex in V. We regard the vertices of U as corresponding to the rows, the vertices of V as corresponding to the columns, and the edges of $K_{m+1,n+1}$ as corresponding to the positions of an $(m+1)\times(n+1)$ matrix. Thus for $i=0,\ldots,m$ and $j=0,\ldots,n$, the edge joining u_i and v_j corresponds to the position (i,j). A *spanning tree* Γ of $K_{m+1,n+1}$ is a connected subgraph of $K_{m+1,n+1}$, consisting of all of the vertices and some of the edges of $K_{m+1,n+1}$, and having no cycles. Thus a spanning tree has $m+n+1$ edges, and these correspond to a set of $m+n+1$ rookwise connected positions of an $(m+1)\times(n+1)$ matrix with no cycles. A particular spanning tree of $K_{m+1,n+1}$ is the spanning tree Γ_0 whose edges correspond to the positions in row 0 and column 0. Thus to prescribe the entries of row 0 and column 0 of the structure matrix T is to prescribe the entries of T at the positions corresponding to a particular spanning tree of $K_{m+1,n+1}$. We show that the entries of the structure matrix T are determined by the entries at the positions corresponding to the edges of any spanning tree. To do this we first establish the following.

LEMMA 4.3. *Let $P=\{(i_1,j_1),(i_1,j_2),(i_2,j_2),\ldots,(i_p,j_p),(i_p,j_1)\}$ be the set of positions of the structure matrix $T=[t_{ij}]$ corresponding to the edges of an elementary cycle of $K_{m+1,n+1}$. Then*

$$\sum_{(i_k,j_l)\in P}(-1)^{k+l}(t_{i_kj_l}-i_kj_l)=0. \tag{4.8}$$

Proof. By (4.1),

$$t_{i_kj_l}-i_kj_l=\sum_{i>i_k} r_i-\sum_{j<j_l} s_j \qquad (k,l=1,\ldots,p). \tag{4.9}$$

Since each row (and column) index either does not appear in P or appears twice, it follows that the term $\sum_{i>i_k} r_i$ (and the term $\sum_{j<j_l} s_j$) of (4.9) occurs twice in (4.8), and the two occurrences have opposite signs. Hence (4.8) holds. ∎

THEOREM 4.4. *The entries of the structure matrix $T=[t_{ij}]$ for $\mathfrak{A}(R,S)$ are determined by the entries of T at the positions corresponding to the edges of a spanning tree of $K_{m+1,n+1}$. Specifically, let Q be the set of positions of T corresponding to the edges of a spanning tree Γ of $K_{m+1,n+1}$. Let (i,j) be a position of T not in Q so that there exists a unique set of positions $P=\{(i,j)=(i_1,j_1),(i_1,j_2),\dots,(i_p,j_1)\}$ corresponding to the edges of an elementary cycle of Γ such that $P'=P-\{(i,j)\}\subseteq Q$. Then*

$$t_{ij}=ij-\sum_{(i_k,j_l)\in P'}(-1)^{k+l}(t_{i_kj_l}-i_kj_l). \tag{4.10}$$

Proof. We apply Lemma 4.3 to P to obtain (4.8), which we then solve for t_{ij}. ■

The recurrence formulas (4.3) and (4.4) can be regarded as a special case of (4.10), obtained by choosing for Γ the spanning tree with edges corresponding to $\{(0,1),\dots,(0,n),(1,0),\dots,(m,0),(k,l)\}$. We note also that applying (4.10) to the spanning tree Γ_0 and using (4.2), we obtain (4.1) for $k=1,\dots,m$ and $l=1,\dots,n$.

Let Γ be the spanning tree of $K_{m+1,n+1}$ whose edges correspond to a set Q of positions of an $(m+1)\times(n+1)$ matrix. For each $(k,l)\in Q$ let c_{kl} be a nonnegative integer. Then Theorem 4.4 can be used to obtain necessary and sufficient conditions for there to exist monotone, nonnegative integral vectors $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ such that $\mathfrak{A}(R,S)$ is nonempty and the structure matrix $T=[t_{ij}]$ for $\mathfrak{A}(R,S)$ satisfies $t_{kl}=c_{kl}$ for $(k,l)\in Q$. Using the notation of that theorem, we obtain the conditions that for each position (i,j) of T not in Q, $c_{ij}\geqslant 0$, where

$$c_{ij}=ij-\sum_{(i_k,j_l)\in P'}(-1)^{k+l}(c_{i_kj_l}-i_kj_l), \tag{4.11}$$

$c_{0n}=0=c_{m0}$, and $c_{00}-c_{01}\geqslant\cdots\geqslant c_{0n-1}-c_{0n}$, $c_{00}-c_{10}\geqslant\cdots\geqslant c_{m-1,0}-c_{m0}$. Some simplification occurs for special spanning trees, and we discuss briefly two instances, the first of which is a generalization of the spanning tree Γ_0.

Let e and f be integers with $0\leqslant e\leqslant m$ and $0\leqslant f\leqslant n$, and let Γ be the spanning tree of $K_{m+1,n+1}$ whose edges correspond to the following set Q of positions of an $(m+1)\times(n+1)$ matrix:

$$Q=\{(e,0),\dots,(e,n)\}\cup\{(0,f),\dots,(m,f)\}.$$

For each $(k,l)\in Q$ let c_{kl} be a nonnegative integer, and define $a_i=c_{i-1,f}-c_{if}$

for $i=1,\dots,m$ and $b_j=c_{e,j-1}-c_{e,j}$ for $j=1,\dots,n$. Defining c_{ij} by (4.11) for each $(i,j)\notin Q$ with $0\leqslant i\leqslant m$, $0\leqslant j\leqslant n$, we see that

$$c_{00}-c_{01}=b_1+e,\dots,\qquad c_{0,n-1}-c_{0n}=b_n+e,$$

$$c_{00}-c_{10}=a_1+f,\dots,\qquad c_{0,m-1}-c_{0m}=a_m+f.$$

Thus an $\mathfrak{A}(R,S)$ with structure matrix $T=[t_{ij}]$ satisfying $t_{kl}=c_{kl}$ for all $(k,l)\in Q$ must satisfy

$$R=(a_1+f,\dots,a_m+f),\qquad S=(b_1+e,\dots,b_n+e).$$

It follows that there exist monotone R and S such that $\mathfrak{A}(R,S)\neq\varnothing$ with structure matrix $T=[t_{ij}]$ satisfying $t_{kl}=c_{kl}$ for all $(k,l)\in Q$ if and only if

$$a_1\geqslant\cdots\geqslant a_m,\qquad b_1\geqslant\cdots\geqslant b_n,$$

$$(b_1+e,\dots,b_n+e)\prec(a_1+f,\dots,a_n+f),$$

$$c_{0f}-c_{ef}+c_{en}=e(n-f),\qquad c_{e0}-c_{ef}+c_{mf}=(m-e)f.$$

The last condition is to assure $c_{0n}=0$, $c_{m0}=0$.

Now let $m=n$, and let Γ be the spanning tree of $K_{n+1,n+1}$ whose edges correspond to the following set Q of positions of an $(n+1)\times(n+1)$ matrix:

$$Q=\{(0,n),(1,n),(1,n-1),\dots,(m,1),(m,0)\}.$$

For each $(k,l)\in Q$ let c_{kl} be a nonnegative integer where $c_{0n}=c_{n0}=0$. Suppose there exists nonnegative integral vectors $R=(r_1,\dots,r_n)$ and $S=(s_1,\dots,s_n)$ such that the structure matrix $T=[t_{ij}]$ for $\mathfrak{A}(R,S)$ satisfies $t_{kl}=c_{kl}$ for all $(k,l)\in Q$. Let $a_1=c_{0n}-c_{1n},\dots,$ $a_n=c_{n-1,1}-c_{n1}$, and let $b_1=c_{n0}-c_{n1},\dots,$ $b_n=c_{1,n-1}-c_{1n}$. Then using (4.11) we obtain

$$t_{0,n-1}-t_{0n}=b_n+1,\quad t_{0,n-2}-t_{0,n-1}=b_{n-1}+2,\dots,\quad t_{00}-t_{01}=b_1+n,$$

$$t_{n-1,0}-t_{n0}=a_n+1,\quad t_{n-2,0}-t_{n-1,0}=a_{n-1}+2,\dots,\quad t_{00}-t_{10}=a_1+n.$$

Thus $R=(a_1+n,\dots,a_n+1)$ and $S=(b_1+n,\dots,b_n+1)$. It follows that there exist monotone R and S with $\mathfrak{A}(R,S)\neq\varnothing$ such that the structure matrix $T=[t_{ij}]$ satisfies $t_{kl}=c_{kl}$ for all $(k,l)\in Q$ if and only if

$$a_1+n\geqslant\cdots\geqslant a_n+1\geqslant 0,$$

$$b_1+n\geqslant\cdots\geqslant b_n+1\geqslant 0,$$

and

$$(b_1+n,\ldots,b_n+1)<(a_1+n,\ldots,a_n+1)^*.$$

We conclude this section by mentioning two problems about which we have essentially no information.

PROBLEM 4.5. *Determine the rank of the structure matrix T for $\mathfrak{A}(R,S)$.* Does it have some combinatorial significance?

PROBLEM 4.6. *Determine the eigenvalues of the structure matrix T for $\mathfrak{A}(R,S)$ in case $m=n$.* Are they of some combinatorial significance?

5. INVARIANT SETS

Throughout this section $R=(r_1,\ldots,r_m)$ *and* $S=(s_1,\ldots,s_n)$ *are nonnegative integral vectors such that* $\mathfrak{A}(R,S)\neq\varnothing$. They are not assumed to be monotone, although at times we shall make this assumption. We say that $\mathfrak{A}(R,S)$ is *column compound* if there exists I with $\varnothing\neq I\subsetneqq\{1,\ldots,m\}$ such that the column sum vector of $A[I,\{1,\ldots,n\}]$ does not depend on the particular choice of $A\in\mathfrak{A}(R,S)$. Similarly, $\mathfrak{A}(R,S)$ is *row compound* if there exists J with $\varnothing\neq J\subsetneqq\{1,\ldots,n\}$ such that the row sum vector of $A[\{1,\ldots,m\},J]$ does not depend on the particular choice of $A\in\mathfrak{A}(R,S)$. We say that $\mathfrak{A}(R,S)$ is *compound* if it is either row or column compound, and *elementary* otherwise. A simple example of a compound $\mathfrak{A}(R,S)$ is obtained by choosing $R=S=(3,2,2)$. Then $\mathfrak{A}(R,S)$ contains only the two matrices

$$\begin{bmatrix}1&1&1\\1&1&0\\1&0&1\end{bmatrix},\qquad\begin{bmatrix}1&1&1\\1&0&1\\1&1&0\end{bmatrix};$$

hence $\mathfrak{A}(R,S)$ is row compound ($J=\{1\}$) and column compound ($I=\{1\}$).

Suppose $\mathfrak{A}(R,S)$ is both row and column compound. It then follows that there exist nonempty $I\subsetneqq\{1,\ldots,m\}$ and nonempty $J\subsetneqq\{1,\ldots,n\}$ such that the row sum vector and column sum vector of $A[I,J]$ do not depend on the particular choice of $A\in\mathfrak{A}(R,S)$. In particular it follows that $\sigma_1(A[I,J])$ is an invariant for $\mathfrak{A}(R,S)$. For $K\subseteq\{1,\ldots,m\}$ and $L\subseteq\{1,\ldots,n\}$, we say that

$K \times L$ is an *invariant set* [17] for $\mathfrak{A}(R,S)$ provided

$$\sigma_1(A[K,L]) = \sigma_1(B[K,L])$$

for all $A, B \in \mathfrak{A}(R,S)$. If $K \times L$ is an invariant set for $\mathfrak{A}(R,S)$, then so are $K \times \bar{L}$, $\bar{K} \times L$, and $\bar{K} \times \bar{L}$, where $\bar{K} = \{1,\dots,m\} - K$, $\bar{L} = \{1,\dots,n\} - L$. Whenever $K = \varnothing$, $K = \{1,\dots,m\}$, $L = \varnothing$, or $L = \{1,\dots,n\}$, then $K \times L$ is an invariant set called a *trivial invariant set*; all other invariant sets are *nontrivial*. From the discussion above it follows that an $\mathfrak{A}(R,S)$ which is both row and column compound has a nontrivial invariant set (cf. Theorem 5.5 and Corollary 5.6). An *invariant position* is an invariant set of cardinality 1. Thus the position (k,l) is invariant provided all or none of the matrices in $\mathfrak{A}(R,S)$ have their (k,l)-entry equal to 1. Suppose $A = [a_{ij}] \in \mathfrak{A}(R,S)$ and $a_{kl} = 1$, where (k,l) is an invariant position of $\mathfrak{A}(R,S)$. Then no sequence of interchanges applied to A can move the 1 in the (k,l)-position, yet every matrix in $\mathfrak{A}(R,S)$ results from A by a sequence of interchanges. Ryser [53] has then called the 1 in the (k,l)-position an *invariant 1* of $\mathfrak{A}(R,S)$. An *invariant 0* is defined similarly. We refer to these as an *invariant 1-position* and *invariant 0-position*, respectively.

LEMMA 5.1. *Suppose $\mathfrak{A}(R,S)$ is row compound, and let J_1 and J_2 be minimal nonempty subsets of $\{1,\dots,n\}$ such that the row sum vector of $A[\{1,\dots,m\},J_k]$ is the same for each $A \in \mathfrak{A}(R,S)$ $(k=1,2)$. Then either $J_1 = J_2$ or $J_1 \cap J_2 = \varnothing$.*

Proof. Suppose $J = J_1 \cap J_2 \neq \varnothing$. Suppose there exist $A = [a_{ij}]$ and $B = [b_{ij}]$ in $\mathfrak{A}(R,S)$ and i with $1 \leqslant i \leqslant m$ such that row i of $A[\{1,\dots,m\},J]$ contains more 1's than row i of $B[\{1,\dots,m\},J]$. Thus there exists $j \in J$ such that $a_{ij} = 1$ while $b_{ij} = 0$. Since $A[\{1,\dots,m\},J_1]$ and $B[\{1,\dots,m\},J_1]$ have the same number of 1's in row i, there exists $j_1 \in J_1 - J$ such that $a_{ij_1} = 0$ while $b_{ij_1} = 1$. The matrix $C = B - A$ has a -1 in the (i,j)-position and a 1 in the (i,j_1)-position. It follows that there exist pairwise disjoint, elementary circuit matrices $C_1,\dots,C_q$ such that $C = C_1 + \cdots + C_q$ and such that the (i,j) entry of C_1 is -1 and the (i,j_1) entry of C_1 is 1. It now follows that $D = A + C_1 \in \mathfrak{A}(R,S)$, and row i of $D[\{1,\dots,m\},J_2]$ contains fewer 1's than row i of $A[\{1,\dots,m\},J_2]$. This is a contradiction, and the lemma is proved. ■

From Lemma 5.1 and its analogue for columns we obtain the following. Suppose $I_1,\dots,I_p$ are the minimal nonempty subsets of $\{1,\dots,m\}$ such that for $i = 1,\dots,p$ the column sum vector of $A[I_i,\{1,\dots,n\}]$ is the same for all $A \in \mathfrak{A}(R,S)$. Suppose $J_1,\dots,J_q$ are the minimal nonempty subsets of $\{1,\dots,n\}$

such that for $j=1,\dots,q$ the row sum vector of $A[\{1,\dots,m\},J_j]$ is the same for all $A\in\mathfrak{A}(R,S)$. Then $I_1,\dots,I_p$ is a partition of $\{1,\dots,m\}$ and $J_1,\dots,J_q$ a partition of $\{1,\dots,n\}$. Thus there exist nonnegative integral vectors $R^{(1)},\dots,R^{(p)}$ and $S^{(1)},\dots,S^{(q)}$ such that for all $A\in\mathfrak{A}(R,S)$, we have $A[I_i,J_j]\in\mathfrak{A}(R^{(i)},S^{(j)})$ for $i=1,\dots,p$ and $j=1,\dots,q$.

We call $I_1,\dots,I_p$ the *row components* and $J_1,\dots,J_q$ the *column components* of $\mathfrak{A}(R,S)$. The sets $I_i\times J_j$ $(i=1,\dots,p,\ j=1,\dots,q)$ are called the *components* of $\mathfrak{A}(R,S)$. It follows that $\mathfrak{A}(R,S)$ is row compound (column compound) if and only if it has at least two column components (row components) and is compound if and only if it has at least two components.

The following theorem, due to Haber [28], is an improvement of a theorem of Ryser [53]. Haber's proof is given in [56, pp. 69–70].

THEOREM 5.2. *Assume R and S are monotone. Suppose (i,j) is an invariant 1-position of $\mathfrak{A}(R,S)$. Then there exist integers e,f with $i\leqslant e\leqslant m$, $j\leqslant f\leqslant n$ such that every matrix $A\in\mathfrak{A}(R,S)$ has the form*

$$A=\begin{bmatrix} J_{ef} & A_1\\ A_2 & 0\end{bmatrix}, \tag{5.1}$$

where J_{ef} is the $e\times f$ matrix all of whose entries equal 1, *and* 0 *is the $(m-e)\times(n-f)$ zero matrix.*

A similar result holds for invariant 0-positions: *if (i,j) is an invariant 0-position, then there exist integers e,f with $0\leqslant e<i$, $0\leqslant f<j$ such that every matrix $A\in\mathfrak{A}(R,S)$ has the form* (5.1).

If some matrix $A\in(\mathfrak{A}(R,S)$ has the form (5.1) where e,f are integers with $0\leqslant e\leqslant m$, $0\leqslant f\leqslant n$, then every matrix in $\mathfrak{A}(R,S)$ has the form (5.1), and it follows that each position (i,j) with $1\leqslant i\leqslant e$, $1\leqslant j\leqslant f$ is an invariant 1-position, while each position (i,j) with $e<i\leqslant m$, $f<j\leqslant n$ is an invariant 0-position. We also note that since for $A\in\mathfrak{A}(R,S)$

$$t_{ef}=\sigma_0(A[\{1,\dots,e\},\{1,\dots,f\}])+\sigma_1(A[\{e+1,\dots,m\},\{f+1,\dots,n\}]),$$

one and hence all matrices $A\in\mathfrak{A}(R,S)$ have the form (5.1) if and only if $t_{ef}=0$. Finally note that in (5.1) both J_{ef} and 0 are vacuous if and only if $e=0$ and $f=n$, or $e=m$ and $f=0$. The following corollaries of Theorem 5.2 are now clear.

COROLLARY 5.3. *Let $mn>1$. If $\mathfrak{A}(R,S)$ has an invariant position, then $\mathfrak{A}(R,S)$ is compound.*

COROLLARY 5.4. *Assume R and S are monotone. Then $\mathfrak{A}(R,S)$ has an invariant position if and only if $t_{ef}=0$ for some pair of integers e,f with $0 \leqslant e \leqslant m$, $0 \leqslant f \leqslant n$, and (e,f) different from $(0,n)$ and $(m,0)$.*

Finally we note that if $t_{ef}=0$ [the form (5.1) holds], then $r_f^* \geqslant e$ and $t_{r_f^*,f}=0$, or equivalently

$$s_1 + \cdots + s_f = r_1^* + \cdots + r_f^*. \tag{5.2}$$

Conversely, if (5.2) holds, then every matrix $A \in \mathfrak{A}(R,S)$ has the form (5.1) where $e = r_f^*$.

While invariant positions are special invariant sets, the following theorem of Brualdi and Ross [17] shows that invariant sets are closely tied to invariant positions.

THEOREM 5.5. *$\mathfrak{A}(R, S)$ has a nontrivial invariant set if and only if it has an invariant position.*

Suppose $\mathfrak{A}(R,S)$ is row compound, so that there exists J with $\varnothing \neq J \subsetneq \{1,\dots,n\}$ such that the row sum vector of $A[\{1,\dots,m\},J]$ is the same for all $A \in \mathfrak{A}(R,S]$. It then follows that $\{(i,j): j \in J\}$ is a nontrivial invariant set for each i with $1 \leqslant i \leqslant m$. Similarly, if $\mathfrak{A}(R,S)$ is column compound, it has a nontrivial invariant set. Hence we have the following corollary, which strengthens Corollary 5.3.

COROLLARY 5.6. *Let $mn > 1$. $\mathfrak{A}(R, S)$ is compound if and only if it has an invariant position.*

We are now in a position to obtain the row and column components of $\mathfrak{A}(R, S)$ from the structure matrix. We assume without loss of generality that R and S are monotone. We first prove the following.

THEOREM 5.7. *Let R and S be monotone and let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R, S)$. Then $\mathfrak{A}(R, S)$ is row compound if and only if $t_{ef}=0$ for some integers e, f with $0 \leqslant e \leqslant m$ and $0<f<n$.*

Proof. Suppose $t_{ef}=0$, where $0 \leqslant e \leqslant m$, $0<f<n$. Then every matrix $A \in \mathfrak{A}(R, S)$ has the form (5.1). Let $K=\{1,\dots, f\}$. Then $\varnothing \neq K \subsetneq \{1,\dots,n\}$, and the row sum vector of $A[\{1,\dots,m\}, K]$ is the same for each matrix $A \in \mathfrak{A}(R, S)$. Now suppose that $\mathfrak{A}(R, S)$ is row compound. Then $n \geqslant 2$ and $\mathfrak{A}(R, S)$ has a nontrivial invariant set and hence by Theorem 5.5 an invariant position. Applying Corollary 5.4, we obtain a pair of integers e, f with $t_{ef}=0$

where $(e,f)\neq(0,n),(m,0)$. If $0<f<n$, the proof is complete. Suppose first that $f=n$, so that $1\leqslant e\leqslant m$. Then the first e rows of each matrix in $\mathfrak{A}(R,S)$ contain only 1's. If $e=m$, then $t_{m1}=0$ and the proof is complete. Now let $1\leqslant e<m$. Let $R'=(r_{e+1},\dots,r_m)$ and $S'=(s_1-e,\dots,s_n-e)$. Then $\mathfrak{A}(R',S')$ is row compound, and we may argue by induction on the number of rows that there exist integers e',f' with $0\leqslant e'\leqslant m-e$ and $0<f'<n$ such that the structure matrix $T'=[t'_{ij}]$ of $\mathfrak{A}(R',S')$ satisfies $t'_{e'f'}=0$. But then $t_{m-e+e',f'}=0$, and the proof is complete in this case. The case $f=0$ can be argued in a similar way to complete the proof of the theorem. ∎

We note that in a similar way one proves that $\mathfrak{A}(R,S)$ is column compound if and only if $t_{ef}=0$ for some integers e,f with $0<e<m$ and $0\leqslant f\leqslant n$.

THEOREM 5.8. *Let R and S be monotone, and let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Define*

$$F=\{f:0<f<n,\ t_{ef}=0\quad \text{for some}\quad e\quad \text{with}\quad 0\leqslant e\leqslant m\},$$

and let the integers in F be arranged in increasing order as $f_1<f_2<\cdots<f_{q-1}$. Then the column components of $\mathfrak{A}(R,S)$ are $J_1=\{1,\dots,f_1\}$, $J_2=\{f_1+1,\dots,f_2\},\dots,$ $J_q=\{f_{q-1}+1,\dots,n\}$.

Proof. We first define $f_0=0$, $f_q=n$. Let $1\leqslant i\leqslant q-1$, and let e_i be an integer such that $t_{e_if_i}=0$. Then every $A\in\mathfrak{A}(R,S)$ has the form

$$A=\begin{bmatrix} J_{e_if_i} & A_1\\ A_2 & 0\end{bmatrix}.$$

Hence the row sum vector of $A[\{1,\dots,m\},\{1,\dots,f_i\}]$ is the same for each $A\in\mathfrak{A}(R,S)$. Since this is true for each i with $1\leqslant i\leqslant q-1$, it now follows that for $j=1,\dots,q$, the row sum vector of $A[\{1,\dots,m\},J_j]$ is the same for each $A\in\mathfrak{A}(R,S)$.

Let $1\leqslant j\leqslant q$, and suppose that there exists a nonempty set K with $K\subsetneqq J_j$ such that the row sum vector R' of $A[\{1,\dots,m\},K]$ is the same for each $A\in\mathfrak{A}(R,S)$. Let e_j be the maximum integer such that $t_{e_jf_j}=0$, and let e_{j-1} be the maximum integer such that $t_{e_{j-1}f_{j-1}}=0$. Then $e_{j-1}\geqslant e_j$. If $e_{j-1}=e_j$, then $t_{e_jf}=0$ for each f with $f_{j-1}<f<f_j$ and we have a contradiction. Suppose $e_{j-1}>e_j$. Let $S'=(s_j-e_j:j\in J_j)$ and let $R'=(r_i-f_{j-1}:e_j+1\leqslant i\leqslant e_{j-1})$. It then follows that $\mathfrak{A}(R',S')$ is row compound, and it follows by applying Theorem 5.7 to $\mathfrak{A}(R',S')$ that there exists e and f with $0\leqslant e\leqslant m$, $f_{j-1}<f<f_j$ such that

$t_{ef}=0$. This contradicts the definition of F. Hence $J_1,\ldots,J_q$ are the column components of $\mathfrak{A}(R,S)$. ■

In an analogous way we obtain the following

THEOREM 5.9. *Let R and S be monotone, and let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Define*

$$E=\{e: 0<e<m,\ t_{ef}=0 \text{ for some } f \text{ with } 0\leqslant f\leqslant n\},$$

and let the integers in E be arranged in increasing order as $e_1<e_2<\cdots<e_{p-1}$. Then the row components of $\mathfrak{A}(R,S)$ are $I_1=\{1,\ldots,e_1\}$, $I_2=\{e_1+1,\ldots,e_2\},\ldots,$ $I_p=\{e_{p-1}+1,\ldots,m\}$.

Once again we assume that R and S are monotone and that $T=[t_{ij}]$ is the structure matrix for $\mathfrak{A}(R,S)$. Since the entries in each column of T form a convex sequence, the zeros in each column of T occur consecutively. Likewise the zeros in each row of T occur consecutively. Suppose i,j,k,l are integers with $0\leqslant i<k\leqslant m$ and $0\leqslant j<l\leqslant n$. Then it follows from the definition of T that not both $t_{ij}=0$ and $t_{kl}=0$. From this we obtain that the zero positions of T can be partitioned into maximal rook paths of the type indicated below:

$$\begin{bmatrix} & & & & & 0 & 0 & 0 \\ & & & & & 0 & & \\ & & & \cdots & 0 & 0 & & \\ & & \vdots & & & & & \\ & & 0 & & & & & \\ 0 & 0 & 0 & & & & & \\ 0 & & & & & & & \end{bmatrix}. \tag{*}$$

Suppose the initial and terminal positions (from right to left) of these paths of type (*) are

$$\begin{array}{rcl} i_0=0,\, j_0=n & \text{and} & k_0, l_0, \\ i_1, j_1 & \text{and} & k_1, l_1, \\ & \vdots & \\ i_{a+1}, j_{a+1} & \text{and} & k_{a+1}=m,\ l_{a+1}=0. \end{array}$$

Then $i_0 \leqslant k_0 \leqslant i_1 - 2$, $i_1 \leqslant k_1 \leqslant i_2 - 2, \ldots, i_{a+1} \leqslant k_{a+1}$ and $j_0 \geqslant l_0 \geqslant j_1 + 2$, $j_1 \geqslant l_1 \geqslant j_2 + 2, \ldots, j_{a+1} \geqslant l_{a+1}$. Moreover it follows that all positions (i, j) with $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$ are invariant positions of $\mathfrak{A}(R, S)$ except the positions in the sets $K_1 \times L_1, \ldots, K_a \times L_a$, where

$$K_u = \{k_{u-1} + 1, \ldots, i_u\}, \qquad u = 1, \ldots, a,$$

$$L_u = \{l_{u-1}, \ldots, j_u + 1\}, \qquad u = 1, \ldots, a.$$

Moreover, the latter sets are invariant sets with no invariant positions. Let $u = 1, \ldots, a$, and let $i \in K_u$. Then each of the positions (i, j) with $j \notin L_u$ is an invariant position of $\mathfrak{A}(R, S)$, so that $\{i\} \times L_u$ is an invariant set and hence a minimal invariant set by Theorem 5.5. Likewise for $u = 1, \ldots, a$ and $j \in L_u$, $K_u \times \{j\}$ is a minimal invariant set.

In particular we have proved the following:

THEOREM 5.10. *Let R and S be monotone. Then there exist pairwise disjoint subsets $K_1, \ldots, K_a$ of $\{1, \ldots, m\}$ and pairwise disjoint subsets $L_1, \ldots, L_a$ of $\{1, \ldots, n\}$ such that the minimal invariant sets of $\mathfrak{A}(R, S)$ are*

$$\{i\} \times L_u \quad \text{for each} \quad i \in K_u \quad (u = 1, \ldots, a), \tag{5.3}$$

$$K_u \times \{j\} \quad \text{for each} \quad j \in L_u \quad (u = 1, \ldots, a), \tag{5.4}$$

$$\{(i, j)\} \quad \text{for each} \quad (i, j) \notin (K_1 \times L_1) \cup \cdots \cup (K_a \times L_a). \tag{5.5}$$

In particular the invariant sets in (5.3) *and* (5.5) *partition* $\{1, \ldots, m\} \times \{1, \ldots, n\}$, *as do the invariant sets in* (5.4) *and* (5.5).

We call the sets $K_1 \times L_1, \ldots, K_a \times L_a$ the *nontrivial components* of $\mathfrak{A}(R, S)$. It follows from Theorems 5.8 and 5.9 that the nontrivial components are indeed components of $\mathfrak{A}(R, S)$ as defined earlier. The nontrivial components contain no invariant positions of $\mathfrak{A}(R, S)$, while each position not in a nontrivial component is an invariant position. Let $A \in \mathfrak{A}(R, S)$, and suppose the row and column sum vectors of $A[K_i, L_i]$ are $R^{(i)}$ and $S^{(i)}$ for $i = 1, \ldots, a$. Then $R^{(i)}$ and $S^{(i)}$ are independent of $A \in \mathfrak{A}(R, S)$, and $\mathfrak{A}(R^{(i)}, S^{(i)})$ has no invariant positions. Moreover, given $A_i \in \mathfrak{A}(R^{(i)}, S^{(i)})$ for $i = 1, \ldots, a$, there is a unique matrix $A \in \mathfrak{A}(R, S)$ with $A[K_i, L_i] = A_i$ for $i = 1, \ldots, a$.

We conclude with an example illustrating the above ideas. Let $m = n = 11$ and $R = (10, 10, 9, 7, 6, 6, 5, 5, 2, 2, 1)$ and $S = (11, 9, 9, 8, 8, 5, 5, 3, 3, 1, 1)$. Using R and S, we construct row 0 and column 0 of the structure matrix $T = [t_{ij}]$, and then using (4.5) we determine the other entries of T column by column to obtain (5.6).

$$T=\begin{bmatrix}
63 & 52 & 43 & 34 & 26 & 18 & 13 & 8 & 5 & 2 & 1 & 0 \\
53 & 43 & 35 & 27 & 20 & 13 & 9 & 5 & 3 & 1 & 1 & 1 \\
43 & 34 & 27 & 20 & 14 & 8 & 5 & 2 & 1 & 0 & 1 & 2 \\
34 & 26 & 20 & 14 & 9 & 4 & 2 & 0 & 0 & 0 & 2 & 4 \\
27 & 20 & 15 & 10 & 6 & 2 & 1 & 0 & 1 & 2 & 5 & 8 \\
21 & 15 & 11 & 7 & 4 & 1 & 1 & 1 & 3 & 5 & 9 & 13 \\
15 & 10 & 7 & 4 & 2 & 0 & 1 & 2 & 5 & 8 & 13 & 18 \\
10 & 6 & 4 & 2 & 1 & 0 & 2 & 4 & 8 & 12 & 18 & 24 \\
5 & 2 & 1 & 0 & 0 & 0 & 3 & 6 & 11 & 16 & 23 & 30 \\
3 & 1 & 1 & 1 & 2 & 3 & 7 & 11 & 17 & 23 & 31 & 39 \\
1 & 0 & 1 & 2 & 4 & 6 & 11 & 16 & 23 & 30 & 39 & 48 \\
0 & 0 & 2 & 4 & 7 & 10 & 16 & 22 & 30 & 38 & 48 & 58
\end{bmatrix}. \tag{5.6}$$

The zero positions of T partition themselves into 4 maximal rook paths with 3 "gaps." Hence $\mathfrak{A}(R,S)$ has 3 nontrivial components, namely,

$$\{1,2\}\times\{10,11\}, \quad \{5,6\}\times\{6,7\}, \quad \{9,10\}\times\{2,3\}.$$

Every matrix in $\mathfrak{A}(R,S)$ then takes the form

$$\left[\begin{array}{ccccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \multicolumn{2}{c}{A_3} \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & & \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & \multicolumn{2}{c}{A_2} & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & & & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & \multicolumn{2}{c}{A_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right].$$

The matrices A_1, A_2, A_3 are 2×2 matrices with row and column sum vector $(1,1)$.

6. TERM RANK

Let A be an $m\times n$ matrix of 0's and 1's. Then k 1's of A are termed *independent* provided no two of them are in the same row or column of A. The *term rank* of A, denoted by $\rho(A)$, equals the maximum integer k for which there exists k independent 1's of A. According to the König-Egérvary theorem (see [56, p. 55]) $\rho(A)$ also equals the minimum number of rows and columns of A which together contain all the 1's of A. The term rank of a matrix is not altered when the rows and the columns are permuted.

Throughout this section we let $R=(r_1,\dots,r_m)$ *and* $S=(s_1,\dots,s_n)$ *be nonnegative integral vectors such that* $\mathfrak{A}(R,S)\neq\varnothing$, and we discuss the term rank of matrices in $\mathfrak{A}(R,S)$. It is a consequence of the discussion of the previous paragraph that there is no loss in generality in assuming that R and S are monotone. Hence we assume *R and S are monotone* throughout this section. Let $T=[t_{ij}]$ denote the structure matrix for $\mathfrak{A}(R,S)$ so that T is a nonnegative matrix. The *maximum term rank* of the matrices in $\mathfrak{A}(R,S)$ is denoted by $\bar{\rho}(R,S)$, while the *minimum term rank* is denoted by $\tilde{\rho}(R,S)$:

$$\bar{\rho}(R,S)=\max\{\rho(A):A\in\mathfrak{A}(R,S)\},$$

$$\tilde{\rho}(R,S)=\min\{\rho(A):A\in\mathfrak{A}(R,S)\}.$$

Formulas for both $\bar{\rho}(R,S)$ and $\tilde{\rho}(R,S)$ are available, and we shall discuss these. But first we note that it follows from the König-Egérvary theorem quoted above that an interchange can alter the term rank by at most 1. Hence if B and C are matrices in $\mathfrak{A}(R,S)$ with $\rho(B)=\tilde{\rho}(R,S)$ and $\rho(C)=\bar{\rho}(R,S)$, and if $B=A_0,A_1,\dots,A_p=C$ is a sequence of matrices in $\mathfrak{A}(R,S)$ such that A_{i+1} is obtainable from A_i by an interchange $(i=0,\dots,p-1)$, then for each k between $\tilde{\rho}(R,S)$ and $\bar{\rho}(R,S)$ there is a matrix A_i in the sequence with $\rho(A_i)=k$. Thus *all integral values intermediate to* $\tilde{\rho}(R,S)$ *and* $\bar{\rho}(R,S)$ *are attained by the term rank of matrices in* $\mathfrak{A}(R,S)$ [54; 56, p. 70].

The formula below for the maximum term rank is due to Ryser [53; 54; 56, p. 75].

THEOREM 6.1. $\bar{\rho}(R,S)=\min\{t_{kl}+k+l:k=0,1,\dots,m;\ l=0,1,\dots,n\}$.

The proof of Theorem 6.1 given by Ryser [56, p. 75] is based on two other results that can be found in [56]. The first of these is due to Haber [28].

THEOREM 6.2. *Let* $\bar{\rho}=\bar{\rho}(R,S)$. *Then there is a matrix* $A_{\bar{\rho}}\in\mathfrak{A}(R,S)$ *with* 1*'s in the positions* $(1,\bar{\rho}),(2,\bar{\rho}-1),\dots,(\bar{\rho},1)$.

The second result is a decomposition theorem for matrices in $\mathfrak{A}(R,S)$ and is proved in [56, pp. 72–74].

THEOREM 6.3. *Let R and S be monotone. Then there exist integers e and f with $0\leqslant e\leqslant m$, $0\leqslant f\leqslant n$ such that for each $A\in\mathfrak{A}(R,S)$,*

$$A=\begin{bmatrix} A_1 & A_3 \\ A_4 & A_2 \end{bmatrix},$$

where A_1 is $e\times f$, and $\sigma_0(A_1)+\sigma_1(A_2)=\bar{\rho}-(e+f)$.

Of course, Theorem 6.3 follows from Theorem 6.1 by taking e and f such that $\bar{\rho}(R,S)=t_{ef}+e+f$. If A is the matrix $A_{\bar{\rho}}$ of Theorem 6.2, then A_1 is the $e\times f$ matrix of all 1's. Recently Brualdi and Ross [16] have given a much different and somewhat more transparent proof of Theorem 6.1. This proof is based on the following [16].

THEOREM 6.4. *Let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$, and let $\bar{R}=(\bar{r}_1,\dots,\bar{r}_m)$, where $\bar{r}_i=r_i$ or r_i-1 $(i=1,\dots,m)$ and $\bar{S}=(\bar{s}_1,\dots,\bar{s}_n)$ with $\bar{s}_j=s_j$ or s_j-1 $(j=1,\dots,n)$. Then there exists a matrix $A\in\mathfrak{A}(R,S)$ and a matrix $B\in\mathfrak{A}(\bar{R},\bar{S})$ such that $B\leqslant A$ if and only if both $\mathfrak{A}(R,S)$ and $\mathfrak{A}(\bar{R},\bar{S})$ are nonempty.*

Suppose $A\in\mathfrak{A}(R,S)$ and $B\in\mathfrak{A}(\bar{R},\bar{S})$, where $B\leqslant A$. Then the matrix $A-B$ is a matrix of 0's and 1's whose 1's form an independent set. Hence $\rho(A)\geqslant\sigma_1(A-B)$. Conversely, if $A\in\mathfrak{A}(R,S)$ and $\rho(A)=p$, then there exists $\bar{R}=(\bar{r}_1,\dots,\bar{r}_m)$ with $\bar{r}_i=r_i$ or r_i-1 $(i=1,\dots,m)$, $\bar{S}=(\bar{s}_1,\dots,\bar{s}_n)$ with $\bar{s}_j=s_j$ or s_j-1 $(j=1,\dots,n)$, and a matrix $B\in\mathfrak{A}(\bar{R},\bar{S})$ such that $B\leqslant A$, $\sigma_1(A-B)=p$, and the 1's of $A-B$ form an independent set. A simple argument using interchanges (see the first paragraph of the proof of our Theorem 6.2 in [56, pp. 70–71]) shows that if R and S are monotone, there exists a matrix A in $\mathfrak{A}(R,S)$ such that $\rho(A[\{1,\dots,\bar{\rho}\},\{1,\dots,\bar{\rho}\}])=\bar{\rho}$. It follows that for monotone vectors $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$, $\bar{\rho}(R,S)$ equals the maximum integer q such that $\mathfrak{A}(R_q,S_q)\neq\varnothing$, where $R_q=(r_1-1,\dots,r_q-1,r_{q+1},\dots,r_m)$ and $S_q=(s_1-1,\dots,s_q-1,s_{q+1},\dots,s_n)$. Unfortunately the vectors R_q and S_q need not be monotone. If they were, Theorem 6.1 would then follow directly from the equivalence of the nonemptiness of $\mathfrak{A}(R_q,S_q)$ and the nonnegativity of the structure matrix for $\mathfrak{A}(R_q,S_q)$. It seems that a more indirect proof like that given in [16] is required.

Haber [29] obtained a formula for the minimum term rank $\tilde{\rho}(R,S)$. Both his formula and proof are quite complicated. Recently we obtained [12] a simpler formula with a proof relying on the supply-demand theorem of network flow theory. Let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. For integers e and f with $0\leqslant e\leqslant m$ and $0\leqslant f\leqslant n$ define

$$\phi(e,f)=\min\{t_{k_1,f+l_2}+t_{e+k_2,l_1}+(e-k_1)(f-l_1)\},$$

where the minimum is taken over all integers k_1,k_2,l_1,l_2 such that

$$0\leqslant k_1\leqslant e\leqslant e+k_2\leqslant m,$$

$$0\leqslant l_1\leqslant f\leqslant f+l_2\leqslant n.$$

THEOREM 6.5. $\tilde{\rho}(R,S)=\min\{e+f:\phi(e,f)\geqslant t_{ef},\ 0\leqslant e\leqslant m,\ 0\leqslant f\leqslant n\}$.

Given an $m\times n$ matrix A of 0's and 1's, then the formulas given in Theorems 6.1 and 6.5 provide upper and lower bounds on the term rank of A in terms of its row and column sum vector.

Algorithms for the construction of matrices in $\mathfrak{A}(R,S)$ having term rank $\bar{\rho}(R,S)$ and $\tilde{\rho}(R,S)$, respectively, have been proposed by Haber [28], and we discuss such algorithms briefly. We continue to assume that R and S are monotone.

Let $\bar{\rho}=\bar{\rho}(R,S)$. According to Theorem 6.2 there is a matrix in $\mathfrak{A}(R,S)$ having 1's in positions $(1,\bar{\rho}),(2,\bar{\rho}-1),\dots,(\bar{\rho},1)$. Let $R=(r_1,\dots,r_m)$, $S=(s_1,\dots,s_n)$, and let $R'(r_1-1,\dots,r_{\bar{\rho}}-1,r_{\bar{\rho}+1},\dots,r_n)$, $S'(s_1-1,\dots,s_{\bar{\rho}}-1,s_{\bar{\rho}+1},\dots,s_n)$. It suffices to construct a matrix $A'\in\mathfrak{A}(R',S')$ with 0's in positions $(1,\bar{\rho}),(2,\bar{\rho}-1),\dots,(\bar{\rho},1)$. The matrix $A\in\mathfrak{A}(R,S)$ obtained from A' by replacing the 0's in these positions with 1's has term rank at least $\bar{\rho}$ and hence equal to $\bar{\rho}$. The matrix A' can be constructed by using algorithms for the construction of feasible flows in supply-demand networks [19]. One sets up a network of sources $x_1,\dots,x_m$ with "supplies" given by the vector R' and sinks with "demands" given by the vector S'. We have an arc from x_i to y_j of capacity 0 if $(i,j)\in\{(1,\bar{\rho}),\dots,(\bar{\rho},1)\}$ and 1 otherwise. A feasible flow in this network corresponds to a matrix $A'\in\mathfrak{A}(R',S')$. Haver [28] has shown that the following algorithm always leads to a matrix B in $\mathfrak{A}(R,S)$ with term rank equal to $\bar{\rho}$. Let $R_n=R$. For this construction we regard R as a column vector and S as a row vector. Suppose $\bar{\rho}<n$. Let ε_n be a column vector of s_n 1's and $m-s_n$ 0's whose 1's are placed as high up as possible consistent with $R_n-\varepsilon_n$ having nonincreasing components. Let $R_{n-1}=R_n-\varepsilon_n$. We construct ε_{n-1} and R_{n-2} in a similar way from R_{n-1} and continue until we obtain $\varepsilon_n,\dots,\varepsilon_{\bar{\rho}+1}$, and $R_{\bar{\rho}}$. The column vectors $\varepsilon_n,\dots,\varepsilon_{\bar{\rho}+1}$ are columns $n,\dots,\bar{\rho}+1$ of the matrix B. Let $S_m=(s_1,\dots,s_{\bar{\rho}})$. Assuming $\bar{\rho}<m$, we apply a similar procedure using the components $r_{\bar{\rho}+1},\dots,r_m$ of R and S_m to obtain $S_{\bar{\rho}}$, and

$\delta_m,\ldots,\delta_{\bar{\rho}+1}$, which are rows $m,\ldots,\bar{\rho}+1$ of B. We are left with constructing the leading $\bar{\rho}\times\bar{\rho}$ submatrix C of B, which has row sum vector $R_{\bar{\rho}}$ and column sum vector $S_{\bar{\rho}}$. To construct C we put a 1 in position $(1,\bar{\rho})$. The remaining 1's in row 1 of C are placed as far to the left as possible and the remaining 1's of C in column $\bar{\rho}$ as high up as possible, consistent with the statement that the matrix obtained from C by deleting row 1 and column $\bar{\rho}$ has a monotone row and column sum vector. We continue like this to construct a row and column at each step until C is determined. An example of the result of this construction is the matrix

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Here $R=(5,3,3,2)$ and $S=(3,3,2,2,2,1)$; $\bar{\rho}(R,S)=4$.

Let $\tilde{\rho}=\tilde{\rho}(R,S)$. The proof of Theorem 6.5 given in [12] begins by showing that $\tilde{\rho}$ equals the minimum of the quantity $e+f$ taken over all pairs e,f such that there exists a matrix $A\in\mathfrak{A}(R,S)$ with $A[\{e+1,\ldots,m\},\{f+1,\ldots,n\}]=0$. Then it is proved that such a matrix exists if and only if $\phi(e,f)\geqslant t_{ef}$. Thus let e,f be integers with $0\leqslant e\leqslant m$, $0\leqslant f\leqslant n$ such that $\phi(e,f)\geqslant t_{ef}$ and $\tilde{\rho}(R,S)=e+f$. A matrix A in $\mathfrak{A}(R,S)$ with term rank $\tilde{\rho}$ corresponds to a feasible flow in the following supply-demand network: there are sources $x_1,\ldots,x_m$ with "supplies" given by $R=(r_1,\ldots,r_m)$, sinks $y_1,\ldots,y_n$ with "demands" given by $S=(s_1,\ldots,s_n)$, and arcs from the x_i to the y_j with capacity 0 if $e+1\leqslant i\leqslant m$ and $f+1\leqslant j\leqslant n$ and capacity 1 otherwise. Such a feasible flow can be constructed using the algorithms described in [19].

We now turn to the investigation of those R and S for which every matrix in $\mathfrak{A}(R,S)$ has the same term rank, that is, for which $\tilde{\rho}(R,S)=\bar{\rho}(R,S)$. Let k be an integer with $1\leqslant k\leqslant n$. If $R=S=K_n$, where K_n is the n-tuple $(k,\ldots,k)$, then it follows from the König-Egérvary theorem that every matrix in $\mathfrak{A}(R,S)$ has term rank equal to n. If S is conjugate to R, then $\mathfrak{A}(R,S)$ contains only one matrix and thus $\tilde{\rho}(R,S)=\bar{\rho}(R,S)$. These two examples show that neither of the two assumptions in the following theorem of Ryser [53] can be dropped. We include the simple proof.

THEOREM 6.5. *Suppose R and S are positive integral vectors, $\mathfrak{A}(R,S)$ has no invariant 1-positions, and $\bar{\rho}(R,S)<\min\{m,n\}$. Then $\tilde{\rho}(R,S)<\bar{\rho}(R,S)$.*

Proof. By Theorem 6.1 there exist integers e and f with $0\leqslant e\leqslant m$, $0\leqslant f\leqslant n$ such that $\bar{\rho}(R,S)=t_{ef}+e+f$. Since $\bar{\rho}(R,S)<m,n$, we have $e<m$

and $f<n$; since R and S are positive, $e>0$ and $f>0$. Let $A\in\mathfrak{A}(R,S)$ have term rank $\bar{\rho}(R,S)$. Then it follows that $A[\{1,\dots,e\},\{1,\dots,f\}]$ is a matrix of all 1's, while $A[\{e+1,\dots,m\},\{f+1,\dots,n\}]$ has exactly t_{ef} 1's. Since the $(1,1)$-position is not an invariant 1-position, there exists a matrix $B=[b_{ij}]\in\mathfrak{A}(R,S)$ with $b_{11}=0$, so that $B[\{e+1,\dots,m\},\{f+1,\dots,n\}]$ contains fewer than t_{ef} 1's. It follows from the König-Egérvary theorem that $\tilde{\rho}(R,S)\leqslant\rho(B)<\bar{\rho}(R,S)$. ■

Ryser [56, p. 76] has asked for a classification of all $\mathfrak{A}(R,S)$ for which $\tilde{\rho}(R,S)=\bar{\rho}(R,S)$. The next theorem gives an answer to this question. First we prove the following.

Lemma 6.7. *Suppose the $m\times n$ matrix A of 0's and 1's has the form*

$$A=\begin{bmatrix} J_{kl} & X_1 \\ X_2 & 0 \end{bmatrix},$$

where J_{kl} is the $k\times l$ matrix of all 1's. Then $\rho(A)=\min\{k+\rho(X_2),\ l+\rho(X_1)\}$. Let $A\in\mathfrak{A}(R,S)$, $X_1\in\mathfrak{A}(R^{(1)},S^{(1)})$, and $X_2\in\mathfrak{A}(R^{(2)},S^{(2)})$. Then

$$\bar{\rho}(R,S)=\min\{k+\bar{\rho}(R^{(2)},S^{(2)}),l+\bar{\rho}(R^{(1)},S^{(1)})\},$$

$$\tilde{\rho}(R,S)=\min\{k+\tilde{\rho}(R^{(2)},S^{(2)}),l+\tilde{\rho}(R^{(1)},S^{(1)})\}.$$

Proof. It follows easily that $\rho(A)\geqslant\rho(X_1)+\rho(X_2)+\min\{k-\rho(X_1),l-\rho(X_2)\}=\min\{k+\rho(X_2),l+\rho(X_1)\}$. By the König-Egérvary theorem equality holds. The lemma now follows. ■

Theorem 6.8. *Let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Then $\tilde{\rho}(R,S)=\bar{\rho}(R,S)$ if and only if there exist integers e,f with $0\leqslant e\leqslant m$, $0\leqslant f\leqslant n$ such that the following hold:*

(i) $t_{ef}=0$.

(ii) *Let* $R^{(1)}=(r_1-f,\dots,r_e-f)$, $S^{(1)}=(s_{f+1},\dots,s_n)$, $R^{(2)}=(r_{e+1},\dots,r_m)$, *and* $S^{(2)}=(s_1-e,\dots,s_f-e)$. *Then* $\tilde{\rho}(R^{(1)},S^{(1)})=e$ *and* $\tilde{\rho}(R^{(2)},S^{(2)})=f$.

Proof. First suppose there exist integers e, f satisfying (i) and (ii). Then every matrix $A\in\mathfrak{A}(R,S)$ has the form

$$A=\begin{bmatrix} J_{ef} & A_1 \\ A_2 & 0 \end{bmatrix},$$

where J_{ef} is the $e\times f$ matrix of all 1's, $A_1\in\mathfrak{A}(R^{(1)},S^{(1)})$, and $A_2\in\mathfrak{A}(R^{(2)},S^{(2)})$. Since $\tilde{\rho}(R^{(1)},S^{(1)})=e$ and $\tilde{\rho}(R^{(2)},S^{(2)})=f$, it follows that $\rho(A_1)=e$ and $\rho(A_2)=f$ and hence $\rho(A)=e+f$. Hence $\tilde{\rho}(R,S)=\bar{\rho}(R,S)=e+f$.

Now suppose that $\tilde{\rho}(R,S)=\bar{\rho}(R,S)$. We prove that there exist e, f such that (i) and (ii) are satisfied by induction on $m+n$. The case $m+n=2$ is easy, so we assume $m+n>2$. First assume that $\mathfrak{A}(R,S)$ has no invariant 1-positions. Let m' be the number of positive coordinates of R, and n' the number of positive coordinates of S. Then it follows from Theorem 6.6 that $\tilde{\rho}(R,S)=\bar{\rho}(R,S)=\min(m',n')$. If $m'\leqslant n'$, then (i) and (ii) hold for $e=m'$, $f=0$; if $n'\leqslant m'$, then (i) and (ii) hold for $e=0$, $f=n'$. We now assume that $\mathfrak{A}(R,S)$ has an invariant 1-position. It follows from Theorem 5.2 that there exist positive integers k and l such that $t_{kl}=0$, so that every matrix $A\in\mathfrak{A}(R,S)$ has the form

$$A=\begin{bmatrix} J_{kl} & X_1\\ X_2 & 0\end{bmatrix},$$

where J_{kl} is the $k\times l$ matrix of all 1's. Let $\bar{\rho}_i$ and $\tilde{\rho}_i$ denote the maximum and minimum term rank for matrices of 0's and 1's having the same row and column sum vector as X_i $(i=1,2)$. By Lemma 6.7, $\bar{\rho}(R,S)=\min\{k+\bar{\rho}_2,l+\bar{\rho}_1\}$ and $\tilde{\rho}(R,S)=\min\{k+\tilde{\rho}_2,l+\tilde{\rho}_1\}$. There are four possibilities depending on the values of these two minimums.

First let $\bar{\rho}(R,S)=k+\bar{\rho}_2\leqslant l+\bar{\rho}_1$ and $\tilde{\rho}(R,S)=k+\tilde{\rho}_2\leqslant l+\tilde{\rho}_1$. Since $\bar{\rho}(R,S)=\tilde{\rho}(R,S)$, $\bar{\rho}_2=\tilde{\rho}_2$. By induction it follows that there exist integers p and q such that every matrix $A\in\mathfrak{A}(R,S)$ has the form

$$A=\left[\begin{array}{cc|c} \multicolumn{2}{c|}{J_{kl}} & X_1\\ \hline J_{pq} & X_3 & \\ X_4 & 0 & 0\end{array}\right],$$

where $\tilde{\rho}_3=p$ and $\tilde{\rho}_4=q$, and hence $\tilde{\rho}_2=p+q$. Here $\tilde{\rho}_3$ and $\tilde{\rho}_4$ are the minimal term ranks for matrices of 0's and 1's having the same row and column sum vector as X_3 and X_4 respectively. Let $e=k+p$ and $f=q$. Then $t_{ef}=0$. Moreover $\tilde{\rho}(R,S)=k+\tilde{\rho}_2=k+p+q$. It follows that the minimal term rank for matrices of 0's and 1's having the same row and column sum vector as

$$\begin{bmatrix} J_{k,l-q} & X_1\\ X_3 & 0\end{bmatrix}$$

equals $k+p$. Hence (i) and (ii) hold in this case.

Now let $\bar{\rho}(R,S)=k+\bar{\rho}_2\leqslant l+\bar{\rho}_1$ and $\tilde{\rho}(R,S)=l+\tilde{\rho}_1\leqslant k+\tilde{\rho}_2$. Then

$$\tilde{\rho}(R,S)=l+\tilde{\rho}_1\leqslant k+\tilde{\rho}_2\leqslant k+\bar{\rho}_2=\bar{\rho}(R,S)=\tilde{\rho}(R,S).$$

Hence $\tilde{\rho}_2=\bar{\rho}_2$. An application of the inductive hypothesis completes the proof as above. The remaining two possibilities are handled in a similar way. The theorem now follows. ∎

We note that in Theorem 6.8, $\mathfrak{A}(R,S)$ has invariant positions unless $e=0$, $f=n$ or $e=m$, $f=0$. If $e=0$, $f=n$, then $\tilde{\rho}(R,S)=\bar{\rho}(R,S)=n$. If $e=m$, $f=0$, then $\tilde{\rho}(R,S)=\bar{\rho}(R,S)=m$. Thus Theorem 6.8 contains Theorem 6.6.

We now consider the following question. Let m and n be positive integers, and let k and l be positive integers with $k\leqslant l$. When do there exist nonnegative integral vectors $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ such that $\tilde{\rho}(R,S)=k$ and $\bar{\rho}(R,S)=l$? A complete answer to this question is given by the next two theorems.

THEOREM 6.9. *If $\tilde{\rho}(R,S)$ is even, then*

$$\bar{\rho}(R,S)\leqslant\tilde{\rho}(R,S)+\left(\frac{\tilde{\rho}(R,S)}{2}\right)^2.$$

If $\tilde{\rho}(R,S)$ is odd, then

$$\bar{\rho}(R,S)\leqslant\tilde{\rho}(R,S)+\frac{\tilde{\rho}(R,S)-1}{2}\,\frac{\tilde{\rho}(R,S)+1}{2}.$$

Proof. As already remarked (and an elementary argument with interchanges shows), there exist nonnegative integers e and f with $e+f=\tilde{\rho}(R,S)$ and a matrix $A\in\mathfrak{A}(R,S)$ all of whose 1's are contained in the first e rows and first f columns. Let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Then

$$\begin{aligned}t_{ef}&=\sigma_0(A[\{1,\dots,e\},\{1,\dots,f\}])+\sigma_1(A[\{e+1,\dots,m\},\{f+1,\dots,n\}])\\&\leqslant ef+0=ef.\end{aligned}$$

By (the easy part of) Theorem 6.1

$$\bar{\rho}(R,S)\leqslant t_{ef}+e+f\leqslant ef+\tilde{\rho}(R,S).$$

The theorem now follows. ∎

THEOREM 6.10. *Let m and n be positive integers, and let k and l be integers. Then there exist nonnegative integral vectors R,S with $\mathfrak{A}(R,S)\neq\varnothing$ such that $\tilde{\rho}(R,S)=k$ and $\bar{\rho}(R,S)=l$ if and only if*

$$0\leqslant k\leqslant l\leqslant\begin{cases}\min\left\{k+\dfrac{k^2}{4},m,n\right\}, & k \text{ even},\\[2ex] \min\left\{k+\dfrac{k^2-1}{4},m,n\right\}, & k \text{ odd}.\end{cases}\tag{6.1}$$

Proof. Since the term rank of an $m\times n$ matrix cannot exceed m or n, it follows from Theorem 6.9 that for given R,S with $\tilde{\rho}(R,S)=k$ and $\bar{\rho}(R,S)=l$, (6.1) holds. Now suppose (6.1) holds. First assume that k is even, say $k=2p$. Let $q=l-k$, so that $q\leqslant\min\{p^2,m-k,n-k\}$. Let $u=m-k-q$ and $v=n-k-q$, so that $u\geqslant 0$ and $v\geqslant 0$. Consider the matrix

$$A=\begin{bmatrix} J_{pp} & J_{pp} & X & J_{pv}\\ J_{pp} & 0 & 0 & 0\\ Y & 0 & I_q & 0\\ J_{up} & 0 & 0 & 0\end{bmatrix},\tag{6.2}$$

where I_q is the identity matrix of order q, the matrices J_{ij} are $i\times j$ matrices of all 1's, X is a $p\times q$ matrix with exactly one 0 in each column and at most p 0's in each row, and Y is a $q\times p$ matrix with exactly one 0 in each row and at most p 0's in each column. Let R and S be the monotone row and column sum vectors of A, and let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Clearly there is a set of $2p+q=l$ 1's of A with no two from the same row or column, and it follows from the König-Egérvary theorem that $\rho(A)=l$. We also note from A that

$$t_{pp}+p+p=q+2p=l.$$

Hence it follows from Theorem 6.1 that $\bar{\rho}(R,S)=l$.

Suppose there exists a matrix $B\in\mathfrak{A}(R,S)$ with $\rho(B)<k$. Hence there exist e rows and f columns of B which together contain all the 1's of B where $e+f<k=2p$. We may suppose that $e<p$. Since each column sum of B is at least p, it follows that $f=n$, which contradicts $k\leqslant n$. It follows that $\rho(B)\geqslant k$ and hence $\tilde{\rho}(R,S)\geqslant k$. But there exists a sequence of interchanges which

transforms the matrix A of (6.2) into a matrix C of the form

$$\begin{bmatrix} Z & J_{p,n-v} \\ J_{m-v,p} & 0 \end{bmatrix}.$$

By the König-Egérvary theorem, $\rho(C) \leqslant 2p = k$ and hence $\rho(C) = k$. It follows that $\tilde{\rho}(R,S) = k$.

The case k odd can be handled in a similar way, and the theorem follows. ∎

We conclude this section with some questions. The first part of the first question is due to Haber [28].

PROBLEM 6.10. *What is the term rank of the special matrix* $\tilde{A}$ *of* $\mathfrak{A}(R,S)$? *Do there exist integers* e *and* f *with* $e+f=\rho(\tilde{A})$ *such that* e *of the rows of* $\tilde{A}$ *with largest row sum and* f *of the columns with largest column sum contain all the* 1's *of* $\tilde{A}$?

We note that $\rho(\tilde{A})$ need not be $\tilde{\rho}(R,S)$ or $\bar{\rho}(R,S)$. For example, let $R = S = (2,1,1)$. Then

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

so that $\rho(\tilde{A}) = 3$. The matrix $A \in \mathfrak{A}(R,S)$ given by

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

shows $\tilde{\rho}(R,S) = 2$. Now let $R = (5,5,2,2,2,2)$ and $S = (5,3,3,3,2,2)$. Then

$$\tilde{A} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Then $\rho(\tilde{A}) = 5$. However, the matrix A obtained from $\tilde{A}$ by the $(2,3;2,3)$-interchange satisfies $\rho(A) = 6$, so that $\bar{\rho}(R,S) = 6$.

If $R=(r_1,\dots,r_m)$, and $S=(s_1,\dots,s_n)$ is conjugate to R, then $\mathfrak{A}(R,S)$ contains only one matrix, which must then be the special matrix $\tilde{A}$ for $\mathfrak{A}(R,S)$. In this case it is not difficult to show by induction the following: Let the distinct components of R be $a_1,a_2,\dots,a_p$, where $a_1>\cdots>a_p$ and a_i occurs e_i times in R $(i=1,\dots,p)$. Then

$$\rho(\tilde{A})=\min\{e_1+\cdots+e_i+a_{i+1}:i=0,1,\dots,p\}.$$

PROBLEM 6.11. *Let $G(R,S)$ be the interchange graph of $\mathfrak{A}(R,S)$. Investigate the distance from a matrix $A\in\mathfrak{A}(R,S)$ to one of maximum (respectively, minimum) term rank.* If $A,B\in\mathfrak{A}(R,S)$ have term rank equal to $\bar{\rho}(R,S)$ [respectively, $\tilde{\rho}(R,S)$], is there a chain joining A and B such that each vertex of the chain is a matrix with term rank equal to $\bar{\rho}(R,S)$ [respectively, $\tilde{\rho}(R,S)$]? This last question can be rephrased as: Do the matrices of maximum term rank generate a connected subgraph of $G(R,S)$?

7. WIDTHS

We assume throughout this section that $R=(r_1,\dots,r_m)$ *and* $S=(s_1,\dots,s_n)$ *are monotone, nonnegative integral vectors such that* $\mathfrak{A}(R,S)\neq\varnothing$. Let α be an integer with $1\leqslant\alpha\leqslant r_m$, and let $A=[a_{ij}]\in\mathfrak{A}(R,S)$. The *$\alpha$-width* $\varepsilon_\alpha(A)$ of A is the smallest number ε of columns that can be selected from A so that the resulting $m\times\varepsilon$ submatrix has at least α 1's in each row. Clearly, $\varepsilon_\alpha(A)$ is an integer with $\alpha\leqslant\varepsilon_\alpha(A)\leqslant n$. The α-width has the following interpretation. Let $Y=\{y_1,\dots,y_n\}$ be a set of n elements, and let $\mathfrak{X}=(X_1,\dots,X_m)$ be a family of subsets of Y such that $y_j\in X_i$ if and only if $a_{ij}=1$ $(i=1,\dots,m,\ j=1,\dots,n)$. Thus A is an incidence matrix for the family $\mathfrak{X}$ of subsets of Y. The α-width of A equals the smallest cardinality of a subset Z of Y such that Z contains at least α elements of each set $X_1,\dots,X_m$. When $\alpha=1$, we have the smallest number of elements in a set of representatives for $\mathfrak{X}$. More specific interpretations of the α-width can be found in [23].

Now let ε be an integer with $1\leqslant\varepsilon\leqslant n$. Then the pair α, ε is a *compatible pair* for $\mathfrak{A}(R,S)$ provided there exists an $A\in\mathfrak{A}(R,S)$ which has an $m\times\varepsilon$ submatrix all of whose row sums are at least α. The *minimum α-width* of the matrices in $\mathfrak{A}(R,S)$ is denoted by $\tilde{\varepsilon}_\alpha(R,S)$, while the *maximum α-width* is denoted by $\bar{\varepsilon}_\alpha(R,S)$:

$$\tilde{\varepsilon}_\alpha(R,S)=\min\{\varepsilon_\alpha(A):A\in\mathfrak{A}(R,S)\},$$

$$\bar{\varepsilon}_\alpha(R,S)=\max\{\varepsilon_\alpha(A):A\in\mathfrak{A}(R,S)\}.$$

The integer $\tilde{\varepsilon}_\alpha(R,S)$ equals the smallest integer ε such that α,ε is a compatible pair for $\mathfrak{A}(R,S)$. A formula for the minimum α-width has been obtained by Fulkerson and Ryser [23, 24] (see Theorem 6.7). However, no formula for the maximum α-width is known, and for good reason. Ryser [56, pp. 126–127] has shown that a solution of the maximum 1-width problem would settle the existence question for finite projective planes. Specifically, let n be an integer at least 2, and let $\mathfrak{A}$ denote the collection of all $(n^2+n+1)\times(n^2+n+1)$ matrices of 0's and 1's with row and column sums all equal to n^2. Then the maximum 1-width is 3 or 2 depending on whether or not there exists a projective plane of order n.

Let α, ε be a compatible pair for $\mathfrak{A}(R,S)$, and consider all matrices $A\in\mathfrak{A}(R,S)$ with $\varepsilon_\alpha(A)\leqslant\varepsilon$. For each such A we consider the collection of all $m\times\varepsilon$ submatrices X with row sums at least α. We define $\delta_{\alpha,\varepsilon}(A)$ to be the smallest integer k such that some submatrix X of A has exactly k row sums equal to α. We call $\delta_{\alpha,\varepsilon}(A)$ *the multiplicity of* α *with respect to* ε *for* A. Clearly, $\delta_{\alpha,\varepsilon}(A)=0$ if and only if $\varepsilon_\alpha(A)<\varepsilon$. The *multiplicity of* α *with respect to* ε *for* $\mathfrak{A}(R,S)$ is the minimum value $\delta_{\alpha,\varepsilon}(R,S)$ of $\delta_{\alpha,\varepsilon}(A)$ taken over all $A\in\mathfrak{A}(R,S)$ with $\varepsilon_\alpha(A)\leqslant\varepsilon$. It follows that $\delta_{\alpha,\varepsilon}(R,S)=0$ if and only if $\tilde{\varepsilon}_\alpha(R,S)<\varepsilon$.

The canonical matrix $\tilde{A}\in\mathfrak{A}(R,S)$ defined in Sec. 2 plays a special role in the determination of the minimum α-width, so that we first obtain a characterization of it. Let $A\in\mathfrak{A}(R,S)$ and for each $j=1,\dots,n$ let R_j be the row sum vector, written as a column vector, of the submatrix formed by the first j columns. The $m\times n$ matrix $M(A)=[R_1\cdots R_n]$ is called the *partial row sum matrix* of A. Note that $R_n^t=R$. We let $\tilde{M}=M(\tilde{A})=[\tilde{R}_1\cdots\tilde{R}_n]$. It follows from the constructive definition of $\tilde{A}$ that $\tilde{R}_j$ is monotone for $j=1,\dots,n$.

An inductive proof of the following lemma can be found in [23].

LEMMA 7.1. *Let $U=(u_1,\dots,u_m)$ and $V=(v_1,\dots,v_m)$ be monotone, nonnegative integral vectors such that $U\prec V$. Let U' be obtained from U by reducing by 1 the k positive components in positions $i_1,\dots,i_k$. Let V' be obtained from V by reducing by 1 the k positive components in positions $j_1,\dots,j_k$. If $i_1\leqslant j_1,\dots,i_k\leqslant j_k$, then $U'\prec V'$.*

The next lemma is readily proved.

LEMMA 7.2. *Let $U=(u_1,\dots,u_m)$ and $V=(v_1,\dots,v_m)$ be nonnegative real vectors such that $U\prec V$ and $V\prec U$. Then there exists a permutation σ of $\{1,\dots,m\}$ such that $U=(v_{\sigma(1)},\dots,v_{\sigma(m)})$. If, in addition, U and V are both monotone, then $U=V$.*

It follows from the above lemma that the relation $\prec$ of majorization is a partial order on the set of monotone, nonnegative vectors $U=(u_1,\ldots,u_m)$. One direction of the following theorem is proved in [24].

THEOREM 7.3. *Let $A^* \in \mathfrak{A}(R,S)$ have partial row sum matrix $M(A^*)=[R_1^* \ \cdots \ R_n^*]$, where $R_1^*,\ldots,R_n^*$ are monotone. Let $A\in\mathfrak{A}(R,S)$ have partial row sum matrix $M(A)=[R_1 \ \cdots \ R_n]$. Then $R_i^* \prec R_i$ $(i=1,\ldots,n)$ for all $A\in\mathfrak{A}(R,S)$ if and only if $A^*=\tilde{A}$.*

Proof. We first show by induction that $\tilde{R}_i \prec R_i$ for $i=1,\ldots,n$. Since $R^t=R_n=\tilde{R}_n$, this is clearly true for $i=n$. Suppose $2\leqslant i\leqslant n$, and assume $\tilde{R}_i\prec R_i$. The vector R_{i-1} is obtained from R_i by reducing s_i components by 1, while a rearrangement of $\tilde{R}_{i-1}$ is obtained by reducing the first s_i components of $\tilde{R}_i$ by 1. Thus by Lemma 7.1, $\tilde{R}_{i-1}\prec R_{i-1}$, so that $\tilde{R}_1\prec R_1,\ldots,\tilde{R}_n\prec R_n$.

Now suppose that $R_i^*\prec R_i$ $(i=1,\ldots,n)$ for each matrix $A\in\mathfrak{A}(R,S)$. By taking $A=\tilde{A}$ we conclude that $R_i^*\prec\tilde{R}_i$ $(i=1,\ldots,n)$. By the above, $\tilde{R}_i\prec R_i^*$ $(i=1,\ldots,n)$. Since R_i^* and $\tilde{R}_i$ are both monotone, it follows by Lemma 6.2 that $R_i^*=\tilde{R}_i$ for $i=1,\ldots,n$ and hence that $A^*=\tilde{A}$. ∎

LEMMA 7.4. *Suppose α, ε is a compatible pair for $\mathfrak{A}(R,S)$. Then there exists a matrix $B\in\mathfrak{A}(R,S)$ such that the submatrix formed by its first ε columns has at least α 1's in each row and exactly $\delta_{\alpha,\varepsilon}(R,S)$ rows with precisely α 1's.*

Proof. Since α, ε is a compatible pair for $\mathfrak{A}(R,S)$, there exists a matrix $A=[a_{ij}]\in\mathfrak{A}(R,S)$ having an $m\times\varepsilon$ submatrix $E=A[\{1,\ldots,m\},\ \{k_1,\ldots,k_\varepsilon\}]$ all of whose row sums are at least α. We choose A and E so that E has exactly $\delta_{\alpha,\varepsilon}(R,S)$ rows with α 1's. Suppose there is a j with $1\leqslant j\leqslant\varepsilon$ such that $j\notin\{k_1,\ldots,k_\varepsilon\}$. Then there exists a k with $\varepsilon<k\leqslant n$ such that $k\in\{k_1,\ldots,k_\varepsilon\}$. If for each $i=1,\ldots,m$, $a_{ik}=1$ implies $a_{ij}=1$, then in E we may replace column k of A by column j. Suppose $p_1,\ldots,p_t$ are those integers such that $a_{p_ij}=0$ while $a_{p_ik}=1$ $(i=1,\ldots,t)$. Since S is monotone, there exist integers $q_1,\ldots,q_t$ such that $a_{q_ij}=1$ while $a_{q_ik}=0$ $(i=1,\ldots,t)$. Replacing 0's by 1's and 1's by 0's in rows $p_1,\ldots,p_t,q_1,\ldots,q_t$ of columns j and k, we obtain a matrix $A'\in\mathfrak{A}(R,S)$ such that the $m\times\varepsilon$ submatrix $F=A'[\{1,\ldots,m\},\ \{k_1,\ldots,k_\varepsilon,j\}-\{k\}]$ has at least α 1's in each row. Moreover, since $s_j\geqslant s_k$, the number of rows of F with precisely α 1's is at most $\delta_{\alpha,\varepsilon}(R,S)$, and hence equals $\delta_{\alpha,\varepsilon}(R,S)$. We may repeat until we obtain a matrix B with the desired properties. ∎

LEMMA 7.5. *Let α, ε be a compatible pair for $\mathfrak{A}(R,S)$. Then the $m\times\varepsilon$ submatrix E of $\tilde{A}$ formed by the first ε columns has at least α 1's in each row and exactly $\delta_{\alpha,\varepsilon}(R,S)$ rows with precisely α 1's.*

Proof. Let B be the matrix of Lemma 7.4, and let the row sum vector of the first ε columns be $R_\varepsilon=(u_1,\dots,u_m)^t$. By Theorem 7.3, $\tilde{R}_\varepsilon \prec R_\varepsilon$, where $\tilde{R}_\varepsilon=(v_1,\dots,v_m)^t$. Hence $v_1+\cdots+v_m=u_1+\cdots+u_m$ and $v_1+\cdots+v_{m-1}\leqslant u_1+\cdots+u_{m-1}$, so that $v_m\geqslant u_m$. Since $u_m\geqslant\alpha$ and $\tilde{R}_\varepsilon$ is monotone, $v_i\geqslant\alpha$ for $i=1,\dots,m$. Suppose E had $t>\delta_{\alpha,\varepsilon}(R,S)$ rows with exactly α 1's. Let $(u_1^*,\dots,u_m^*)^t$ be R_ε rearranged in nonincreasing order. Then by Theorem 7.3, $v_1+\cdots+v_{m-t}\leqslant u_1^*+\cdots+u_{m-t}^*$, so that

$$t\alpha=v_{m-t+1}+\cdots+v_m\geqslant u_{m-t+1}^*+\cdots+u_m^*\geqslant t\alpha.$$

It follows that $u_{m-t+1}^*+\cdots+u_m^*=\alpha$. Hence the submatrix of B formed by the first ε columns has t rows with exactly α 1's, contradicting Lemma 7.4. ∎

The following theorem is by Fulkerson and Ryser [24].

THEOREM 7.6. *Let α be an integer with $1\leqslant\alpha\leqslant r_m$. Then*

$$\tilde{\varepsilon}_\alpha(R,S)=\varepsilon_\alpha(\tilde{A}).$$

Moreover, the αth 1 in row m of $\tilde{A}$ occurs in column $k=\tilde{\varepsilon}_\alpha(R,S)$, and $\delta_{\alpha,k}(R,S)$ equals the number of components of $\tilde{R}_k$ equal to α.

Proof. The theorem is a direct consequence of Lemma 7.5 and the fact that the partial row sum vectors of $\tilde{A}$ are monotone. ∎

Theorem 7.6 affords a simple procedure to determine the minimum α-width of $\mathfrak{A}(R,S)$. Once the matrix $\tilde{A}$ is constructed the minimum α-width can be determined from its last row.

Let e,f,ε be integers with $0\leqslant e\leqslant m$ and $0\leqslant\varepsilon\leqslant f\leqslant n$, and let

$$N(\varepsilon,e,f)=(r_{e+1}+\cdots+r_m)-(s_{\varepsilon+1}+\cdots+s_f)+e(f-\varepsilon). \tag{7.1}$$

Let $A\in\mathfrak{A}(R,S)$, and suppose that

$$A=\begin{bmatrix} * & Y & * \\ X & * & Z \end{bmatrix},$$

where X has size $(m-e)\times\varepsilon$ and Y has size $e\times(f-\varepsilon)$. Then

$$N(\varepsilon,e,f)=\sigma_1(X)+\sigma_0(Y)+\sigma_1(Z).$$

If $T=[t_{ij}]$ is the structure matrix for $\mathfrak{A}(R,S)$, then it is easily verified that

$$N(\varepsilon,e,f)=t_{ef}+t_{e0}-t_{e\varepsilon},\qquad N(0,e,f)=t_{ef}.$$

Thus the numbers (7.1) are generalizations of the invariants t_{ij} for $\mathfrak{A}(R,S)$. The following formula is due to Fulkerson and Ryser [23].

THEOREM 7.7. *The α-width $\tilde{\varepsilon}_\alpha(R,S)$ of $\mathfrak{A}(R,S)$ equals the smallest nonnegative integer ε such that $N(\varepsilon,e,f)\geqslant\alpha(m-e)$ for all integers e,f with $0\leqslant e\leqslant m$, $\varepsilon\leqslant f\leqslant n$.*

We observe the following consequence of Theorems 7.6 and 7.7. The matrix $\tilde{A}$ was defined by indicating how to construct its columns in sequence, beginning with the last column and ending with the first. By using Theorems 7.6 and 7.7 we can also construct in sequence the rows of $\tilde{A}$, beginning with the last row and ending with the first. Let $R=(r_1,\dots,r_m)$, and let $1\leqslant\alpha\leqslant r_m$. Then the r_m 1's in row m of $\tilde{A}$ occur in columns $\tilde{\varepsilon}_\alpha(R,S)$, $\alpha=1,\dots,r_m$. We may now repeat on row $m-1$ and in this way construct all of $\tilde{A}$.

The final theorem of Fulkerson and Ryser [23] follows, since an interchange cannot alter the width of a matrix by more than one.

THEOREM 7.8. *Let ε be an integer with $\tilde{\varepsilon}_\alpha(R,S)\leqslant\varepsilon\leqslant\bar{\varepsilon}_\alpha(R,S)$. Then there exists a matrix $A\in\mathfrak{A}(R,S)$ with $\varepsilon_\alpha(A)=\varepsilon$.*

Further information on widths and multiplicities can be found in [23, 24, 25]. In particular Theorem 4.4 of [23] (cf. (2.13) of [24]) furnishes a formula for the multiplicities $\delta_{\alpha,\varepsilon}(R,S)$. An upper bound for the 1-width of an $m\times n$ matrix of 0's and 1's with row sums at least r and column sums at most c was obtained by Stein [60]. Tarakanov [61] obtained a bound for square matrices with constant row and column sums. Tarakanov's method was later used by Henderson and Dean [38, 39] to obtain improved bounds in terms of r and c. Information on the 1-width of incidence matrices of Steiner triple systems is obtained in [22] and [25].

Finally we note the following concerning Theorem 7.3. Let $A,B\in\mathfrak{A}(R,S)$, and suppose the partial row sum matrices of A and B are, respectively, $[R_1\ \cdots\ R_n]$ and $[R'_1\ \cdots\ R'_n]$. We say that A is *majorized* by B, denoted $A\prec B$, provided $R_i\prec R'_i$ for $i=1,\dots,n$. This relation on the matrices in $\mathfrak{A}(R,S)$ is, in general, only a quasi partial order, since we may have $A\prec B$

and $B \prec A$ for distinct A and B. For instance, if

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

then $A \prec B$ and $B \prec A$. According to Theorem 6.3 the matrix $\tilde{A}$ is a minimal element of this partial order and is the unique minimal element each of whose partial row sum vectors is monotone. In contrast, there need not be a unique maximal element each of whose partial row sum vectors is monotone. To obtain an example, let $R=(3, 3, 3, 2, 1)$ and $S=(4, 2, 2, 2, 2)$. Consider the matrices in $\mathfrak{A}(R, S)$

$$A_1=\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad A_2=\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

whose partial row sum matrices are, respectively,

$$M_1=\begin{bmatrix} 1 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 3 & 3 \\ 1 & 1 & 1 & 2 & 3 \\ 1 & 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad M_2=\begin{bmatrix} 1 & 2 & 3 & 3 & 3 \\ 1 & 2 & 2 & 3 & 3 \\ 1 & 1 & 2 & 3 & 3 \\ 1 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Suppose $B \in \mathfrak{A}(R, S)$ and $A_1 \prec B$. Using the third column of M_1, we conclude that the third partial row sum vector of B has at least two components equal to 3. Hence without loss of generality we may assume B has the form

$$B=\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ * & 0 & 0 & * & * \\ * & 0 & 0 & * & * \\ * & 0 & 0 & * & * \end{bmatrix}.$$

Since the sum of the elements of row 3 of B is 3, the third row of B is determined. Since the fourth partial row sum vector of B majorizes the fourth column of M_1, we now conclude that $B=A_1$. Hence A_1 is a maximal element and each of its partial row sum vectors is monotone. In a similar way one proves A_2 is maximal, and each partial row sum vector of A_2 is monotone.

We conclude with some problems.

PROBLEM 7.9. *Investigate the maximal elements of* $\mathfrak{A}(R,S)$ *each of whose partial row sum vectors are monotone.*

PROBLEM 7.10. *Suppose* $A,B\in\mathfrak{A}(R,S)$ *have the property that for each* $i=1,\ldots,n$ *the ith partial row sum vector of B is a rearrangement of the ith partial row sum vector of A. Is there a sequence of interchanges that transforms A to B such that each intermediary matrix also has this property?* How do the combinatorial invariants of A compare to those of B? For example, does $\rho(A)=\rho(B)$?

PROBLEM 7.11. *For which R and S does* $\tilde{\varepsilon}_\alpha(R,S)=\bar{\varepsilon}_\alpha(R,S)$ *for* $\alpha=1,\ldots,r_m$?

8. IRREDUCIBLE AND FULLY INDECOMPOSABLE MATRICES

In this section we shall be dealing almost exclusively with square matrices. Let $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ be nonnegative integral vectors with $\mathfrak{A}(R,S)\neq\varnothing$, and let $A=[a_{ij}]\in\mathfrak{A}(R,S)$. We now associate with A a *directed graph* $\Gamma(A)$. We take $X=\{x_1,\ldots,x_n\}$ to be a set of n distinct elements which are the *vertices* of $\Gamma(A)$. There is an *arc* (x_i,x_j) from x_i to x_j if and only if $a_{ij}=1$ $(i,j=1,\ldots,n)$. Since there may be 1's on the main diagonal of A, $\Gamma(A)$ may have *loops*, arcs joining a vertex to itself. The number of arcs from vertex x_i, the *outdegree* of x_i, equals r_i, while the number of arcs to x_i, the *indegree* of x_i, equals s_i. Thus R and S are, respectively, the *outdegree and indegree sequences* of the vertices of $\Gamma(A)$. Conversely, given a directed graph Γ with vertex set X whose outdegree sequence is R and whose indegree sequence is S, we can reverse the above steps to construct a matrix $A(\Gamma)\in\mathfrak{A}(R,S)$. We note that $A(\Gamma(A))=A$ and $\Gamma(A(\Gamma))=\Gamma$.

A directed graph Γ is *strongly connected* provided for each ordered pair of vertices x,y there is a directed path from x to y. It is well known [5, p. 30; 31; 63, p. 20] that Γ is strongly connected if and only if the matrix $A(\Gamma)$ is irreducible. Recall that a square matrix A is *reducible* if there is a permutation matrix P such that PAP^t has the form

$$\begin{bmatrix} X_1 & 0 \\ X_{21} & X_2 \end{bmatrix}, \tag{8.1}$$

where X_1 and X_2 are square, nonvacuous matrices. A matrix is *irreducible* if it is not reducible. Irreducible matrices are of fundamental importance in the spectral theory of nonnegative matrices [5,63]. Given a square matrix A, there exists a permutation matrix Q and an integer $t \geqslant 1$ such that QAQ^t has the form

$$\begin{bmatrix} A_1 & 0 & \cdots & 0 \\ A_{21} & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ A_{t1} & A_{t2} & \cdots & A_t \end{bmatrix}, \tag{8.2}$$

where $A_1,\dots,A_t$ are irreducible matrices. The matrices $A_1,\dots,A_t$ are unique up to simultaneous row and column permutations and are called the *irreducible components* of A [they correspond to the strongly connected components of $\Gamma(A)$]. A property more restrictive than that of irreducibility is that of full indecomposability. The $n\times n$ matrix A is called *partly decomposable* if either $n=1$ and $A=[0]$, or $n>1$ and there exist permutation matrices P and Q such that PAQ has the form (8.1). A matrix is *fully indecomposable* if it is not partly decomposable. Fully indecomposable matrices are of primary importance in the study of doubly stochastic matrices. This is because the zero-nonzero pattern of a doubly stochastic matrix is that of a direct sum of fully indecomposable matrices. In general, given an $n\times n$ matrix A with term rank $\rho(A)$ equal to n, there exist permutation matrices U and V such that UAV has the form (8.2) where $A_1,\dots,A_t$ are fully indecomposable. The matrices $A_1,\dots,A_t$ are then unique up to row and column permutations and are called the *fully indecomposable components* of A. The matrix A has *total support* provided that in (8.2), $A_{ij}=0$ whenever $1\leqslant j<i\leqslant t$. Clearly, a fully indecomposable matrix is irreducible but the converse need not hold. However the following is true. Let A be a fully indecomposable matrix of order n. Then it follows from the Kőnig-Egérvary theorem that $\rho(A)=n$. Hence there exists a permutation matrix P such that PA has only 1's on its main diagonal. Then [15, Lemma 2.3; 10, Lemma 2.4] A is fully indecomposable if and only if PA (or $PA-I_n$) is irreducible.

The property of irreducibility is not invariant under row and column permutations. Consequently we cannot assume that R and S are monotone without losing some generality. But we can assume that one of R and S is monotone without loss of generality. This is because for a permutation matrix Q, A is irreducible if and only if QAQ^t is. The property of full indecomposability is invariant under arbitrary row and column permutations, so that when considering this property we may assume without loss of generality that both R and S are monotone.

Beineke and Harary [2] have obtained necessary and sufficient conditions for nonnegative integral vectors R and S to be the indegree and outdegree sequence of a strongly connected directed graph with no loops—equivalently, an irreducible matrix in $\mathfrak{A}(R,S)$ with trace zero. Brualdi [11] has obtained necessary and sufficient conditions for the existence of a fully indecomposable matrix in $\mathfrak{A}(R,S)$. Below we prove one theorem part of which is a direct generalization of the Beineke-Harary theorem and obtain it along with the theorem on fully indecomposable matrices. Since the main part of the proof of the theorem [that (8.3.3) implies (8.3.2)] parallels that given for fully indecomposable matrices in [11], we shall be somewhat brief.

We require the following two lemmas, which are readily proved in terms of the strong connectivity of the corresponding directed graphs.

LEMMA 8.1. *Let C and D be irreducible matrices of 0's and 1's, and let $B=C\oplus D$. Then the matrix obtained from B by a $(p,q;\ u,v)$-interchange where $p(q)$ corresponds to a row of $C(D)$ and $u(v)$ corresponds to a column of $C(D)$ is irreducible.*

LEMMA 8.2. *Let C and D be irreducible matrices of 0's and 1's, and let $B=[b_{ij}]$ be the matrix*

$$\begin{bmatrix} C & 0 \\ X & D \end{bmatrix},$$

where column q of X is not zero. Suppose $b_{pq}=1$ is an entry of C and $b_{pr}=0$ is an entry of 0. Then the matrix obtained from B by changing b_{pq} to 0 and b_{pr} to 1 is irreducible.

We denote by $\tilde{\sigma}(R,S)$ the smallest trace of a matrix in $\mathfrak{A}(R,S)$.

THEOREM 8.3. *Let $n\geqslant 2$, and let $R=(r_1,\dots,r_n)$ and $S=(s_1,\dots,s_n)$ be nonnegative integral vectors such that S is monotone and $\mathfrak{A}(R,S)\neq\varnothing$. Then the following are equivalent:*

(8.3.1) *There exists an irreducible matrix in $\mathfrak{A}(R,S)$.*

(8.3.2) *There exists an irreducible matrix in $\mathfrak{A}(R,S)$ having trace equal to $\tilde{\sigma}(R,S)$.*

(8.3.3) *$r_i>0$ and $s_i>0$ for $i=1,\dots,n$, and for $k=1,\dots,n-1$,*

$$\sum_{i=1}^{k} s_i < \sum_{i=1}^{k} r_i + \sum_{i=k+1}^{n} \min\{k,r_i\}$$

(8.3.4) *$r_i>0$ and $s_i>0$ for $i=1,\dots,n$, and $t(K,L)>0$ whenever K and L are pairwise disjoint, nonempty subsets of $\{1,\dots,n\}$.*

Proof. Trivially, (8.3.2) implies (8.3.1). Suppose (8.3.1) holds, and let $A \in \mathfrak{A}(R,S)$ be irreducible. Since an irreducible matrix of order $n \geqslant 2$ has no row or column of all 0's, $r_i > 0$ and $s_i > 0$ for $i = 1,\ldots,n$. Let K and L be pairwise disjoint nonempty subsets of $\{1,\ldots,n\}$. If $t(K,L) = 0$, then there exists a permutation matrix P such that PAP^t has the form

$$\begin{array}{cc} & \begin{matrix} \overbrace{}^{K} & & \overbrace{}^{L} \end{matrix} \\ K\{ & \left[\begin{array}{c|c|c} & & J \\ \hline 0 & 0 & \end{array}\right] \end{array}$$

where J is a matrix of all 1's, and hence A is reducible. Hence $t(K,L) > 0$ and (8.3.4) holds.

Now assume (8.3.4) holds and let $1 \leqslant k \leqslant n-1$. Suppose that

$$\sum_{i=1}^{k} s_i = \sum_{i=1}^{k} r_i + \sum_{i=k+1}^{n} \min\{k, r_i\}.$$

Considering a matrix $A \in \mathfrak{A}(R,S)$, we see that $r_i \leqslant k$ for $i = 1,\ldots,k$. Let $K = \{i : r_i \geqslant k \text{ and } k < i \leqslant n\}$. Then since $s_n > 0$, it follows that $K \neq \varnothing$ and $t(K,L) = 0$ where $L = \{1,\ldots,k\}$. This contradicts (8.3.4), so that (8.3.3) holds. It remains to show that (8.3.3) implies (8.3.2).

Assume (8.3.3) holds. Let A be a matrix in $\mathfrak{A}(R,S)$ with trace equal to $\tilde{\sigma}(R,S)$, and having the smallest number t of irreducible components of all matrices in $\mathfrak{A}(R,S)$ with trace $\tilde{\sigma}(R,S)$. If $t = 1$, then (8.3.2) holds. So we suppose that $t > 1$. Let Q be a permutation matrix so that QAQ^t has the form (8.2), where $A_1,\ldots,A_t$ are irreducible matrices of orders $n_1,\ldots,n_t$, respectively. We call A_i a *trivial component* of A if $n_i = 1$, and $A_i = [0]$ and a *nontrivial component* otherwise $(i = 1,\ldots,t)$. Since $r_i > 0$ and $s_i > 0$ for $i = 1,\ldots,n$, it follows that A_1 and A_t are nontrivial components. We note that it follows from Lemma 8.1 that if A_i and A_j are nontrivial where $i < j$, then A_{ji} is a matrix of all 1's. First suppose that A has no trivial components. Then QAQ^t has the form

$$\begin{bmatrix} A_1 & 0 \\ J & A_1' \end{bmatrix},$$

and hence (8.3.3) is contradicted when $k = n_1$. Thus we may suppose that A has at least one trivial component. We assume that A has been chosen to have the additional property that the trivial components occur as early as possible in the sequence $A_1,\ldots,A_t$. Let $i > 1$ be the integer such that $A_1,\ldots,A_{i-1}$ are nontrivial and A_i is trivial. Let $k < t$ be the largest integer such that $A_i, \ldots, A_k$ are trivial and

$$\begin{bmatrix} A_i & \cdots & 0 \\ \vdots & & \vdots \\ A_{ki} & \cdots & A_k \end{bmatrix} = 0.$$

Let X be the matrix

$$\begin{bmatrix} A_{k+1,1} & \cdots & A_{k+1,i-1} \\ \vdots & & \vdots \\ A_{t1} & \cdots & A_{t,i-1} \end{bmatrix}. \tag{8.3}$$

If X is a matrix of all 1's, then we contradict (8.3.3) when $k = n_1 + \ldots + n_{i-1}$. Thus we may suppose that X has at least one zero. Consider the first row of X that contains a zero. It must be part of a row of A which contains a trivial component A_j of A. Let the zero occur in a column of A which meets a column of A_r. Suppose A_{jp} contains a 1 for some p with $r < p < i$ or $k < p < j$. Then it follows from Lemma 8.2 that we may apply an interchange to combine A_r and A_p into a single irreducible component without increasing the trace, contradicting our choice of t. Hence we may assume that A_{jp} is a zero matrix for each p with $r < p < i$ or $k < p < j$. It follows that we may assume that $j = k + 1$.

Now by repeated use of Lemma 8.2, it follows that there is a sequence of interchanges which does not increase the trace and which brings A to the form

$$\begin{bmatrix} B & 0 & 0 & 0 \\ C & D & 0 & 0 \\ E & F & 0 & 0 \\ J & J & G & H \end{bmatrix}$$

where the diagonal blocks are square, D has zero trace, and J denotes a matrix of all 1's. If B is $u \times u$ and D is $v \times v$, we contradict (8.3.3) with $k = u + v$. Hence (8.3.2) holds and the proof of the theorem is complete. ■

Corollary 8.4. *Suppose there does not exist and irreducible matrix in $\mathfrak{A}(R,S)$. Then there exists a fixed integer k with $1 \leqslant k \leqslant n-1$ and a fixed permutation matrix P such that for all $A \in \mathfrak{A}(R,S)$, PAP^t has the form*

$$\begin{bmatrix} A_1 & 0 \\ A_{21} & A_2 \end{bmatrix}, \tag{8.4}$$

where A_1 is a $k \times k$ matrix.

COROLLARY 8.5. *Suppose $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ are monotone, positive integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$. Then there exists an irreducible matrix in $\mathfrak{A}(R,S)$.*

Proof. Let $1\leqslant k\leqslant n-1$. Suppose

$$\sum_{i=1}^{k} s_i=\sum_{i=1}^{k} r_i+\sum_{i=k+1}^{n}\min\{k,r_i\}. \tag{8.5}$$

Let $A\in\mathfrak{A}(R,S)$. Then it follows from (8.5) that $r_1\leqslant k$ and hence, since R is monotone, that $r_i\leqslant k$ for $i=1,\ldots,n$. But then by (8.5), $\sum_{i=1}^{k}s_i=\sum_{i=1}^{n}r_i$, so that $s_n=0$, a contradiction. Hence (8.3.3) is satisfied, and by Theorem 8.3 $\mathfrak{A}(R,S)$ contains an irreducible matrix. ■

COROLLARY 8.6. *Suppose $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ are monotone, nonnegative integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$. Let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Then the following are equivalent:*

(8.6.1) *There exists an irreducible matrix in $\mathfrak{A}(R^*,S)$ where $R^*=(r_n,\ldots,r_1)$.*

(8.6.2) *There exists an irreducible matrix in $\mathfrak{A}(R',S)$ for each rearrangement R' of R.*

(8.6.3) *$r_i>0$ and $s_i>0$ for $i=1,\ldots,n$, and $t_{kl}>0$ for positive integers k and l with $k+l\leqslant n$.*

Proof. Clearly (8.6.2) implies (8.6.1). Let K and L be pairwise disjoint, nonempty subsets of $\{1,\ldots,n\}$ with $|K|=k$ and $|L|=l$. Thus k and l are positive integers with $k+l\leqslant n$. Let $R'=(r_1',\ldots,r_n')$ be a rearrangement of R, and let $R^*=(r_1^*,\ldots,r_n^*)=(r_n,\ldots,r_1)$ be the nondecreasing rearrangement of R. Then

$$\sum_{i\notin K} r_i'-\sum_{j\in L} s_j+|K||L|\geqslant\sum_{i=k+1}^{n} r_i-\sum_{j=1}^{l} s_j+kl=t_{kl},$$

where the inequality is equality when $R'=R^*$, and $K=\{n-k+1,\ldots,n\}$, $L=\{1,\ldots,l\}$. Hence it follows from Theorem 8.3 that (8.6.3) implies both (8.6.1) and (8.6.2). It also follows from Theorem 8.3 and the above that (8.6.1) implies (8.6.3). This proves the corollary. ■

In connection with the above results, we remark that it has been proved in [10] that if A is an $n\times n$ matrix of 0's and 1's with positive row and column sum vectors, then there is a permutation matrix P such that PA is irreducible.

The following result, when restricted to condition (8.3.3), is proved by Beineke and Harary [2] in the context of directed graphs. It is an immediate consequence of Theorem 8.3. Let $\mathfrak{A}_0(R,S)$ denote the subset of matrices in $\mathfrak{A}(R,S)$ having only 0's on the main diagonal.

THEOREM 8.7. *Let $n \geqslant 2$, and let $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ be nonnegative integral vectors such that S is monotone and there exists a matrix in $\mathfrak{A}_0(R,S)$. Then there exists an irreducible matrix in $\mathfrak{A}_0(R,S)$ if and only if* (8.3.3) *or* (8.3.4) *holds.*

COROLLARY 8.8. *Suppose $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ are monotone, nonnegative integral vectors. Let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$, and let $R^*=(r_n,\ldots,r_1)$. Suppose $\mathfrak{A}_0(R^*,S)\neq\varnothing$. Then there exists an irreducible matrix in $\mathfrak{A}_0(R^*,S)$ if and only if $r_i>0$ and $s_i>0$ for $i=1,\ldots,n$, and $t_{kl}>0$ for positive integers k and l with $k+l\leqslant n$.*

The proof is basically the same as that of Corollary 8.6 with condition (8.6.2) omitted.

Necessary and sufficient conditions for $\mathfrak{A}_0(R,S)$ to be nonempty have been obtained by Fulkerson [21; 19, pp. 85–86]. In case R and S are both monotone, these conditions simplify to $0\leqslant r_i,s_i\leqslant n-1$ $(i=1,\ldots,n)$, $\sum_{i=1}^n r_i = \sum_{i=1}^n s_i$, and

$$\sum_{i=1}^{k} s_i \leqslant \sum_{i=1}^{k} \min\{k-1,r_i\} + \sum_{i=k+1}^{n} \min\{k,r_i\} \qquad \text{for} \quad k=1,\ldots,n-1.$$

We now turn to fully indecomposable matrices. The following theorem is proved[1] in [11]. We obtain it here from Corollary 8.8.

THEOREM 8.9. *Let $n\geqslant 2$, and let $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ be monotone, nonnegative integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$. Let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Then there exists a fully indecomposable matrix in $\mathfrak{A}(R,S)$ if and only if*

$$r_i \geqslant 2 \text{ and } s_i \geqslant 2 \qquad \text{for} \quad i=1,\ldots,n \tag{8.6a}$$

and

$$t_{kl}+k+l\geqslant n \qquad (k,l=0,1,\ldots,n), \tag{8.6b}$$

with equality only if $k=0$ or $l=0$.

[1]In Lemma 2 of [11] one needs column q of X to be nonzero. The last two paragraphs of the proof of Theorem 1 of [11] need to be restructured as in the proof of Theorem 8.3 here.

Proof. Suppose there exists a fully indecomposable matrix in $\mathfrak{A}(R,S)$. Then using Theorem 6.1 and the Kőnig-Egérvary theorem, one readily verifies (8.6) (see [11]). Conversely, suppose (8.6) holds. From Theorem 6.1 it follows that $\bar{\rho}(R,S)=n$, and then from Theorem 6.2 it follows that there exists a matrix $A\in\mathfrak{A}(R^*,S)$ with all 1's on the main diagonal, where $R^*=(r_n,\ldots,r_1)$. Let $U=(r_1-1,\ldots,r_n-1)$ and $V=(s_1-1,\ldots,s_n-1)$, and let $T'=[t'_{ij}]$ be the structure matrix for $\mathfrak{A}(U,V)$. Let $U^*=(r_n-1,\ldots,r_1-1)$. It then follows from the above that $\mathfrak{A}_0(U^*,V)\neq\varnothing$, and an easy calculation shows that

$$t'_{kl}=t_{kl}+k+l-n \qquad (k,l=0,1,\ldots,n).$$

It now follows from Corollary 8.8 that there exists an irreducible matrix B in $\mathfrak{A}_0(U^*,V)$. The matrix $B+I_n$ is then a fully indecomposable matrix in $\mathfrak{A}(R^*,S)$. The theorem now follows. ■

The following corollary [11] is a consequence of Theorem 8.9.

COROLLARY 8.10. *Let $R'=(r_1,\ldots,r_m)$ and $S'=(s_1,\ldots,s_m)$ be integral vectors with $r_1\geqslant\ldots\geqslant r_n>1=r_{n+1}=\ldots=r_m$ and $s_1\geqslant\ldots\geqslant s_p>1=s_{p+1}=\ldots=s_m$, and suppose $\mathfrak{A}(R',S')\neq\varnothing$. Then there exists a matrix with total support in $\mathfrak{A}(R',S')$ if and only if $n=p$, and $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ satisfy* (8.6).

Using Corollary 8.4 and the proof of Theorem 8.9, we obtain the following.

COROLLARY 8.11. *Suppose there does not exist a fully indecomposable matrix in $\mathfrak{A}(R,S)$. Then there exists a fixed integer k with $1\leqslant k\leqslant n-1$ and fixed permutation matrices P and Q such that for all $A\in\mathfrak{A}(R,S)$, PAQ has the form* (8.4) *where A_1 is a $k\times k$ matrix.*

We now consider when $\mathfrak{A}(R,S)$ contains a partly decomposable matrix. The following lemma is easily proved using interchanges.

LEMMA 8.12. *Let $R=(r_1,\ldots,r_m)$ and $S=(s_1,\ldots,s_n)$ be monotone, nonnegative integral vectors, and suppose for some positive integers k and l there exists a matrix in $\mathfrak{A}(R,S)$ having a $k\times l$ zero submatrix. Then there exists a matrix $A\in\mathfrak{A}(R,S)$ such that $A[\{m-k+1,\ldots,m\},\{n-l+1,\ldots,n\}]=0$.*

THEOREM 8.13. *Let $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ be monotone, nonnegative integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$. Let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Then every matrix in $\mathfrak{A}(R,S)$ is fully indecomposable if* *and only if*

$$\phi(n-k,k) < t_{n-k,k} \qquad (k=1,\ldots,n-1). \tag{8.7}$$

Proof. The function $\phi(\cdot,\cdot)$ was defined in Sec. 6. According to the proof of Theorem 6.5 given in [12], if e and f are integers with $0 \leqslant e \leqslant n$ and $0 \leqslant f \leqslant n$, there exists a matrix $A \in \mathfrak{A}(R,S)$ such that $A[\{n-e+1,\ldots,n\}, \{n-f+1,\ldots,n\}]=0$ if and only if $\phi(e,f) \geqslant t_{ef}$. From this fact and Lemma 8.12, the theorem follows. ■

It is possible to obtain an analogue of Theorem 8.13 for irreducible matrices, for in the manner of the proof of Theorem 6.5 in [12] one can obtain necessary and sufficient conditions in order that there exist a matrix in $\mathfrak{A}(R,S)$ for which a specified submatrix is a zero matrix. Such conditions are rather complicated due to a lack of monotonicity. However, if there is a compatible rearrangement of the row sum vector and column sum vector so that the components of one are increasing while those of the other are decreasing, simplification occurs.

THEOREM 8.14. *Let $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$ be monotone, nonnegative integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$, and let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. Let $R'=(r_n,\ldots,r_1)$. Then every matrix in $\mathfrak{A}(R',S)$ is irreducible if and only if* (8.14) *holds.*

Proof. It follows from Lemma 8.12 that there exists a reducible matrix in $\mathfrak{A}(R',S)$ if and only if there exists a partly decomposable matrix in $\mathfrak{A}(R,S)$. The theorem now follows from Theorem 8.13. ■

Our final result in this section concerns rectangular matrices. An $m\times n$ matrix A is called *decomposable* if there exist permutation matrices P and Q and integers k and l with $0<k+l<m+n$ such that PAQ has the form

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}, \tag{8.8}$$

where A_1 is a $k\times l$ matrix. Thus in (8.8) A_1 or A_2 may be vacuous but must have at least one row or at least one column. The matrix A is *indecomposable* if it is not decomposable. It follows readily that A is indecomposable if and only if the bipartite graph associated with A is connected.

THEOREM 8.15. *Let $R=(r_1,\ldots,r_m)$ and $S=(s_1,\ldots,s_n)$ be nonnegative integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$. There exists an indecomposable* *matrix in* $\mathfrak{A}(R,S)$ *if and only if*

$$\begin{gathered} r_i>0 \quad (i=1,\dots,m), \qquad s_j>0 \quad (j=1,\dots,n), \\ \sum_{i=1}^{m} r_i \geqslant m+n-1. \end{gathered} \tag{8.9}$$

Proof. Suppose $A\in\mathfrak{A}(R,S)$ is indecomposable. Then the associated bipartite graph G is connected. Hence every vertex of G has degree at least 1, and G has at least $m+n-1$ edges (the number of edges of a tree with $m+n$ vertices [32, pp. 32–33]). Hence (8.9) is satisfied. Now suppose (8.9) holds. Let A be a matrix in $\mathfrak{A}(R,S)$ whose corresponding bipartite graph has the smallest number t of connected components. If $t=1$, A is indecomposable. Thus suppose $t>1$. It follows that there exist permutation matrices P and Q such that

$$PAQ=\begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_t \end{bmatrix}, \tag{8.10}$$

where A_i is a (nonvacuous) $m_i\times n_i$ indecomposable matrix for $i=1,\dots,t$. We obtain from (8.9) that $\sigma(A_i)\geqslant m_i+n_i$ for at least one i. Without loss of generality we may take $i=1$. Then A_1 contains a 1 whose replacement by 0 leaves an indecomposable matrix. An interchange involving this 1 of A_1 and any 1 of A_2 results in a matrix in which the matrix

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

is replaced by a matrix of the form

$$B=\begin{bmatrix} A_1' & E \\ F & A_2' \end{bmatrix},$$

where the bipartite graph associated with A_1' is connected and that associated with A_2' has at most two connected components, and E and F each contain exactly one 1. One readily checks that the bipartite graph associated with B is connected. This contradicts the minimality of t, and the theorem is proved. ■

We conclude this section with three problems.

PROBLEM 8.15. *Let $\bar{\sigma}(R,S)$ be the largest trace of a matrix in $\mathfrak{A}(R,S)$. When does there exist an irreducible matrix in $\mathfrak{A}(R,S)$ with trace equal to $\bar{\sigma}(R,S)$?*

PROBLEM 8.16. *An $n\times n$ fully indecomposable (irreducible) matrix A is called* nearly decomposable (nearly reducible) *if each matrix obtained from A by replacing a 1 with a 0 is partly decomposable (reducible). When does $\mathfrak{A}(R,S)$ contain a nearly decomposable (nearly reducible) matrix?*

PROBLEM 8.17. *Let $R=(r_1,\dots,r_m)$ and $S=(s_1,\dots,s_n)$ be monotone, nonnegative integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$. Let $A\in\mathfrak{A}(R,S)$, and let $A(i,j)$ be the matrix obtained from A by eliminating row i and column j $(i=1,\dots,m,\ j=1,\dots,n)$. When does there exist a matrix in $\mathfrak{A}(R,S)$ such that $A(i,j)$ is indecomposable for each $i=1,\dots,m$ and $j=1,\dots,n$?* It is easy to see that the following are necessary conditions: $r_i\geqslant 2$ $(i=1,\dots,m)$, $s_j\geqslant 2$ $(j=1,\dots,n)$, and $t_{11}\geqslant m+n-3$, where $T=[t_{ij}]$ is the structure matrix for $\mathfrak{A}(R,S)$. Are these conditions sufficient?

9. PERMANENT

Let $A=[a_{ij}]$ be an $n\times n$ matrix. The *permanent* of A is defined by

$$\operatorname{per} A=\sum a_{1i_1}\cdots a_{ni_n},$$

where the summation extends over all permutations $(i_1,\dots,i_n)$ of $1,2,\dots,n$. The permanent has been the subject of a recent book [47], and as a result we shall not discuss its properties or applications here. Our interest lies in certain inequalities for the permanent which are of particular significance to $\mathfrak{A}(R,S)$. Throughout we assume that *$R=(r_1,\dots,r_n)$ and $S=(s_1,\dots,s_n)$ are nonnegative integral vectors such that $\mathfrak{A}(R,S)\neq\varnothing$*, and since the permanent is invariant under arbitrary row and column permutations of a matrix, we also assume that both *R and S are monotone*. We record the following observation.

THEOREM 9.1. *There exists $A\in\mathfrak{A}(R,S)$ with* $\operatorname{per} A>0$ *if and only if $\bar{\rho}(R,S)=n$. Moreover,* $\operatorname{per} A>0$ *for all $A\in\mathfrak{A}(R,S)$ if and only if $\tilde{\rho}(R,S)=n$.*

We first discuss some lower bounds for the permanent of matrices in $\mathfrak{A}(R,S)$. The first inequality was first proved by Jurkat and Ryser [41] and later independently by Ostrand [50] in the context of the number of systems of distinct representatives of a family of sets. The criterion for equality is due to Minc [44].

THEOREM 9.2. *Suppose A is a matrix in $\mathfrak{A}(R,S)$ with* $\operatorname{per} A>0$. *Then*

$$\operatorname{per} A \geqslant \prod_{k=1}^{n} \max\{1, r_k+k-n\}. \tag{9.1}$$

Equality occurs if and only if there is a permutation matrix P such that $AP=A(R;n)$.

A special case of (9.1) is the following inequality of Hall [30]: if $r_i \geqslant r$ for $i=1,\dots,n$, then $\operatorname{per} A \geqslant r!$. Rado [51] also proved an inequality weaker than that of (9.1) but which contains the above inequality of Hall.

A matrix $A \in \mathfrak{A}(R,S)$ may have 1's which do not contribute to its permanent, that is there may be 1's whose replacement by 0's does not decrease the permanent. For such matrices inequalities involving the components of R and S cannot be sharp. It follows from the Kőnig-Egérvary theorem that every 1 of A contributes to its permanent if and only if A is a matrix with total support. Since a nonzero matrix with total support, after row and column permutations, is a direct sum of one or more fully indecomposable, and since the permanent of a direct sum of matrices is the product of their permanents, one often considers only fully indecomposable matrices.

The following inequality was first proved by Minc [45]. The criterion for equality is due to Brualdi and Gibson [14], who also gave a geometric interpretation and proof.

THEOREM 9.3. *Let $A \in \mathfrak{A}(R,S)$ be fully indecomposable. Then*

$$\operatorname{per} A \geqslant \left(\sum_{i=1}^{n} r_i\right)-2n+2. \tag{9.2}$$

Equality occurs in (9.2) *if and only if there exist permutation matrices P and Q and an integer p with $0 \leqslant p \leqslant n-1$ such that PAQ has the form*

$$\begin{bmatrix} X & A_1 \\ A_2 & 0 \end{bmatrix}, \tag{9.3}$$

where X is an $(n-p)\times(p+1)$ matrix, and A_1 and A_2^t have exactly two 1's in each column.

A special case of (9.2) is the following. The criterion for equality is due to Minc [46]. It can be obtained as a special case of the equality criterion in Theorem 9.3.

COROLLARY 9.4. *Let $A\in\mathfrak{A}(R,S)$ be fully indecomposable. Then* $\operatorname{per}A \geqslant r_1$, *with equality if and only if $r_2=\dots=r_n=2$.*

An improvement of (9.2) is due to Gibson [27] and Hartifiel [35]. The preceding inequalities are all in terms of the components of the row sum vector R. There are, of course, corresponding inequalities in terms of the column sum vector S.

We now turn to a brief description of upper bounds. Let $A=[a_{ij}]\in\mathfrak{A}(R,S)$. Since

$$\prod_{i=1}^{n} r_i=\prod_{i=1}^{n}\left(\prod_{j=1}^{n} a_{ij}\right)=\operatorname{per}A+(\text{other nonnegative terms}),$$

it follows the $\operatorname{per}A\leqslant\prod_{i=1}^{n}r_i$. Likewise $\operatorname{per}A\leqslant\prod_{i=1}^{n}s_i$, so that

$$\operatorname{per}A\leqslant\min\left\{\prod_{i=1}^{n} r_i,\ \prod_{i=1}^{n} s_i\right\}. \tag{9.4}$$

The following inequality, due to Jurkat and Ryser [42], improves (9.4).

THEOREM 9.5. *Let $A\in\mathfrak{A}(R,S)$. Then*

$$\operatorname{per}A\leqslant\prod_{i=1}^{n}\min\{r_i,s_i\}. \tag{9.5}$$

Let k be a positive integer with $1\leqslant k\leqslant n$. If $r_i=s_i=k$ for $i=1,\dots,n$, then we denote $\mathfrak{A}(R,S)$ by $\mathfrak{A}(k,n)$. Thus $\mathfrak{A}(k,n)$ consists of all $n\times n$ matrices of 0's and 1's with exactly k 1's in each row and column. Each matrix in $\mathfrak{A}(k,n)$ has a positive permanent, indeed by (9.1) a permanent of at least $k!$. Ryser [55] raised the question of determining the maximum permanent of a matrix in $\mathfrak{A}(k,n)$ and conjectured that if k is a divisor of n, then the maximum permanent is $(k!)^{n/k}$. This number is the permanent of the matrix in $\mathfrak{A}(k,n)$ which is the direct sum of n/k $k\times k$ matrices of all 1's. Ryser's conjecture is an immediate consequence of the following inequality, conjectured by Minc [43] and proved first by Brégman [6] and then by Schrijver [58].

THEOREM 9.6. *For* $A \in \mathfrak{A}(R,S)$,

$$\operatorname{per} A \leqslant \prod_{i=1}^{n} (r_i!)^{1/r_i}.$$

For $k=1$ or 2, the matrices in $\mathfrak{A}(k,n)$ have a very simple structure. Hence the first case of interest occurs when $k=3$. The problem of good lower bounds for the permanent of matrices in $\mathfrak{A}(3,n)$ has interested a number of people [13,34,36,37]. Recently, Voorhoeve [64] obtained by elementary means the following exponential lower bound.

THEOREM 9.7. *Let* $A \in \mathfrak{A}(3,n)$. *Then*

$$\operatorname{per} A \geqslant 6\left(\tfrac{4}{3}\right)^{n-3}. \tag{9.6}$$

An upper bound of a nature quite different from that of Theorem 9.6 has been obtained by Foregger [20].

THEOREM 9.8. *Let* $n \geqslant 2$, *and let* $A \in \mathfrak{A}(R,S)$ *be fully indecomposable. Then*

$$\operatorname{per} A \leqslant 2^{p} + 1, \tag{9.7}$$

where

$$p = \sum_{i=1}^{n} r_i - 2n. \tag{9.8}$$

Equality occurs in (9.7) *if and only if there exist an integer* $r \geqslant p$ *and permutation matrices* P *and* Q *such that* PAQ *has the form*

$$\begin{bmatrix} A_1 & 0 & \cdots & 0 & E_1 \\ E_2 & A_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & E_r & A_r \end{bmatrix},$$

where for $i=1,\dots,r$, E_i *is a matrix with exactly one* 1, *and where* p *of the*

A_i's are matrices of order $n_i \geqslant 2$ of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \end{bmatrix}, \tag{9.9}$$

and the other $r-p$ A_i's equal the 1×1 matrix [1].

For arbitrary matrices in $\mathfrak{A}(R,S)$ we have the following result of Brualdi and Gibson [14].

THEOREM 9.9. *Let $A\in\mathfrak{A}(R,S)$ have total support with t fully indecomposable components. Then*

$$\operatorname{per} A \leqslant 2^q, \tag{9.10}$$

where $q=\sum_{i=1}^{n} r_i - 2n + t$. Equality occurs in (9.10) *if and only if q of the fully indecomposable components of A are matrices of order at least* 2 *of the form* (9.9) *and the other $t-q$ equal the 1×1 matrix* [1].

Geometric interpretations of Theorems 9.8 and 9.9 can be found in [14]. We conclude with some problems.

PROBLEM 9.10. *Determine when equality holds in* (9.5).

PROBLEM 9.11. *Suppose k is not a divisor of n. What is the maximum permanent of a matrix in $\mathfrak{A}(k,n)$?*

PROBLEM 9.12. *Since for every matrix A in $\mathfrak{A}(4,n)$ there is a matrix $B\in\mathfrak{A}(3,n)$ for which $B\leqslant A$, the inequality* (9.6) *also applies to matrices in $\mathfrak{A}(4,n)$. Find an improved exponential bound for the permanent of matrices in $\mathfrak{A}(4,n)$.*

We refer to [47] for the connection between Problem 9.12 and the van der Waerden conjecture for the minimum of the permanent of doubly stochastic matrices.

10. OTHER RESULTS AND PROBLEMS

In this final section we discuss a variety of results and problems. We assume that $R=(r_1,\dots,r_m)$ *and* $S=(s_1,\dots,s_n)$ *are monotone, nonnegative integral vectors such that* $\mathfrak{A}(R,S)\neq\varnothing$, and we let $T=[t_{ij}]$ be the structure matrix for $\mathfrak{A}(R,S)$. For $A=[a_{ij}]\in\mathfrak{A}(R,S)$ we define the *trace* of A by

$$\operatorname{tr}(A)=\sum_{i=1}^{\min\{m,n\}} a_{ii}.$$

Extending the notation introduced in Sec. 8, we let

$$\tilde{\sigma}(R,S)=\min\{\operatorname{tr}(A): A\in\mathfrak{A}(R,S)\}$$

and

$$\bar{\sigma}(R,S)=\max\{\operatorname{tr}(A): A\in\mathfrak{A}(R,S)\}.$$

The following result of Ryser [54] evaluates $\tilde{\sigma}(R,S)$ and $\bar{\sigma}(R,S)$ in terms of the structure matrix T.

THEOREM 10.1. *The minimum and maximum trace for matrices in* $\mathfrak{A}(R,S)$ *are given by*

$$\tilde{\sigma}(R,S)=\max\{\min\{k,l\}-t_{kl}: k=0,\dots,m,\ l=0,\dots,n\}, \tag{10.1}$$

$$\bar{\sigma}(R,S)=\min\{\max\{k,l\}+t_{kl}: k=0,\dots,m,\ l=0,\dots,n\}. \tag{10.2}$$

It is easy to show the expression maximized in (10.1) is a lower bound for $\tilde{\sigma}(R,S)$, and the expression minimized in (10.2) is an upper bound for $\bar{\sigma}(R,S)$. Since the trace of a matrix is not invariant under arbitrary row and column permutations, the assumption that R and S are monotone implies some loss of generality. But this assumption is necessary in order that the formulas in Theorem 10.1 be in terms of the structure matrix for $\mathfrak{A}(R,S)$. In addition, Ryser [54] has shown that given an integer p with $\tilde{\sigma}(R,S)\leqslant p\leqslant\tilde{\sigma}(R,S)$, there exists a matrix $A\in\mathfrak{A}(R,S)$ with $\operatorname{tr}(A)=p$ except in two quite specific circumstances; the only omitted traces are $\tilde{\sigma}(R,S)+1$ and $\bar{\sigma}(R,S)-1$.

Now let e and f be integers with $0\leqslant e\leqslant m$ and $0\leqslant f\leqslant n$, and let $H_0(e,f)$ equal the maximum number of 0's which a matrix in $\mathfrak{A}(R,S)$ can have in its leading $e \times f$ submatrix:

$$H_0(e,f) = \max\{\sigma_0(A[\{1,\dots,e\},\{1,\dots,f\}]) : A \in \mathfrak{A}(R,S)\}.$$

Haber [29] has evaluated $H_0(e,f)$ in the following way. Let

$$\psi(e,f) = \min\{t_{ij} : e \leqslant i \leqslant m, f \leqslant j \leqslant n\},$$

and recall the numbers $\phi(e,f)$ defined in Sec. 6.

THEOREM 10.2. *For $0 \leqslant e \leqslant m$ and $0 \leqslant f \leqslant n$,*

$$H_0(e,f) = \min\{\phi(e,f), \psi(e,f)\}.$$

Let $A \in \mathfrak{A}(R,S)$. Then for integers e,f with $0 \leqslant e \leqslant m$ and $0 \leqslant f \leqslant n$,

$$t_{ef} = \sigma_0(A[\{1,\dots,e\},\{1,\dots,f\}]) + \sigma_1(A[\{e+1,\dots,m\},\{f+1,\dots,n\}])$$

and

$$\sum_{i=1}^{e} r_i - ef + \sigma_0(A[\{1,\dots,e\},\{1,\dots,f\}])$$
$$= \sigma_1(A[\{1,\dots,e\},\{f+1,\dots,n\}]).$$

Hence from Theorem 10.2 one obtains formulas for

$$\min\{\sigma_1(A[\{e+1,\dots,m\},\{f+1,\dots,n\}]) : A \in \mathfrak{A}(R,S)\}$$

and

$$\max\{\sigma_1(A[\{1,\dots,e\},\{f+1,\dots,n\}]) : A \in \mathfrak{A}(R,S)\}.$$

Moreover, since

$$\sigma_0(A[\{1,\dots,e\},\{1,\dots,f\}]) + \sigma_1(A[\{1,\dots,e\},\{1,\dots,f\}]) = ef,$$

one also obtains a formula for

$$\min\{\sigma_1(A[\{1,\dots,e\},\{1,\dots,f\}]) : A \in \mathfrak{A}(R,S)\}.$$

Let

$$H_1(e,f)=\max\{\sigma_1(A[\{1,\ldots,e\},\{1,\ldots,f\}]) : A\in\mathfrak{A}(R,S)\}.$$

PROBLEM 10.3. *Find a formula for* $H_1(e,f)$.

We note that

$$H_1(e,f)\leqslant \min\left\{\sum_{i=1}^{e}\min\{r_i,f\},\ \sum_{j=1}^{f}\min\{s_j,e\}\right\}. \tag{10.3}$$

Concerning the preceding problem, one can say the following: There is no matrix $B\in\mathfrak{A}(R,S)$ such that

$$H_1(e,f)=\sigma_1(B[\{1,\ldots,e\},\{1,\ldots,f\}]) \tag{10.4}$$

for all $e=0,\ldots,m$ and all $f=0,\ldots,n$. To see this we take $R=(3,3,3,2,1)$ and $S=(4,2,2,2,2)$. Let the matrix $A\in\mathfrak{A}(R,S)$ be defined by

$$A=\begin{bmatrix}1&1&1&0&0\\1&1&1&0&0\\1&0&0&1&1\\1&0&0&1&0\\0&0&0&0&1\end{bmatrix}.$$

Using (10.3) one readily checks that

$$H_1(e,f)=\sigma_1(A[\{1,\ldots,e\},\{1,\ldots,f\}])$$

for all pairs (e,f) except $(3,4)$. Let $A'\in\mathfrak{A}(R,S)$ be defined by

$$A'=\begin{bmatrix}1&1&1&0&0\\1&1&0&1&0\\1&0&1&1&0\\1&0&0&0&1\\0&0&0&0&1\end{bmatrix}.$$

It follows from (10.3) that

$$H_1(3,4)=\sigma_1(A'[\{1,2,3\},\{1,2,3,4\}])=9.$$

Suppose there were a matrix $B \in \mathfrak{A}(R,S)$ such that (10.4) held for $e,f=0,\ldots,5$. Then it follows from the above remarks that B would have the form

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & & \\ 1 & 0 & 0 & & \\ 0 & 0 & 0 & & \end{bmatrix}$$

and hence $\sigma_1(B[\{1,2,3\},\{1,2,3,4\}]) \leqslant 8$. Hence no such B exists. Perhaps, however, the following is true.

Problem 10.4. *Given $\mathfrak{A}(R,S)$ and an integer f with $0 \leqslant f \leqslant n$, does there exist a matrix $B \in \mathfrak{A}(R,S)$ such that* (10.4) *holds for all $e=0,\ldots,m$?*

Let $A=[a_{ij}] \in \mathfrak{A}(R,S)$, and let β be an integer with $1 \leqslant \beta \leqslant r_1$. The *$\beta$-column matching number* $\mu_\beta(A)$ of A is defined to be the maximum number μ of columns of A that can be selected so that the resulting $m \times \mu$ submatrix has at most β 1's in each row. Clearly, $\mu_\beta(A)$ is an integer with $\beta \leqslant \mu_\beta(A) \leqslant n$. There is the following important interpretation. Let $E=\{e_1,\ldots,e_m\}$ be a set of m elements and consider the collection $\mathcal{F}(R,S)$ of families (or hypergraphs) $H=(F_1,\ldots,F_n)$ of subsets (or edges) of E such that F_j contains s_j elements $(j=1,\ldots,n)$ and e_i is a member of r_i sets of H $(i=1,\ldots,m)$. Let $H=(F_1,\ldots,F_n)$ be the family corresponding to A so that (cf. Sec. 1) $a_{ij}=1$ if and only if $e_i \in F_j$ $(i=1,\ldots,m;\ j=1,\ldots,n)$. Then $\mu_\beta(A)$ equals the maximum number of sets of the family H which can be selected so that each element of E occurs in at most β sets. In particular, $\mu_1(A)$ *equals the maximum number of pairwise disjoint sets of H.* In the terminology of hypergraphs [4], a collection of pairwise disjoint edges of H is called a *matching* of H, and the maximum number of edges in a matching is its *matching number* $\tau(H)$. Thus $\tau(H)=\mu_1(A)$. For $1 \leqslant \beta \leqslant r_1$, we let

$$\tilde{\mu}_\beta(R,S)=\min\{\,\mu_\beta(A): A \in \mathfrak{A}(R,S)\}$$

and

$$\bar{\mu}_\beta(R,S)=\max\{\,\mu_\beta(A): A \in \mathfrak{A}(R,S)\}.$$

A particular case of interest occurs when S is the n-tuple $(2,\ldots,2)$. Then $\mathcal{F}(R,S)$ becomes the collection of graphs with degree sequence R and n edges. In this case $\bar{\mu}_1(R,S)$ equals the maximum matching number of all such graphs. A formula for the matching number of a graph is due to Tutte [62] and Berge [3].

Fulkerson and Ryser [24] have shown that the minimum width numbers $\tilde{\varepsilon}_{\alpha}(R,S)$ of Sec. 7 determine the maximum matching numbers $\bar{\mu}_{\beta}(R,S)$. First we extend the domain of the parameter β in $\mu_{\beta}(A)$ to $0 \leqslant \beta \leqslant r_1$ by defining $\mu_0(A)$ to be 0. We also extend the domain of α in the α-width $\varepsilon_{\alpha}(A)$ to $0 \leqslant \alpha \leqslant r_m$ by defining $\varepsilon_0(A)$ to be 0. Let $R^c = (n-r_1, \ldots, n-r_n)$ and $S^c = (m - s_1, \ldots, m-s_n)$. Since the components of R and S are not increasing, the components of R^c and S^c are not decreasing. Let $0^c = 1$ and $1^c = 0$. Then there is a one-to-one correspondence between matrices in $\mathfrak{A}(R,S)$ and those in $\mathfrak{A}(R^c, S^c)$ given by $A \leftrightarrow A^c$ where $A = [a_{ij}] \in \mathfrak{A}(R,S)$ and $A^c = [a_{ij}^c] \in \mathfrak{A}(R^c, S^c)$. Under this correspondence, the following is readily verified [24]: Let $0 \leqslant \beta \leqslant r_1$ and $0 \leqslant \alpha \leqslant n - r_1$. Then $\mu_{\beta}(A) \geqslant \alpha + \beta$ if and only if $\varepsilon_{\alpha}(A^c) \leqslant \alpha + \beta$. From this one obtains the following [24].

THEOREM 10.5. *Let $A \in \mathfrak{A}(R,S)$ and let β be an integer with $0 \leqslant \beta \leqslant r_1$. If α is the largest integer such that $0 \leqslant \alpha \leqslant n - r_1$ and $\varepsilon_{\alpha}(A^c) - \alpha \leqslant \beta$, then $\mu_{\beta}(A) = \alpha + \beta$. Conversely, if $\mu_{\beta}(A) = \alpha + \beta$, then α is the largest integer such that $0 \leqslant \alpha \leqslant n - r_1$ and $\varepsilon_{\alpha}(A^c) - \alpha \leqslant \beta$.*

Similarly, let $A \in \mathfrak{A}(R,S)$, and let α be an integer with $0 \leqslant \alpha \leqslant n - r_1$. If β is the smallest integer such that $0 \leqslant \beta \leqslant r_1$ and $\mu_{\beta}(A) - \beta \geqslant \alpha$, then $\varepsilon_{\alpha}(A^c) = \alpha + \beta$. Conversely, if $\varepsilon_{\alpha}(A^c) = \alpha + \beta$, then β is the smallest integer such that $0 \leqslant \beta \leqslant r_1$ and $\mu_{\beta}(A) - \beta \geqslant \alpha$.

COROLLARY 10.6. *Let β be an integer with $0 \leqslant \beta \leqslant r_1$. Let α be the largest integer such that $0 \leqslant \alpha \leqslant n - r_1$ and $\tilde{\varepsilon}_{\alpha}(R^c, S^c) - \alpha \leqslant \beta$. Then $\bar{\mu}_{\beta}(R,S) = \alpha + \beta$.*

From Theorems 7.6 and 10.5 one obtains the following.

COROLLARY 10.7. *Let β be an integer with $0 \leqslant \beta \leqslant r_1$. For the canonical matrix $\tilde{A}$ of $\mathfrak{A}(R^c, S^c)$, the following hold:*

(10.7.1) $\bar{\mu}_{\beta}(R,S) = \mu_{\beta}(\tilde{A}^c)$.

(10.7.2) *If the $(\beta+1)$st zero of the last row of $\tilde{A}$ occurs in column j, then $\bar{\mu}_{\beta}(R,S) = j - 1$.*

We now make some comments concerning Theorem 6.4. Anstee [1] has observed that the technique of proof of that theorem [16] can be used to obtain the following interesting result.

THEOREM 10.8. *Let $R = (r_1, \ldots, r_m)$ and $S = (s_1, \ldots, s_n)$ and let k be a nonnegative integer. Let $\bar{R} = (\bar{r}_1, \ldots, \bar{r}_m)$ and $\bar{S} = (\bar{s}_1, \ldots, \bar{s}_n)$ where $\bar{r}_i = r_i - k$ or*

$r_i - k - 1$ for $i = 1,\ldots,m$ and where $\bar{s}_j \leqslant s_j$ for $j = 1,\ldots,n$. Then there exists a matrix $A \in \mathfrak{A}(R,S)$ and a matrix $B \in \mathfrak{A}(\bar{R},\bar{S})$ such that $B \leqslant A$ if and only if both $\mathfrak{A}(R,S)$ and $\mathfrak{A}(\bar{R},\bar{S})$ are nonempty.

In connection with Theorems 6.4 and 10.8, the following has been conjectured independently by the author and Anstee [1].

CONJECTURE 10.9. Let $R, \bar{R}, S, \bar{S}$ be nonnegative integral vectors. Then there exists a matrix $A \in \mathfrak{A}(R,S)$ and a matrix $B \in \mathfrak{A}(\bar{R},\bar{S})$ such that $A + B \in \mathfrak{A}(R+\bar{R}, S+\bar{S})$ if and only if $\mathfrak{A}(R,S)$, $\mathfrak{A}(\bar{R},\bar{S})$, and $\mathfrak{A}(R+\bar{R}, S+\bar{S})$ are all nonempty.

Although there are general conditions for the existence of a matrix in $\mathfrak{A}(R,S)$ with 0's in prescribed places (cf. Sec. 2), for certain patterns of 0's these conditions simplify. Such is the case for the following theorem of Anstee [1].

THEOREM 10.10. *Let $R = (r_1,\ldots,r_n)$ and $S = (s_1,\ldots,s_n)$ be monotone, nonnegative integral vectors. Then there exists a triangular matrix in $\mathfrak{A}(R,S)$ of the form*

$$\begin{bmatrix} * & * & \cdots & * & * \\ * & * & \cdots & * & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ * & * & \cdots & 0 & 0 \\ * & 0 & \cdots & 0 & 0 \end{bmatrix}$$

if and only if $\mathfrak{A}(R,S) \neq \varnothing$, and $r_i, s_i \leqslant n - i + 1$ for $i = 1,\ldots,n$.

We now discuss another natural parameter. We denote the *rank* of a matrix A by $\nu(A)$. If $\mathfrak{A}(R,S) \neq \varnothing$, then we define

$$\tilde{\nu}(R,S) = \min\{\nu(A) : A \in \mathfrak{A}(R,S)\}$$

and

$$\bar{\nu}(R,S) = \max\{\nu(A) : A \in \mathfrak{A}(R,S)\}.$$

In a private communication, Ryser has posed the following.

PROBLEM 10.11. *Find a general formula for $\tilde{\nu}(R,S)$.*

We may also pose the following problem, which, however, does not seem tractable.

PROBLEM 10.12. *Find a general formula for* $\bar{\nu}(R,S)$.

Let k be an integer with $1 \leqslant k \leqslant n$, and let R and S both equal the n-tuple $(k,\ldots,k)$. Let $\bar{\nu}(k,n)=\bar{\nu}(R,S)$ and $\tilde{\nu}(k,n)=\tilde{\nu}(R,S)$. Then the results of Houck and Paul [40] are equivalent to the following.

THEOREM 10.13. *For* $1 \leqslant k \leqslant n$,

$$\bar{\nu}(k,n)=\begin{cases} 1 & \text{if} \quad k=n, \\ 3 & \text{if} \quad k=2 \text{ and } n=4, \\ n & \text{otherwise.} \end{cases}$$

A less ambitious problem than Problem 10.12 is the following.

PROBLEM 10.14. *Find a formula for* $\tilde{\nu}(k,n)$.

Some progress on Problems 10.11 and 10.12 perhaps can be made by finding good bounds for $\bar{\nu}(R,S)$ and $\tilde{\nu}(R,S)$. As Ryser has observed, it is not difficult to show that given an integer t with $\tilde{\nu}(R,S) \leqslant t \leqslant \bar{\nu}(R,S)$, there exists a matrix $A \in \mathfrak{A}(R,S)$ with $\nu(A)=t$.

Now let $A=[a_{ij}]$ be an $n \times n$ nonnegative matrix, and let $A_{(0,1)}=[b_{ij}]$ be the $n \times n$ matrix of 0's and 1's where $b_{ij}=1$ if and only if $a_{ij}>0$. If A has r_i positive entries in row i and s_i positive entries in column i $(i=1,\ldots,n)$, then $A_{(0,1)} \in \mathfrak{A}(R,S)$ where $R=(r_1,\ldots,r_n)$ and $S=(s_1,\ldots,s_n)$. Sinkhorn and Knopp [59] have proved the following theorem. Another proof and a generalization were obtained by Engel and Schneider [18].

THEOREM 10.14. *Let* $A=[a_{ij}]$ *be an* $n \times n$ *fully indecomposable, nonnegative matrix. Then there exist diagonal matrices* D_1 *and* D_2 *with positive main diagonals such that* $D_1 A D_2 = A_{(0,1)}$ *if and only if there exists a positive number* c *such that* $a_{1i_1} \cdots a_{ni_n}=0$ *or* c *for each permutation* $i_1,\ldots,i_n$ *of* $1,\ldots,n$.

To conclude we mention that for given R and S, $\mathfrak{A}(R,S)$ can be enlarged to include all $m \times n$ nonnegative matrices with row sum vector R and column sum vector S. The resulting collection $\mathfrak{A}^*(R,S)$ of matrices (there is no need to assume R and S are integral now) is a convex subset of Euclidean mn-space and has been extensively investigated. One reason for interest in $\mathfrak{A}^*(R,S)$ is that it forms the domain for the important Hitchcock transportation problem [19, p. 95–111]. A special case of great interest is the set of $n \times n$ *doubly stochastic* matrices which results when both R and S equal the n-tuple $(1,\ldots,1)$. We shall not include an extensive bibliography but mention the articles [7], [9], [14], [42] and Chapter 4 of the recent book [5]. Additional references can be found in these sources.

REFERENCES

1 R. Anstee, Doctoral Dissertation, California Institute of Technology, 1980.
2 L. W. Beineke and F. Harary, Local restrictions for various classes of directed graphs, *J. London Math. Soc.* 40:87–95 (1965).
3 C. Berge, Sur le couplage maximum d'un graphe, *C. R. Acad. Sci. Paris* 247:258–259 (1958).
4 C. Berge, *Graphs and Hypergraphs*, Dunod, Paris, 1970.
5 A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic, New York, 1979.
6 L. M. Brégman, Certain properties of nonnegative matrices and their permanents, *Dokl. Adad. Nauk SSSR* 211:27–30 (1973) [*Soviet Math. Dokl.* 14:945–949 (1973)].
7 R. A. Brualdi, Convex sets of nonnegative matrices, *Canad. J. Math.* 20:144–157 (1968).
8 R. A. Brualdi, Existence theorem for infinite integral matrices, *Mathematika* 17:91–97 (1970).
9 R. A. Brualdi, Combinatorial properties of symmetric nonnegative matrices, in *Proceedings of the International Conference on Combinatorial Theory*, Rome, Sept. 1973, Academia Nazionale dei Lincei, Roma, pp. 99–120.
10 R. A. Brualdi, Matrices permutation equivalent to irreducible matrices and applications, *Linear and Multilinear Algebra* 7:1–12 (1979).
11 R. A. Brualdi, Matrices of 0's and 1's with total support, *J. Combinatorial Theory* Ser. A, 28:249–256 (1980).
12 R. A. Brualdi, On Haber's minimum term rank formula, submitted for publication.
13 R. A. Brualdi, On 1-factors of cubic bipartite graphs, *Ars Combinatoria*, to appear.
14 R. A. Brualdi and P. M. Gibson, The convex polyhedron of doubly stochastic matrices: I. Applications of the permanent function, *J. Combinatorial Theory Ser. A* 22:194–230 (1977).
15 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).
16 R. A. Brualdi and J. A. Ross, On Ryser's maximum term rank formula, *Linear Algebra and Appl.* 29:33–38 (1980).
17 R. A. Brualdi and J. A. Ross, Invariant sets for classes of matrices of zeros and ones, *Proc. Amer. Math. Soc.*, to appear in 1980.
18 G. M. Engel and H. Schneider, Cyclic and diagonal products on a matrix, *Linear Algebra and Appl.* 7:301–335 (1973).
19 L. R. Ford, Jr. and D. R. Fulkerson, *Flows in Networks*, Princeton U. P., Princeton, N.J., 1962.
20 T. H. Foregger, An upper bound for the permanent of a fully indecomposable matrix, *Proc. Amer. Math. Soc.* 49:319–324 (1975).
21 D. R. Fulkerson, Zero-one matrices with zero trace, *Pacific J. Math.* 10:831–836 (1960).
22 D. R. Fulkerson, G. L. Nemhauser, and L. E. Trotter, Jr., Two computationally difficult set covering problems that arise in computing the 1-width of incidence

matrices of Steiner triple systems, MRC Technical Summary Report #1411, Madison, Wis., Feb. 1974.
23 D. R. Fulkerson and H. J. Ryser, Widths and heights of (0, 1)-matrices, *Canad. J. Math.* 13:239–255 (1961).
24 D. R. Fulkerson and H. J. Ryser, Multiplicities and minimal widths for (0, 1)-matrices, *Canad. J. Math.* 14:498–508 (1962).
25 D. R. Fulkerson and H. J. Ryser, Width sequences for special classes of (0, 1)-matrices, *Canad. J. Math.* 15:371–396 (1963).
26 D. Gale, A theorem on flows in networks, *Pacific J. Math.* 7:1073–1082 (1957).
27 P. M. Gibson, A lower bound for the permanent of a (0, 1)-matrix, *Proc. Amer. Math. Soc.* 33:245–246 (1972).
28 R. M. Haber, Term rank of 0, 1 matrices, *Rend. Sem. Mat. Univ. Padova* 30:24–51 (1960).
29 R. M. Haber, Minimal term rank of a class of (0, 1)-matrices, *Canad. J. Math.* 15:188–192 (1963).
30 M. Hall, Jr., Distinct representatives of subsets, *Bull. Amer. Math. Soc.* 54:922–926 (1948).
31 F. Harary, A graph theoretic method for the complete reduction of a matrix with a view toward finding its eigenvalues, *J. Math. and Physics* 38:104–111 (1959).
32 F. Harary, *Graph Theory*, Addison-Wesley, Reading, Mass., 1969.
33 G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge U. P., 1934.
34 D. J. Hartfiel, A canonical form for fully indecomposable (0, 1)-matrices, *Canad. Math Bull.* 18:223–227 (1975).
35 D. J. Hartfiel, A lower bound on the permanent of a (0, 1)-matrix, *Proc. Amer. Math. Soc.* 39:83–85 (1973).
36 D. J. Hartfiel and J. W. Crosby, On the permanent of a certain class of (0, 1)-matrices, *Canad. Math. Bull.* 14:507–511 (1971).
37 D. J. Hartfiel and J. W. Crosby, A lower bound for the permanent on $U_n(k, k)$, *J. Combinatory Theory* 12:283–288 (1972).
38 J. R. Henderson and R. A. Dean, The 1-width of (0, 1)-matrices having constant row sum 3, *J. Combinatorial Theory Ser. A* 16:355–370 (1974).
39 J. R. Henderson and R. A. Dean, A general upper bound for 1-widths, *J. Combinatorial Theory Ser. A*, 18:236–238 (1975).
40 D. J. Houck and M. E. Paul, Nonsingular 0-1 matrices with constant row and column sums, *Linear Algebra and Appl.* 22:263–266 (1978).
41 W. B. Jurkat and H. J. Ryser, Matrix factorizations of determinants and permanents, *J. Algebra* 3:1–27 (1966).
42 W. B. Jurkat and H. J. Ryser, Term ranks and permanents of nonnegative matrices, *J. Algebra* 5:342–357 (1967).
43 H. Minc, Upper bounds for permanents of (0, 1)-matrices, *Bull. Amer. Math. Soc.* 69:789–791 (1963).
44 H. Minc, A lower bound for permanents of (0, 1)-matrices, *Proc. Amer. Math. Soc.* 18:1128–1132 (1967).
45 H. Minc, On lower bounds for permanents of (0, 1)-matrices, *Proc. Amer. Math. Soc.* 22:117–123 (1969).

46 H. Minc, (0, 1)-matrices with minimal permanents, *Israel J. Math.* 15:27–30 (1973).
47 H. Minc, *Permanents*, Encyclopedia of Mathematics and its Applications, Vol. 6, Addison-Wesley, Reading, Mass., 1978.
48 L. Mirsky, Combinatorial theorems and integral matrices, *J. Combinatorial Theory* 5:30–44 (1968).
49 L. Mirsky, *Transversal Theory*, Mathematics in Science and Engineering, Vol. 75, Academic, New York, 1971.
50 P. A. Ostrand, Systems of distinct representatives, II, *J. Math. Anal. Appl.* 32:1–4 (1970).
51 R. Rado, On the number of systems of distinct representatives of sets, *J. London Math. Soc.* 42:321–324 (1967).
52 H. J. Ryser, Combinatorial properties of matrices of zeros and ones, *Canad. J. Math.* 9:371–377 (1957).
53 H. J. Ryser, The term rank of a matrix, *Canad. J. Math.* 10:57–65 (1958).
54 H. J. Ryser, Traces of matrices of zeros and ones, *Canad. J. Math.* 12:463–476 (1960).
55 H. J. Ryser, Matrices of zeros and ones, *Bull. Amer. Math. Soc.* 66:442–464 (1960).
56 H. J. Ryser, *Combinatorial Mathematics*, Carus Mathematical Monograph No. 14, Math. Assoc. of Amer., Washington, 1963.
57 H. J. Ryser, Matrices of zeros and ones in combinatorial mathematics, in *Recent Advances in Matrix Theory*, Univ. of Wis. Press, Madison, 1964, pp. 103–124.
58 A. Schrijver, A short proof of Minc's conjecture, *J. Combinatorial Theory Ser. A* 25:80–83 (1978).
59 R. Sinkhorn and P. Knopp, Problems involving diagonal products in nonnegative matrices, *Trans. Amer. Math. Soc.* 136:67–75 (1969).
60 S. K. Stein, Two combinatorial covering theorems, *J. Combinatorial Theory Ser. A* 16:391–397 (1974).
61 V. E. Tarakanov, Maximal depth of a class of (0, 1)-matrices, *Math. USSR—Sb.* 4:3–12 (1968).
62 W. T. Tutte, The factorization of linear graphs, *J. London Math. Soc.* 22:107–111 (1947).
63 R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962.
64 M. Voorhoeve, A lower bound for the permanents of certain (0, 1)-matrices, *Indag. Math.* 41:83–86 (1979).
65 D. W. Walkup, Minimal interchanges of (0, 1)-matrices and disjoint circuits in a graph, *Canad. J. Math.* 17:831–838 (1965).

Received 15 February 1980; revised 14 April 1980