1

Incidence Matrices

1.1 Fundamental Concepts

Let

$$A = [a_{ij}], \quad (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$$

be a matrix of m rows and n columns. We say that A is of *size* m by n, and we also refer to A as an m by n matrix. In the case that $m = n$ then the matrix is square of *order* n. It is always assumed that the entries of the matrix are elements of some underlying field F. Evidently A is composed of m row vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ over F and n column vectors $\beta_1, \beta_2, \ldots, \beta_n$ over F, and we write

$$A = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} = \begin{bmatrix} \beta_1 & \beta_2 & \ldots & \beta_n \end{bmatrix}.$$

It is convenient to refer to either a row or a column of the matrix as a *line* of the matrix. We use the notation A^T for the transpose of the matrix A. We always designate a zero matrix by O, a matrix with every entry equal to 1 by J, and the identity matrix of order n by I. In order to emphasize the size of these matrices we sometimes include subscripts. Thus $J_{m,n}$ denotes the all 1's matrix of size m by n, and this is abbreviated to J_n if $m = n$. The notations $O_{m,n}$, O_n and I_n have similar meanings. In displaying a matrix we often use $*$ to designate a submatrix of no particular concern. The $n!$ permutation matrices of order n are obtained from I_n by arbitrary permutations of the lines of I_n. A permutation matrix P of order n satisfies the matrix equations

$$PP^T = P^T P = I_n.$$

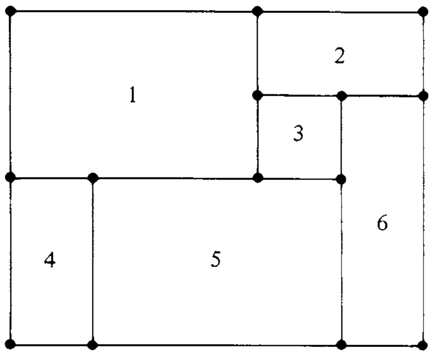

Figure 1.1

In most of our discussions the underlying field F is the field of real numbers or the subfield of the rational numbers. Indeed we will be greatly concerned with matrices whose entries consist exclusively of the integers 0 and 1. Such matrices are referred to as (0,1)-*matrices*. and they play a fundamental role in combinatorial mathematics.

We illustrate this point by an example that reformulates an elementary problem in geometry in terms of (0,1)-matrices. Let a rectangle R in the plane be of integral height m and of integral length n. Let all of R be partitioned into t smaller rectangles. Each of these rectangles is also required to have integral height and integral length. We number these smaller rectangles in an arbitrary manner $1, 2, \ldots, t$. An example with $m = 4$, $n = 5$ and $t = 6$ is illustrated in Figure 1.1.

We associate with the partitioned rectangle of Figure 1.1 the following two (0,1)-matrices of sizes 4 by 6 and 6 by 5, respectively:

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

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

The number of 1's in column i of X is equal to the height of rectangle i, and all of the 1's in column i occur consecutively. The topmost and bottommost 1's in column i of X locate the position of rectangle i with respect to the top and bottom horizontal lines of the full rectangle. The matrix Y behaves in much the same way but with respect to rows. Thus the number of 1's in row j of Y is equal to the length of rectangle j, and all the 1's in row j of Y occur consecutively. The first and last 1's in row j of Y locate the position of rectangle j with respect to the left and right vertical lines of the full rectangle. It follows from the definition of matrix multiplication that the product of the matrices X and Y satisfies

$$XY = J. \tag{1.1}$$

The matrix J in (1.1) is the matrix of 1's of size 4 by 5.

This state of affairs is valid in the general case. Thus the partitioning of a rectangle in the manner described is precisely equivalent to a matrix equation of the general form (1.1) with the following constraints. The matrices X and Y are (0,1)-matrices of sizes m by t and t by n, respectively, and J is the matrix of 1's of size m by n. The 1's in the columns of X and the 1's in the rows of Y are required to occur consecutively. If the original problem is further restricted so that all rectangles involved are squares, then we must require in addition that $m = n$ and that the sum of column i of X is equal to the sum of row i of Y, $(i = 1, 2, \ldots, t)$.

We next describe in more general terms the basic link between (0,1)-matrices and a wide variety of combinatorial problems. Let

$$X = \{x_1, x_2, \ldots, x_n\}$$

be a nonempty set of n elements. We call X an *n-set*. Now let $X_1, X_2, \ldots, X_m$ be m not necessarily distinct subsets of the n-set X. We refer to this collection of subsets of an n-set as a *configuration* of subsets. Vast areas of modern combinatorics are concerned with the structure of such configurations. We set $a_{ij} = 1$ if $x_j \in X_i$, and we set $a_{ij} = 0$ if $x_j \notin X_i$. The resulting (0,1)-matrix

$$A = [a_{ij}], \quad (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$$

of size m by n is the *incidence matrix* for the configuration of subsets $X_1, X_2, \ldots, X_m$ of the n-set X. The 1's in row α_i of A display the elements in the subset X_i, and the 1's in column β_j display the occurrences of the element x_j among the subsets. Thus the lines of A give us a complete description of the subsets and the occurrences of the elements within these subsets. This representation of our configuration in terms of the (0,1)-matrix A is of the utmost importance because it allows us to apply the powerful techniques of matrix theory to the particular problem under investigation.

Let A be a (0,1)-matrix of size m by n. The *complement* C of the incidence matrix A is obtained from A by interchanging the roles of the 0's and 1's and satisfies the matrix equation

$$A + C = J.$$

We note that the matrices O and J of size m by n are complementary and correspond to the configurations with the empty set repeated m times and the full n-set repeated m times, respectively. A second incidence matrix associated with the (0,1)-matrix A of size m by n is the transposed matrix A^T of size n by m. The configuration of subsets associated with a transposed incidence matrix is called the *dual* of the configuration.

Suppose that we have subsets $X_1, X_2, \ldots, X_m$ of an n-set X and subsets $Y_1, Y_2, \ldots, Y_m$ of an n-set Y. Then these two configurations of subsets are regarded as the same or *isomorphic* provided that we may relabel the subsets $X_1, X_2, \ldots, X_m$ and the elements of the n-set X so that the resulting configuration coincides with the configuration $Y_1, Y_2, \ldots, Y_m$ of the n-set Y. This means that our original configurations are the same apart from the notation in which they are written.

The above isomorphism concept for configurations of subsets has a direct interpretation in terms of the incidence matrices that represent the configurations. Thus suppose that A and B are two (0,1)-matrices of size m by n that represent incidence matrices for subsets $X_1, X_2, \ldots, X_m$ of an n-set X and for subsets $Y_1, Y_2, \ldots, Y_m$ of an n-set Y, respectively. Then these configurations of subsets are isomorphic if and only if A is transformable to B by line permutations. In other words the configurations of subsets are isomorphic if and only if there exist permutation matrices P and Q of orders m and n, respectively, such that

$$PAQ = B.$$

In many combinatorial investigations we are often primarily concerned with those properties of a (0,1)-matrix that remain invariant under arbitrary permutations of the lines of the matrix. The reason for this is now apparent because such properties of the matrix become invariants of isomorphic configurations.

If two configurations of subsets are isomorphic, then their associated incidence matrices of size m by n are required to satisfy a number of necessary conditions. Thus the row sums including multiplicities are the same for both matrices and similarly for the column sums. The ranks of the two matrices must also coincide. The incidence matrices may be tested for invariants like these. But thereafter it may still be an open question as to whether or not the given configurations are isomorphic. Suppose, for example, that A is a (0,1)-matrix of order n such that all of the line sums of A are equal to the positive integer k. We may ask if the configuration associated with this incidence matrix is isomorphic to its dual. This will be the case if and only if there exist permutation matrices P and Q of order n such that

$$PAQ = A^T.$$

We now look at some examples of the isomorphism problem which are readily solvable. Suppose that the configuration is represented by a permutation matrix of order n. Then clearly the configuration may be represented equally well by the identity matrix I_n, and thus any two such configurations are isomorphic.

Suppose next that the configuration is represented by a (0,1)-matrix A of order n such that all of the line sums are equal to 2. This restriction on A allows us to replace A under line permutations by a direct sum of the form

$$A_1 \oplus A_2 \oplus \cdots \oplus A_e.$$

Each of the components of this direct sum has line sums equal to 2, and is "fully indecomposable" in the sense that it cannot be further decomposed by line permutations into a direct sum. Each fully indecomposable component is itself normalized by line permutations so that the 1's appear on the main diagonal and in the positions directly above the main diagonal with an additional 1 in the lower left corner. For example, a normalized fully indecomposable component of order 5 is given by the matrix

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

We have constructed a canonical form for A in the sense that if

$$B_1 \oplus B_2 \oplus \cdots \oplus B_f$$

is a second decomposition for A, then we have $e = f$ and the A_i are equal to the B_j in some ordering. The essential reasoning behind this is as follows. We first label all of the 1's in A from 1 to $2n$ in an arbitrary manner. Under line permutations two labeled 1's in a line always remain within the same line. Consider the component A_1 and its two labeled 1's in the (1,1) and (1,2) positions of A_1. These labeled 1's occur in some row of a component of the second decomposition, say in component B_i. But then the labeled 1 in the (2,2) position of A_1 also occurs in B_i and similarly for the labeled 1 in the (2,3) position of A_1. In this way we see that all of the labeled 1's in A_1 occur in B_i. But then A_1 and B_i are equal because both matrices are fully indecomposable and normalized. We may then identify the labeled 1's in the component A_2 with the labeled 1's in another component B_j of the second decomposition and so on.

The above canonical form for A implies that we have devised a straightforward procedure for deciding the isomorphism problem for configurations whose incidence matrices have all of their line sums equal to 2. The situation for (0,1)-matrices that have all of their line sums equal to 3 is vastly more complicated. Such matrices may already possess a highly intricate internal structure.

In the study of configurations of subsets two broad categories of problems emerge. The one deals with the structure of very general configurations, and the other deals with the structure of much more restricted configurations. In the present chapter we will see illustrations of both types of problems. We will begin with the proof of a minimax theorem that holds for an arbitrary (0,1)-matrix. Later we will also discuss certain (0,1)-matrices of such a severely restricted form that their very existence is open to question.

1.2 A Minimax Theorem

We now prove the fundamental minimax theorem of König[1936]. This theorem has a long history and many ramifications which are described in detail in the book by Mirsky[1971]. The theorem deals exclusively with properties of a (0,1)-matrix that remain invariant under arbitrary permutations of the lines of the matrix.

Theorem 1.2.1. *Let A be a $(0,1)$-matrix of size m by n. The minimal number of lines in A that cover all of the 1's in A is equal to the maximal number of 1's in A with no two of the 1's on a line.*

Proof. We use induction on the number of lines in A. The theorem is valid in case that $m = 1$ or $n = 1$. Hence we take $m > 1$ and $n > 1$. We let ρ' equal the minimal number of lines in A that cover all of the 1's in A, and we let ρ equal the maximal number of 1's in A with no two of the 1's on a line. We may conclude at once from the definitions of ρ and ρ' that $\rho \leq \rho'$. Thus it suffices to prove that $\rho \geq \rho'$. A minimal covering of the 1's of A is called *proper* provided that it does not consist of all m rows of A or of all n columns of A. The proof of the theorem splits into two cases.

In the first case we assume that A does not have a proper covering. It follows that we must have $\rho' = \min\{m, n\}$. We permute the lines of A so that the matrix has a 1 in the (1,1) position. We delete row 1 and column 1 of the permuted matrix and denote the resulting matrix of size $m-1$ by $n-1$ by A'. The matrix A' cannot have a covering composed of fewer than $\rho' - 1 = \min\{m-1, n-1\}$ lines because such a covering of A' plus the two deleted lines would yield a proper covering for A. We now apply the induction hypothesis to A' and this allows us to conclude that A' has $\rho' - 1$ 1's with no two of the 1's on a line. But then A has ρ' 1's with no two of the 1's on a line and it follows that $\rho \geq \rho'$.

In the alternative case we assume that A has a proper covering composed of e rows and f columns where $\rho' = e + f$. We permute lines of A so that these e rows and f columns occupy the initial positions. Then our permuted matrix assumes the following form

$$\begin{bmatrix} * & A_1 \\ A_2 & O \end{bmatrix}.$$

In this decomposition O is the zero matrix of size $m - e$ by $n - f$. The matrix A_1 has e rows and cannot be covered by fewer than e lines and the matrix A_2 has f columns and cannot be covered by fewer than f lines. This is the case because otherwise we contradict the fact that $\rho' = e + f$ is the minimal number of lines in A that cover all of the 1's on A. We may apply the induction hypothesis to both A_1 and A_2 and this allows us to conclude that $\rho \geq \rho'$. □

The maximal number of 1's in the (0,1)-matrix A with no two of the 1's on a line is called the *term rank* of A. We denote this basic invariant of A by

$$\rho = \rho(A).$$

We next investigate some important applications of the König theorem. Let $X_1, X_2, \ldots, X_m$ be m not necessarily distinct subsets of an n-set X. Let

$$D = (a_1, a_2, \ldots, a_m)$$

be an ordered sequence of m distinct elements of X and suppose that

$$a_i \in X_i, (i = 1, 2, \ldots, m).$$

Then the element a_i *represents* the set X_i, and we say that our configuration of subsets has a *system of distinct representatives* (abbreviated SDR). We call D an SDR for the ordered sequence of subsets $(X_1, X_2, \ldots, X_m)$. The definition of SDR requires $a_i \neq a_j$ whenever $i \neq j$, but X_i and X_j need not be distinct as subsets of X.

The following theorem of P. Hall[1935] gives necessary and sufficient conditions for the existence of an SDR. We derive the Hall theorem from the König theorem. We remark that one may also reverse the procedure and derive the König theorem from the Hall theorem (Ryser[1963]).

Theorem 1.2.2. *The subsets $X_1, X_2, \ldots, X_m$ of an n-set X have an SDR if and only if the set union $X_{i_1} \cup X_{i_2} \cup \cdots \cup X_{i_k}$ contains at least k elements for $k = 1, 2, \ldots, m$ and for all k-subsets $\{i_1, i_2, \ldots, i_k\}$ of the integers $1, 2, \ldots, m$.*

Proof. The necessity of the condition is clear because if a set union $X_{i_1} \cup X_{i_2} \cup \cdots \cup X_{i_k}$ contains fewer than k elements then it is not possible to select an SDR for these subsets.

We now prove the reverse implication. Let A be the (0,1)-matrix of size m by n which is the incidence matrix for our configuration of subsets. Suppose that A does not have the maximal possible term rank m. Then by the König theorem we may cover the 1's in A with e rows and f columns, where $e+f<m$. We permute the lines of A so that these e rows and f columns occupy the initial positions. Then our permuted A assumes the form

$$\begin{bmatrix} * & A_1 \\ A_2 & O \end{bmatrix}.$$

In this decomposition O is the zero matrix of size $m-e$ by $n-f$. The matrix A_2 of size $m-e$ by f has $m-e>f$. But then the last $m-e$ rows of the displayed matrix correspond to subsets of X whose union contains fewer than $m-e$ elements, and this is contrary to our hypothesis. Hence the matrix A is of term rank m, and this in turn implies that our configuration of subsets has an SDR. □

Let $A=[a_{ij}]$ be a matrix of size m by n with elements in a field F and suppose that $m \le n$. Then the *permanent* of A is defined by

$$\text{per}(A) = \sum a_{1i_1} a_{2i_2} \cdots a_{mi_m},$$

where the summation extends over all the m-permutations $(i_1, i_2, \ldots, i_m)$ of the integers $1, 2, \ldots, n$. Thus per(A) is the sum of all possible products of m elements of A with the property that the elements in each of the products lie on different lines of A. This scalar valued function of the matrix A occurs throughout the combinatorial literature in connection with various enumeration and extremal problems. We remark that per(A) remains invariant under arbitrary permutations of the lines of A. Furthermore, in the case of square matrices per(A) is the same as the determinant function apart from a factor ± 1 preceding each of the products in the summation. In the case of square matrices certain determinantal laws have direct analogues for permanents. In particular, the Laplace expansion for determinants has a simple counterpart for permanents. But the basic multiplicative law of determinants

$$\det(AB) = \det(A)\det(B)$$

is flagrantly false for permanents. Similarly, the permanent function is in general greatly altered by the addition of a multiple of one row of a matrix to another. These facts tend to severely restrict the computational procedures available for the evaluation of permanents.

We return to the (0,1)-matrix A of size m by n, and we now assume that $m \le n$. Then it follows directly from the definition of per(A) that per$(A)>0$ if and only if A is of term rank m. The following theorem is also a direct consequence of the terminology involved.

Theorem 1.2.3. *Let A be the incidence matrix for m subsets $X_1, X_2, \ldots, X_m$ of an n-set X and suppose that $m \le n$. Then the number of distinct SDR's for this configuration of subsets is* per(A). □

The permanent function is studied more thoroughly in Chapter 7.

We have characterized a configuration of subsets by means of a (0,1)-matrix. The choice of the integers 0 and 1 is particularly judicious in many situations, and this is already exemplified by Theorem 1.2.3. But the configuration could also be characterized by a $(1,-1)$-matrix or for that matter by a more general matrix whose individual entries possess or fail to possess a certain property. For example, the following assertion is entirely equivalent to our formulation of the König theorem. *Let A be a matrix of size m by n with elements from a field F. The minimal number of lines in A that cover all of the nonzero elements of A is equal to the maximal number of nonzero elements in A with no two of the nonzero elements on a line.* In what follows we apply the König theorem to nonnegative real matrices.

A matrix of order n is called *doubly stochastic* provided that its entries are nonnegative real numbers and all of its line sums are equal to 1. The $n!$ permutation matrices of order n as well as the matrix of order n with every entry equal to $1/n$ are simple instances of doubly stochastic matrices. The following theorem on doubly stochastic matrices is due to Birkhoff[1946].

Theorem 1.2.4. *A nonnegative real matrix A of order n is doubly stochastic if and only if there exist permutation matrices $P_1, P_2, \ldots, P_t$ and positive real numbers $c_1, c_2, \ldots, c_t$ such that*

$$A = c_1P_1 + c_2P_2 + \cdots + c_tP_t \tag{1.2}$$

and

$$c_1 + c_2 + \cdots + c_t = 1. \tag{1.3}$$

Proof. If the nonnegative matrix A satisfies (1.2) and (1.3) then

$$AJ = JA = J$$

and A is doubly stochastic.

We now prove the reverse implication. We assert that the doubly stochastic matrix A has n positive entries with no two of the positive entries on a line. For if this were not the case, then by the König theorem we could cover all of the positive entries in A with e rows and f columns, where $e+f<n$. But then since all of the line sums of A are equal to 1, it follows that $n \le e+f < n$, and this is a contradiction. Now let P_1 be the permutation matrix of order n with 1's in the same positions as those occupied by the n positive entries of A. Let c_1 be the smallest of these n positive entries. Then $A - c_1P_1$ is a scalar multiple of a doubly stochastic matrix, and at least one more 0 appears in $A - c_1P_1$ than in A. Hence we may iterate the argument on $A - c_1P_1$ and eventually obtain the desired decomposition (1.2). We now multiply (1.2) by J and this immediately implies (1.3). □

Corollary 1.2.5. *Let A be a $(0,1)$-matrix of order n such that all of the line sums of A are equal to the positive integer k. Then there exist permutation matrices $P_1, P_2, \ldots, P_k$ such that*

$$A = P_1 + P_2 + \cdots + P_k.$$

Proof. The (0,1)-matrix A is a scalar multiple of a doubly stochastic matrix. This means that the same arguments used in the proof of Theorem 1.2.4 may be applied directly to the matrix A. But we now have each $c_i = 1$ and the entire process comes to an automatic termination in k steps. □

Corollary 1.2.5 has the following amusing interpretation. A dance is attended by n boys and n girls. Each boy has been previously introduced to exactly k girls and each girl has been previously introduced to exactly k boys. No further introductions are allowed. Is it possible to pair the boys and the girls so that the boy and girl of each pair have been previously introduced? We number the boys $1, 2, \ldots, n$ in an arbitrary manner and similarly for the girls. Then we let $A = [a_{ij}]$ denote the (0,1)-matrix of order n defined by $a_{ij} = 1$ provided boy j has been previously introduced to girl i and by $a_{ij} = 0$ in the alternative situation. Then A satisfies all of the requirements of Corollary 1.2.5, and each of the k permutation matrices P_i gives us a desired pairing of boys and girls. The totality of all of the permitted pairings of boys and girls is equal to $\operatorname{per}(A)$. But it should be noted that $\operatorname{per}(A)$ depends not only on n and k, but also on detailed information involving the structure of the previous introductions.

Exercises

1. Derive Theorem 1.2.1 from Theorem 1.2.2.
2. Suppose in Theorem 1.2.2 the set union $X_{i_1} \cup X_{i_2} \cup \cdots \cup X_{i_k}$ always contains at least $k+1$ elements. Let x be any element of X_1. Show that the sets $X_1, X_2, \ldots, X_m$ have an SDR with the property that x represents X_1.
3. Let A be a (0,1)-matrix of order n satisfying the equation $A + A^T = J - I$. Prove that the term rank of A is at least $n - 1$.
4. Let A be an m by n (0,1)-matrix. Suppose that there exist a positive integer p such that each row of A contains at least p 1's and each column of A contains at most p 1's. Prove that $\operatorname{per}(A) > 0$.
5. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_m$ be two partitions of the n-set X into m subsets. Prove that there exists a permutation $j_1, j_2, \ldots, j_m$ of $\{1, 2, \ldots, m\}$ such that

$$X_i \cap Y_{j_i} \neq \emptyset, \quad (i = 1, 2, \ldots, m)$$