ON UNIMODULAR MATRICES

I. HELLER AND A. J. HOFFMAN

1. Introduction and summary. For the purpose of this note a matrix is called unimodular if every minor determinant equals 0, 1 or -1.

I. Heller and C. B. Tompkins [1] have considered a set

$$S = \{u_i, v_j, u_i + v_j, u_i - u_{i*}, v_j - v_{j*}\}$$

where the $u_1, u_2, \cdots, u_m, v_1, v_2, \cdots, v_n$ are linearly independent vectors in $m + n = k$-dimensional space E, and have shown that in the coordinate representation of S with respect to an arbitrary basis in E every nonvanishing determinant of k vectors of S has the same absolute value, and that, with respect to a basis in S, the vectors of S or of any subset of S are the columns of a unimodular matrix. For the purpose of this note the class of unimodular matrices obtained in this fashion shall be denoted as the class T.

A. J. Hoffman and J. B. Kruskal [4] have considered incidence matrices A of vertices versus directed paths of an oriented graph G, and proved that:

(i) if G is alternating, then A is unimodular;

(ii) if the matrix A of *all* directed paths of G is unimodular, then G is alternating. The terms are defined as follows. A graph G is oriented if it has no circular edges, at most one edge between any given two vertices, and each edge is oriented. A path is a sequence of distinct vertices $v_1, v_2, \cdots, v_k$ of G such that, for each i from 1 to $k-1$, G contains an edge connecting v_i with v_{i+1}; if the orientation of these edges is from v_i to v_{i+1}, the path is directed; if the orientation alternates throughout the sequence, the path is alternating. A loop is a sequence of vertices $v_1, v_2, \cdots, v_k$, which is a path except that $v_k = v_1$. A loop is alternating if successive edges are oppositely oriented and the first and last edges are oppositely oriented. The graph is alternating if every loop is alternating. The incidence matrix $A = (a_{ij})$ of the vertices v_i of G versus a set of directed paths $p_1, p_2, \cdots, p_k$ of G is defined by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is in } p_j \\ 0 & \text{otherwise .} \end{cases}$$

The class of unimodular matrices thus associated with alternating graphs shall be denoted by K.

I. Heller [2] and [3] has considered unimodular matrices obtained

Received April 12, 1962. Research sponsored (for both authors) by the RAND Corporation, in conjunction with the Symposium in Combinatorial Analysis, and (for the first author) by the National Science Foundation.

by representing the edges (interpreted as vectors) of an n-simplex in terms of a basis chosen among the edges (in graph theoretical terms: the edges and vertices of the simplex form a complete graph G; a basis is a maximal tree in G, that is, a tree containing all vertices of G), and has shown that:

(i) the matrix representing all edges of the simplex is unimodular and maximal (i.e., will not remain unimodular when a new column is adjoined);

(ii) the columns of every unimodular matrix of n rows and $n(n+1)$ columns represent the edges of an n-simplex.

The class of (unimodular) matrices whose columns are among the edges of a simplex shall be denoted by H. H can also be defined as a class of incidence matrices: A matrix A belongs to H if there is some oriented graph F without loops such that A is the incidence matrix of the edges of F versus a set of path in F. That is,

$$a_{ij} = \begin{cases} 1 & \text{if edge } e_i \text{ is in path } p_j \\ -1 & \text{if } -e_i \text{ is in } p_j \\ 0 & \text{otherwise} . \end{cases}$$

In [2] it has further been shown that:

(iii) there exist unimodular matrices which do not belong to H;

(iv) the classes H and T are identical.

The purpose of the present note is to show that the class K is identical with the set of nonnegative matrices of H.

2. THEOREM. *If a matrix A of n rows and m columns belongs to K (i.e., A is the incidence matrix of the n vertices of some alternating graph G versus a set of m directed paths in G), then A belongs to H (i.e, there is some n-simplex S and a basis B among its edges such that the columns of A represent edges of S in terms of B). Conversely, every non-negative matrix of H belongs to K.*

3. NOTATION. An oriented graph is viewed as a set

$$R = V \cup E , \tag{3.1}$$

where V is the set of vertices $A_1, A_2, \cdots, A_n$, and E is the set of oriented edges e_ν, that is certain ordered pairs (A_i, A_j) with $j \neq i$ of elements of V, such that at most one of the two pairs (A_i, A_j), (A_j, A_i) is in E.

For brevity of notation we define

$$[A_i, A_j] = \{(A_i, A_j), (A_j, A_i)\} . \tag{3.2}$$

The origin and endpoint of an edge e are denoted by ρe and σe:

$$\rho(A, B) = A , \quad \sigma(A, B) = B , \tag{3.3}$$

If A and B are vertices of R, the relation $A \prec B$ (A is immediate predecessor of B), also written as $B \succ A$, is defined by

$$A \prec B \Longleftrightarrow (A, B) \in R . \tag{3.4}$$

Similarly, if a,b are edges of R,

$$a \prec b \Longleftrightarrow \sigma a = \rho b . \tag{3.5}$$

A subset V' of vertices of R defines a subgraph of R

$$R(V') = V' \cup E' \tag{3.6}$$

where $(A, B) \in E' \Longleftrightarrow A \in V', B \in V', (A, B) \in E$.

4. *Proof*. Using the graph-theoretical definition of the class H, the first half of the theorem shall be proved by showing that to each alternating graph G there is an oriented loopless graph F such that the K-matrices associated with G are among the H-matrices associated with F.

A column of a K-matrix is the incidence column K_p of the vertices of G versus a directed path p in G; a column of an H-matrix is the incidence column H_q of the edges of F versus a path q in F. For given G it will therefore be sufficient to show the existence of an F such that

(4.1) to each directed path p in G there is a path $q = \varphi(p)$ in F such that $K_p = H_q$.

This will be shown by constructing an F and a mapping φ of the set of vertices of G onto the set of edges of F in such a way that φ satisfies (4.1), or equivalently, that φ preserves the relation defined in (3.4) and (3.5), that is, for any two distinct vertices A, B of G,

$$A \prec B \text{ (in } G) \Longrightarrow \varphi(A) \prec \varphi(B) \text{ (in } F) . \tag{4.2}$$

The construction of F and φ shall now be carried out under the assumption that G is connected. If G is not connected, the same construction can be applied to each component of G, yielding an F with an equal number of components.

If G has n vertices, take as the vertices of F a set of $n + 1$ distinct elements $P_0, P_1, \cdots, P_n$.

The n edges $e_1, e_2, \cdots, e_n$ of F are defined successively as follows.

First, choose an arbitrary vertex A_1 in G, define

$$\varphi(A_1) = e_1 = (P_0, P_1) , \tag{4.3}$$

and note that:

(i) the subgraph $G_1 = G(A_1)$, consisting of the one vertex A_1 of G, is, trivially, connected;

(ii) the graph $F_1 = \{P_0, P_1, (P_0, P_1)\}$ is connected;

(iii) with respect to G_1 and F_1, φ trivially satisfies (4.2).

Then, assuming $A_\nu \in G$ already chosen and $e_\nu = \varphi(A_\nu)$ defined for $\nu = 1, 2, \cdots, k$ in such a manner that $G_k = G\{A_1, A_2, \cdots, A_k\}$ and $F_k = \{P_0, P_1, \cdots, P_k, e_1, \cdots, e_k\}$ are each connected and φ satisfies (4.2) with respect to G_k and F_k, choose $A_{k+1} \in G$ such that

$$[A_i, A_{k+1}] \cap G \neq 0 \tag{4.4}$$

for some $i \leqq k$ and define

$$\varphi(A_{k+1}) = e_{k+1} = \begin{cases} (\sigma e_i, P_{k+1}) \text{ when } (A_i, A_{k+1}) \in G \\ (P_{k+1}, \rho e_i) \text{ when } (A_{k+1}, A_i) \in G , \end{cases} \tag{4.5}$$

noting that this definition depends on the choice of i since more than one i may satisfy (4.4).

Obviously, G_{k+1} and F_{k+1} are each connected.

To show that φ satisfies (4.2) with respect to G_{k+1} and F_{k+1}, let $A_r \prec A_s$ in G_{k+1}.

If $r \leqq k$ and $s \leqq h$, (4.2) is satisfied according to the induction's hypothesis.

For $\{r, s\} = \{i, k+1\}$, (4.2) is satisfied by definition (4.5). Namely: for $r = i$, $s = k+1$, (4.5) defines $e_{k+1} = (\sigma e_i, P_{k+1})$, hence $\sigma e_i = \rho e_{k+1}$, which by (3.5) means $e_i \prec e_{k+1}$; similarly for $s = i$, $r = k+1$, (4.5) defines $e_{k+1} = (P_{k+1}, \rho e_i)$, hence $\sigma e_{k+1} = \rho e_i$, which means $e_{k+1} \prec e_i$.

There remains the case $\{r, s\} = \{j, k+1\}, j \neq i, j \leqq k$, with

$$[A_j, A_{k+1}] \cap G_{k+1} \neq 0 , \tag{4.6}$$

that is either $A_j \prec A_{k+1}$ or $A_{k+1} \prec A_j$ in G_{k+1}.

In this case A_{k+1}, which by (4.4) has an edge in common with A_i, now also has an edge in common with $A_j \neq A_i$, thus connecting these two distinct vertices of G_k by the path

$$A_i, A_{k+1}, A_j \tag{4.7}$$

in G_{k+1} but outside G_k.

On the other hand, by the induction's hypothesis, G_k is connected. Hence A_i and A_j are connected by a path in G_k

$$A_i, A_{t_1}, A_{t_2}, \cdots, A_{t_\lambda}, A_j \tag{4.8}$$

($\lambda = 0$ not a priori excluded).

The paths (4.7) and (4.8) combine to the loop

$$A_{k+1}, A_i, A_{t_1}, A_{t_2}, \cdots, A_{t_\lambda}, A_j, A_{k+1} \tag{4.9}$$

in G_{k+1}, which is obviously also a loop in G.

Since G is alternating, the loop (4.9) must be alternating. This implies that the number of vertices is even, hence $\lambda = 2\nu + 1$, and that the orientation is either

$$(4.10) \qquad A_{k+1} \prec A_i \succ A_{t_1} \prec A_{t_2} \succ \cdots \prec A_{t_{2\nu}} \succ A_{t_{2\nu+1}} \prec A_j \succ A_{k+1}$$

or the opposite.

Now assume first

$$(4.11) \qquad A_{k+1} \prec A_j ,$$

which implies the orientation (4.10), and consider that part of the loop which is in G_k, namely the path (4.8)

(4.10) and the induction's hypothesis that, relative to G_k and F_k, φ satisfies (4.2), imply

$$(4.12) \qquad e_i \succ e_{t_1} \prec e_{t_2} \succ \cdots \prec e_{t_{2\nu}} \succ e_{t_{2\nu+1}} \prec e_j ,$$

hence

$$(4.13) \qquad \rho e_i = \sigma e_{t_1} = \rho e_{t_2} = \sigma e_{t_3} = \cdots = \rho e_{t_{2\nu}} = \sigma e_{t_{2\nu+1}} = \rho e_j .$$

The definition (4.5) of e_{k+1}, in conjunction with $A_{k+1} \prec A_i$ from (4.10), implies

$$(4.14) \qquad \sigma e_{k+1} = \rho e_i .$$

This together with (4.13) yields

$$(4.15) \qquad \sigma e_{k+1} = \rho e_j , \quad \text{that is } e_{k+1} \prec e_j ,$$

which proves that assumption (4.11) implies (4.15).

Similarly, the assumption $A_{k+1} \succ A_j$ yields $e_{k+1} \succ e_j$, by reversing the relation $\prec$ and interchanging ρ and σ in the above argument.

This completes the proof that to any connected alternating graph G there exists a connected oriented graph F and a mapping φ satisfying (4.2)

That F has no loops (and hence is a tree) is obvious from the fact that its $n + 1$ vertices are connected by n edges. Hence, the incidence matrices of F certainly belong to class H.

If G consists of k components, the construction will yield an F consisting of k trees.

This completes the proof of the theorem's first half, namely that every K-matrix is an H-matrix.

The second half of the theorem, namely that each nonnegative H-matrix is a K-matrix, is due to J. Edmonds. It will be proved by showing that to each loopless oriented F there is an alternating G and a mapping ψ of the edges of F onto the vertices of G that preserves the relation $\prec$, that is, for any two edges a, b of F

$$a \prec b \Longrightarrow \psi(a) \prec \psi(b) . \tag{4.16}$$

This is achieved by the following simple construction.

If F has n edges $e_1, e_2 \cdots, e_n$, choose a set of n elements $A_1, A_2, \cdots, A_n$ as the vertices of G, define ψ by

$$\psi e_i = A_i , \tag{4.17}$$

and define the edges of G by

$$(A_i, A_j) \in G \Longleftrightarrow e_i \prec e_j , \tag{4.18}$$

that is, G shall have an edge oriented from A_i to A_j if and only if $\sigma e_i = \rho e_j$.

Obviously ψ preserves the relation $\prec$, since (4.18) is equivalent to

$$A_i \prec A_j \Longleftrightarrow e_i \prec e_j . \tag{4.19}$$

Note that $\prec$ is also preserved by the inverse of ψ, that is, in the transition from G to F.

Note further that G is oriented (in the sense of the definition given in [4] and cited in § 1 of present note), that is:

(a) each edge of G is oriented, since the edges of G have been defined by (4.18) as oriented edges;

(b) G has no circular edge, since $(A_i, A_i) \in G$ for some i would imply $e_i \prec e_i$, or equivalently $\sigma e_i = \rho e_i$, that is, e_i a circular edge in F, contradicting the assumption on F;

(c) G has at most one edge between any given two vertices: $(A_i, A_j) \in G$ and $(A_j, A_i) \in G$ for some pair i, j, would imply $e_i \prec e_j$ and $e_j \prec e_i$, that is $\sigma e_i = \rho e_j$ and $\sigma e_j = \rho e_i$, hence e_i and e_j would form a 2-loop (with the vertices $\rho e_i, \sigma e_i$), again contradicting the assumption on F.

Finally, to show that G is alternating, note that, by (4.17) and (4.19), G, F and $\varphi = \psi^{-1}$ satisfy the condition (4.1). Thus the incidence matrices (of vertices versus directed paths) associated with G are among the incidence matrices (edges versus paths) associated with F, and hence unimodular. Especially then, the incidence matrix of the vertices versus *all* the directed paths of G is unimodular, which, by the Hoffman-Kruskal Theorem (Theorem 4 in [4], cited in § 1 of this note), implies that G is necessarily alternating.

This completes proof of the theorem.

It is worth noting that the last part of the proof (namely that G is alternating) can easily be established without using the result of [4] (which contains more than is needed here).

References

1. I. Heller and C. Tompkins, *An extension of a theorem of Dantzig's*, Ann. Math. Study No. 38.
2. I. Heller, *On linear systems with integralvalued solutions*, Pacific. J. Math., **7** (1957), 111-111.
3. ———, *Constraint Matrices of Transportation-Type Problems*, Naval Res. Logistics Quarterly, Vol. 4, No. 1.
4. A. J. Hoffman and J. B. Kruskal, *Integral Boundary Points of Convex Polyhedra*, Ann. Math. Study No. 38.

STANFORD UNIVERSITY
IBM RESEARCH CENTER