Discrete Applied Mathematics 15 (1986) 221-239
North-Holland

STRONG UNIMODULARITY FOR MATRICES AND HYPERGRAPHS

Yves CRAMA
RUTCOR - Rutgers Center for Operations Research, Rutgers University, New Brunswick, NJ 08903, USA

Peter L. HAMMER
RUTCOR - Rutgers Center for Operations Research, Rutgers University, New Brunswick, NJ 08903, USA

Toshihide IBARAKI
Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto, Japan

Received 30 April 1986

A 0-1 matrix A is called strongly unimodular if all the bases of (A, I) are triangular. We develop equivalent conditions for strong unimodularity, first in algebraic, then in graph theoretic terms. This provides a link with the theory of unimodular and balanced hypergraphs, and allows us to produce a polynomial-time recognition algorithm for strongly unimodular matrices.

We consider next the constraint matrix of the problem obtained by linearizing a general, unconstrained optimization problem in 0-1 variables. Because that matrix has 0, 1 and -1 entries, we are led to introduce the concept of signed hypergraph in which every edge is affected of a positive or negative sign. Our results on strong unimodularity are extended to the class of signed hypergraphs.

1. Introduction

Let A be an $(m \times n)$ matrix of 0's and 1's, without zero rows or columns, and let b and c be two integer vectors with m and n components respectively. We associate with A, b and c the following integer linear programming problem (ILP):

$$\begin{aligned}&\max cx,\\ &\text{s.t. } x \in P(A,b) = \{x \in \mathbb{R}^n : Ax \leq b,\ x \geq 0\},\\ &x \text{ integer.}\end{aligned} \tag{1}$$

A *basis* of (A, I_m), where I_m denotes the $(m \times m)$ identity matrix, is any $(m \times m)$ nonsingular submatrix of (A, I_m).

The aim of this paper is to investigate conditions on A which ensure that all the bases of (A, I_m) are triangular (we do not distinguish here between a triangular matrix and one which can be put in triangular form after suitable permutations of its rows and columns). When this is the case, we call A *strongly unimodular*.

Clearly, all the vertices of $P(A,b)$ are then integer and the integrality constraints in (1) become superfluous: the ILP can be solved (at least in principle) as a continuous linear programming problem.

The node-edge incidence matrix of a bipartite graph provides a typical example of strongly unimodular matrix. More generally, Hoffman points out in [8] the significance of strongly unimodular matrices in a unifying theory of greedy algorithms.

In Sections 2–6, we develop equivalent conditions for strong unimodularity, first in algebraic, then in graph theoretic terms. This provides a link with the theory of unimodular and balanced hypergraphs introduced by Berge [2, 3]. We also produce a polynomial-time algorithm for the recognition of strongly unimodular matrices.

In Section 7, we consider the ILP obtained by 'linearizing' a 0–1 *polynomial programming problem*. Because the constraint matrix of the resulting problem has 0, 1 and -1 entries, we are led to introduce the concept of *signed hypergraph*, in which every edge is affected of a positive or negative sign. Our results on strong unimodularity are then extended to the class of signed hypergraphs. We obtain a characterization theorem for strongly unimodular problems, similar to that obtained by Hansen and Simeone for unimodular quadratic problems [6].

2. Algebraic conditions

For the sake of clarity, let us repeat here the following definition. A matrix A of 0's and 1's, without zero rows or columns, is called *strongly unimodular* (*SU*) if all the bases of (A, I_m) are triangular.

As discussed informally above, this definition is motivated by the following observation:

Proposition 1. *Every SU matrix is totally unimodular.*

Proof. If A is SU, the extreme points of $P(A,b)$ are integer for every integer right-hand side b. Therefore, by a well-known result of Hoffman and Kruskal [9], A is totally unimodular. □

We now state our first characterization theorem:

Theorem 1. *The following conditions are equivalent:*

(a) *A is SU.*

(b) *Every square nonsingular submatrix of (A, I_m) is triangular.*

(c) *Every square nonsingular submatrix of A is triangular.*

(d) *For every subset J of columns of A, the columns of A_J are linearly dependent, where A_J is the submatrix of A formed by the columns in J, from which all rows containing at most one non zero entry in J have been deleted.*

Proof. (a)$\Rightarrow$(b). Consider C, a square nonsingular submatrix of (A, I_m). Using appropriate columns from I_m, we can construct the basis

$$B = \begin{vmatrix} C & 0 \\ D & I \end{vmatrix} \tag{2}$$

and, by assumption, B can be put in triangular form. Therefore, it is easy to see that C can be put in triangular form too.

(b)$\Rightarrow$(c). Trivial.

(c)$\Rightarrow$(d). Let J and A_J be as in condition (d). If the columns of A_J are linearly independent, then A_J has a $(|J| \times |J|)$ nonsingular submatrix, say B. By condition (c), B is triangular. But this is impossible, since B has at least two nonzero entries in each row.

(d)$\Rightarrow$(a). Consider a basis B of (A, I_m). B can be written as in (2) above, where C is nonsingular. Because its columns are linearly independent, C has at least one row containing exactly one nonzero entry, say $C(i,j) = 1$. Permuting the rows and columns of C, we can bring $C(i,j)$ in position $C(1,1)$, so that C has the form:

$$C = \begin{vmatrix} 1 & 0 \\ C' & C'' \end{vmatrix}$$

where C'' is nonsingular. The same argument can be used repeatedly to show that C is triangular. Therefore, B is triangular too. □

Corollary. *A matrix is SU if and only if its transpose is SU.*

Proof. Follows immediately from condition (c) in Theorem 1. □

Remark. Theorem 1 remains true with the same definition of strong unimodularity when A is not restricted to have 0–1 entries. We will use that observation in Section 7 below.

3. Strongly unimodular hypergraphs

Every $(m \times n)$ 0–1 matrix can be considered as the *incidence matrix* of a *hypergraph* $H = (X, E)$, where:

$X = \{x_1, \ldots, x_n\}$ is the set of vertices of H;
$E = \{e_1, \ldots, e_m\}$ is the set of edges of H;
$x_j \in e_i$ if and only if $A(i,j) = 1$ $(j = 1, \ldots, n;\ i = 1, \ldots, m)$.

In the sequel, we denote generically by x_j and e_i the vertices and the edges of H, without assuming any longer that x_j corresponds precisely to the j-th column of A, or that e_i corresponds to the i-th row of A.

Our terminology for hypergraphs is that of Berge [2, 3]. In particular, we call *sub-*

hypergraph of H induced by $S \subset X$ the hypergraph $H_S = (S, E_S)$ defined by

$$E_S = \{e_i \cap S : e_i \cap S \neq \emptyset,\ i = 1, \ldots, m\}.$$

A hypergraph is *unimodular* if its incidence matrix is totally unimodular [3].

We define now a hypergraph to be *strongly unimodular* (*SU*) if its incidence matrix is SU. Most properties of SU matrices can be meaningfully translated in terms of the associated hypergraphs. We record here some of these facts for further reference.

Proposition 2. *Every SU hypergraph is unimodular.*

Proof. This is just a translation of Proposition 1. □

If S is a subset of vertices, e is an edge contained in S and f is a real-valued map defined on S, we denote by $f(e)$ the quantity $\sum_{x \in e} f(x)$. With that notation, we have:

Proposition 3. *A hypergraph $H = (X, E)$ is SU if and only if, for every induced sub-hypergraph $H_S = (S, E_S)$ of H, there exists a map $w : S \to \mathbb{R}$, not identically zero on S, such that $w(e) = 0$ for all $e \in E_S$ with $|e| \geq 2$.*

Proof. This is another expression of condition (d) in Theorem 1. □

Proposition 4. *A hypergraph $H = (X, E)$ is SU if and only if all its induced sub-hypergraphs are SU.*

Proof. This follows at once from Proposition 3. □

Define now the dual of $H = (X, E)$ as the hypergraph H^* with vertices $e_1^*, \ldots, e_m^*$ and edges $x_1^*, \ldots, x_n^*$, where

$$e_i^* \in x_j^* \quad \text{if and only if} \quad x_j \in e_i.$$

Then we can state:

Proposition 5. *A hypergraph is SU if and only if its dual is SU.*

Proof. Just observe that the incidence matrix of H^* is the transpose of the incidence matrix of H. □

4. Strongly balanced hypergraphs

We need some more definitions. By a *chain of length p* in the hypergraph

$H=(X,E)$, we mean a sequence $(x_1, e_1, x_2, e_2, \ldots, e_p, x_{p+1})$ such that

(1) the x_j's are pairwise distinct vertices of H, for $j \le p$;

(2) the e_j's are pairwise distinct edges of H, for $j \le p$;

(3) $x_j, x_{j+1} \in e_j$, for $j \le p$.

If $p>1$ and $x_1 = x_{p+1}$, the chain is called a *cycle*. A cycle is *odd* if p is odd.

Berge [2, 3] calls a hypergraph H *balanced* if every odd cycle $(x_1, e_1, x_2, \ldots, e_p, x_1)$ of H uses an edge e_j $(j \le p)$ which contains at least three vertices of the cycle. Anstee [1] calls a cycle C *special* if all its edges contain exactly two vertices of C. With this terminology, a balanced hypergraph can also be defined as being a hypergraph without odd special cycles.

We define now a hypergraph H to be *strongly balanced* if every odd cycle of H uses two edges which contain at least three vertices of the cycle. As in the case of balanced hypergraphs, an alternative definition of strong balanced hypergraphs can be given in terms of forbidden cycles. Define a cycle C to be *sparse* if *at most* one of its edges contains exactly three vertices of C, while the other edges contain exactly two vertices of C. Note that every special cycle is sparse. Now:

Proposition 6. *A hypergraph is strongly balanced if and only if it has no odd sparse cycles.*

Proof. The 'only if' statement is obvious.

For the 'if' part, assume that H is a hypergraph without odd sparse cycles, and let C be an odd cycle of H. Because C is not sparse, it uses two edges which contain three of its vertices, or one edge which contains (at least) four of its vertices. We have to show that, in the latter case, C has another edge which contains at least three vertices of C.

By contradiction, assume this is not true. More specifically, assume:

$$C=(x_1, e_1, x_2, \ldots, e_p, x_1), \quad p \text{ odd};$$

$$|e_1 \cap \{x_1, \ldots, x_p\}| \ge 4; \tag{3}$$

$$|e_j \cap \{x_1, \ldots, x_p\}| = 2 \quad \text{for } 2 \le j \le p. \tag{4}$$

Also, assume that C is the shortest odd cycle of H with the property that all its edges except one contain exactly two vertices of it (i.e. the shortest cycle satisfying conditions similar to (3)–(4), for some suitable labelling of the edges).

We already know that x_1 and x_2 are in e_1. Let x_i and x_j be two other vertices of C which also belong to e_1, and assume $i<j$. Consider the following three cases:

(i) If $j-i$ is even, let

$$C^* = (x_i, e_i, x_{i+1}, \ldots, e_{j-1}, x_j, e_1, x_i).$$

(ii) If $j-i$ is odd and $i-2$ is even, let

$$C^* = (x_2, e_2, x_3, \ldots, e_{i-1}, x_i, e_1, x_2).$$

(iii) If $j-i$ is odd and $i-2$ is odd, let

$$C^* = (x_j, e_j, x_{j+1}, \ldots, e_p, x_1, e_1, x_j).$$

In all cases, C^* is an odd cycle, shorter than C. Moreover, all its edges, except maybe for e_1, contain exactly two vertices of the cycle. Hence, by minimality of C, C^* does not satisfy condition (3), which means that C^* is sparse: contradiction. □

As a straightforward consequence of Proposition 6, we get:

Proposition 7. *A hypergraph is strongly balanced if and only if all its induced sub-hypergraphs are strongly balanced.*

Proof. Trivial. □

We are now ready to state one of the main results of this paper, namely the equivalence of the two classes of hypergraphs we have introduced above:

Theorem 2. *A hypergraph is SU if and only if it is strongly balanced.*

This theorem is certainly a surprising result, in view of the very different definitions we gave for the two types of hypergraphs. Its proof relies heavily on the following lemma, which plays a central role in the characterization of strongly balanced hypergraphs.

Lemma 1. *Let $H=(X,E)$ be a strongly balanced hypergraph. Then, there exists a (nonempty) subset S of X such that the induced hypergraph $H_S=(S,E_S)$ is a (connected) bipartite graph, and each edge e of H with $|e|\geq 2$ contains either zero or two vertices of S.*

The proof of this lemma is rather long and technical, and we defer it to Section 8. But we are now in a position to prove Theorem 2.

Proof of Theorem 2. (a) For the 'only if' part of the proof, we assume that the hypergraph H is not strongly balanced and show (using the characterization in Proposition 3) that H is not SU.

If H is not strongly balanced, it must have a sparse odd cycle C. Call S the set of vertices in C.

If C is a special cycle, then the sub-hypergraph induced by S is precisely the odd cycle C. In that case, it is clear that one cannot find weights $w(x_i)$ as in Proposition 3.

If exactly one edge in C contains three vertices of C, the conclusion is easily seen to be the same.

(b) For the 'if' part, assume that $H=(X,E)$ is a strongly balanced hypergraph.

We want to show that there exists $w: X \to \mathbb{R}$, not identically zero, such that:

$$w(e) = 0 \quad \text{for all } e \in E \text{ with } |e| \geq 2. \tag{5}$$

By Proposition 7, the same conclusion will hold for each induced sub-hypergraph of H, and by Proposition 3, H is SU.

Let S be a subset of vertices defined as in Lemma 1. $H_S = (S, E_S)$ being a bipartite graph, there exists a partition of S into S_1 and S_2 such that S_1 and S_2 are stable in H_S.

Define $w: X \to \mathbb{R}$ by

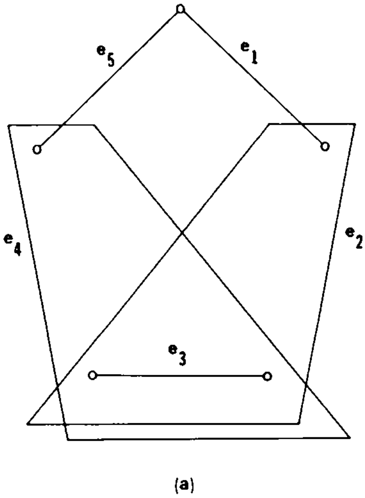
$$w(x) = \begin{cases} 1 & \text{if } x \in S_1, \\ -1 & \text{if } x \in S_2, \\ 0 & \text{otherwise.} \end{cases}$$

We show now that w satisfies condition (5). Let e be an edge of H with $|e| \geq 2$. Trivially, if $e \cap S = \emptyset$, then (5) is satisfied. Else, by choice of S, e contains two vertices of S, say x and y. Because H_S is an induced sub-hypergraph of H, $\{x, y\}$ is an edge of H_S, and therefore: $w(e) = w(x) + w(y) = 0$. $\square$

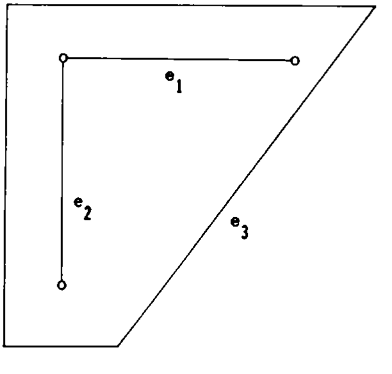
Since the concepts of strong unimodularity and strong balance coincide, we will only use from now on the denomination SU for the hypergraphs we are dealing with. At this point, we have the following implications:

$$H \text{ is without odd cycles} \Rightarrow H \text{ is SU} \Rightarrow H \text{ is unimodular} \Rightarrow H \text{ is balanced.}$$

None of these implications can be reversed. Fig. 1(a) displays an example of a hypergraph which has an odd cycle, but is SU, and Fig. 1(b) an example of a unimodular hypergraph which is not SU. Berge [3] gives an example of a balanced hypergraph which is not unimodular.



(a)



(b)

Fig. 1.

5. A polynomial-time recognition algorithm

We show in this section that there exists a polynomial-time recognition algorithm for SU hypergraphs. To that effect, we first need to prove two theorems which further elucidate the relationship between SU hypergraphs, on the one hand, and balanced or unimodular hypergraphs, on the other hand.

Let us introduce the following notation: if $H=(X,E)$ is a hypergraph, and $x_j \in e_i$ ($j=1,\dots,n$; $i=1,\dots,m$), we denote by $H|(i,j)$ the hypergraph with vertex-set X and with edge-set $\{e_1,\dots,e_i \setminus x_j,\dots,e_m\}$. Similarly, if A is the incidence matrix of H, then $A|(i,j)$ denotes the incidence matrix of $H|(i,j)$, i.e., $A|(i,j)$ is the matrix obtained after replacing the (i,j)-entry of A by 0.

Theorem 3. *H is SU if and only if H and $H|(i,j)$ are balanced, for all i,j such that $x_j \in e_i$ ($j=1,\dots,n$; $i=1,\dots,m$).*

Proof. (Only if) If H is SU, we already know it is balanced. Moreover, any odd cycle of H uses at least two edges containing three vertices of the cycle. Therefore, after deleting a vertex from one edge, every odd cycle still uses at least one edge containing three vertices of the cycle, and the resulting hypergraph is balanced.

(If) If H is not SU, it contains an odd special cycle or an odd sparse cycle. In the first case, H is not balanced. In the second case, let $C=(x_1,e_1,x_2,\dots,e_p,x_1)$ be the odd sparse cycle, and assume without loss of generality that e_1 contains x_1,x_2 and x_j ($3\le j\le p$). Then, C is an odd special cycle in $H|(1,j)$, which therefore is not balanced. □

This result does not provide a 'good' characterization of SU hypergraphs, since there is no known polynomial algorithm to recognize balanced hypergraphs. But it is useful in proving the next theorem:

Theorem 4. *H is SU if and only if H and $H|(i,j)$ are unimodular, for all i,j such that $x_j \in e_i$ ($j=1,\dots,n$; $i=1,\dots,m$).*

Proof. (Only if) Assume H is SU, and let A denote its incidence matrix. We know H is unimodular. Consider i,j such that $x_j \in e_i$, and consider any square submatrix B of A containing the (i,j)-entry of A, a_{ij}. In order to prove that $A|(i,j)$ is totally unimodular, we have to show that the determinant of $B|(i,j)$ is -1, 0 or 1.

If B is nonsingular, it is triangular by definition of SU. Hence, $B|(i,j)$ is triangular too, and $\det(B|(i,j)) \in \{-1,0,1\}$.

So, we can assume that B is singular. If we expand its determinant along the row of B containing the (i,j)-entry of A, we get:

$$0=\det(B)=\det(B|(i,j))+\text{cofactor}(a_{ij}).$$

But cofactor(a_{ij}) is -1, 0 or 1, since A is SU. Therefore, $\det(B|(i,j))$ is -1, 0 or 1 too.

(If) This follows directly from Theorem 3, since every unimodular hypergraph is balanced. □

Now, it is a well-known consequence of Seymour's decomposition theorem for regular matroids [11] that totally unimodular matrices (and hence, unimodular hypergraphs) can be recognized in polynomial time. Therefore, we can derive easily from Theorem 4 the announced result:

Theorem 5. *There exists a polynomial-time recognition algorithm for SU hypergraphs.*

Proof. To determine if H is SU, it is enough by Theorem 4 to check that each of the hypergraphs H and $H|(i,j)$ is unimodular, for all i,j such that $x_j \in e_i$ ($j=1,\dots,n$; $i=1,\dots,m$). Since there are at most $L+1$ hypergraphs to test, where L is the number of nonzero entries in the incidence matrix of H, this task can clearly be performed in polynomial time. □

Remark. Although Theorem 5 states a perfectly satisfactory result from the point of view of complexity theory, the procedure described in its proof is computationally expensive. Conforti and Rao describe in [4] an efficient polynomial-time recognition algorithm, based on a decomposition of SU matrices into matrices without odd cycles.

6. Permanent equitable bicolorings

By definition, a *bicoloring* of a hypergraph $H=(X,E)$ is a map $\lambda : X \to \{-1,1\}$. A bicoloring is *equitable* if, for each edge e of H, $\lambda(e) \in \{-1,0,1\}$.

It is known that a hypergraph is unimodular if and only if all its sub-hypergraphs have an equitable bicoloring (see Berge [3]). We derive here a similar characterization for SU hypergraphs.

Let us call an equitable bicoloring λ *permanent* if, after having switched the 'color' of one vertex x from $\lambda(x)$ to $-\lambda(x)$, we can obtain another equitable bicoloring by switching in the same way the color of at most two vertices in each edge. More formally, the equitable bicoloring λ is permanent if, for each vertex x, there exists an equitable bicoloring μ such that

$$\mu(x) = -\lambda(x), \tag{6}$$

$$\text{for all } e \in E, \quad |\{y \in e : \mu(y) = -\lambda(y)\}| \leq 2. \tag{7}$$

Before stating the characterization theorem, we prove another lemma.

Lemma 2. *If $H=(X,E)$ is SU, there exists a partition $(S_1, S_2, \dots, S_k)$ of X such that* *the induced hypergraphs $H_i=(S_i,E_i)$ are bipartite graphs $(1\le i\le k)$, and, for each edge e in E,*

$$|e\cap S_i|=1$$

holds for at most one i $(1\le i\le k)$.

Proof. Let $X_1=X$, and let $S_1=S$, where S is chosen as in Lemma 1. Then, consider the sub-hypergraph H' induced by $X_2=X_1\setminus S_1$. Because H' is SU, Lemma 1 applies to it and yields a subset S_2 of X_2. Let $X_3=X_2\setminus S_2$, and repeat the procedure inductively. This can be done until we are left with a subset X_k which has at most one common vertex with each edge in E. Then, let $S_k=X_k$, and stop. The partition obtained is easily checked to satisfy the desired conditions. □

Theorem 6. *A hypergraph is SU if and only if all its induced sub-hypergraphs have a permanent equitable bicoloring.*

Proof. (Only if) It is enough to prove that every SU hypergraph has a permanent equitable bicoloring. Let H be SU, and let $(S_1,S_2,\dots,S_k)$ be the partition described in Lemma 2. Also, let H_i denote the bipartite graph induced by S_i, and choose an equitable bicoloring λ_i of H_i $(1\le i\le k)$. Now, the bicoloring of H defined by

$$\lambda(x)=\lambda_i(x) \quad \text{if } x\in S_i$$

is easily seen to be equitable.

In order to show that λ is permanent, we consider an arbitrary vertex x of H, and we exhibit a bicoloring μ which satisfies (6) and (7). Assume that $x\in S_i$; then, μ can be chosen as follows:

$$\begin{aligned}\mu(y)&=\lambda(y) && \text{if } y\notin S_i,\\ \mu(y)&=-\lambda(y) && \text{if } y\in S_i.\end{aligned}$$

(If) We assume that H is not SU, and we show that H has an induced sub-hypergraph without permanent equitable bicolorings.

Let C be a sparse odd cycle of H, and let S be the set of vertices in C. If C is a special cycle, then H_S is a non-bipartite graph, and has no equitable bicolorings.

Assume now that H_S has exactly one edge e containing three vertices, and that λ is an equitable bicoloring of H_S. Then, it is easy to see that, for every vertex x in S, the only equitable bicoloring μ satisfying $\mu(x)=-\lambda(x)$ is defined by

$$\mu(y)=-\lambda(y) \quad \text{for all } y\in S$$

(in other words, if H_S has an equitable bicoloring, then it has exactly two of them: λ and $-\lambda$).

But μ does not satisfy condition (7), since the color of three vertices has been switched in e. Hence, H_S has no permanent equitable bicolorings. □

7. Application to pseudo-Boolean optimization

A *pseudo-Boolean function* is a real-valued function of 0–1 variables. It is well-known that every pseudo-Boolean function has a (unique) representation as a polynomial in its variables [5], and therefore, the most general formulation of an unconstrained optimization problem in 0–1 variables can be stated as:

$$\max f(x) = \sum_{i \in X} a_i x_i + \sum_{k \in P} b_k T_k + \sum_{k \in N} c_k T_k, \tag{8}$$

$$\text{s.t. } x_i \in \{0,1\} \quad \text{for all } i \in X = \{1,2,\dots,n\} \tag{9}$$

where

$$b_k > 0 \quad \text{for all } k \in P,$$

$$c_k < 0 \quad \text{for all } k \in N,$$

$$T_k = \prod_{i \in S(k)} x_i \quad \text{and} \quad S(k) \subset X \quad \text{for all } k \in P \cup N.$$

Introducing new 0–1 variables y_k associated with the nonlinear terms T_k $(k \in P \cup N)$ and constraining them to take the value of the product of the variables in these terms, we arrive at the following linear 0–1 program:

$$\max z(x,y) = \sum_{i \in X} a_i x_i + \sum_{k \in P} b_k y_k + \sum_{k \in N} c_k y_k, \tag{10}$$

$$\text{s.t. } y_k - x_i \le 0 \quad \text{for all } i \in S(k),\ k \in P, \tag{11}$$

$$-y_k + \sum_{i \in S(k)} x_i \le |S(k)| - 1 \quad \text{for all } k \in N, \tag{12}$$

$$x_i \in \{0,1\} \quad \text{for all } i \in X, \tag{13}$$

$$y_k \in \{0,1\} \quad \text{for all } k \in P \cup N. \tag{14}$$

This is the *discrete linear form* of the maximization problem (8)–(9). Let $A(f)$ denote the matrix associated with the constraints (11)–(12). When N is empty, Rhys [10] has observed that $A(f)$ is totally unimodular. Therefore, in that case, (13) and (14) can be replaced by their continuous relaxations:

$$0 \le x_i \le 1 \quad \text{for all } i \in X, \tag{15}$$

$$0 \le y_k \le 1 \quad \text{for all } k \in P \cup N. \tag{16}$$

When N is not empty, the linear programming problem (10)–(12) and (15)–(16) represents a relaxation of (10)–(14), to be called the *continuous linear form* of the problem.

Since $A(f)$ has 0, 1 and -1 entries, it does not directly fit into the framework of Sections 2–6 above. Our aim in this section is to extend our previous results on strong unimodularity, so that they apply to $A(f)$ as well.

Again, we say that $A(f)$ is SU if all the bases of $(A(f), I_m)$ are triangular, where

m is the number of rows of $A(f)$, and we call f itself SU when $A(f)$ is SU. We already know that Theorem 1 remains valid for $A(f)$ (see the remark following Theorem 1).

Now, given a pseudo-Boolean function f in the form (8), consider the *signed hypergraph* $H(f)=(X,E^+,E^-)$ defined by:

$$X=\{1,2,\dots,n\},$$
$$E^+=\{S(k):k\in P\},$$
$$E^-=\{S(k):k\in N\}.$$

We call the edges in E^+ *positive* and the edges in E^- *negative.*

Clearly, the class of signed hypergraphs is in one-to-one correspondence with the class of constraints matrices $A(f)$ arising from pseudo-Boolean optimization problems as described before. A signed hypergraph is called SU if the corresponding matrix $A(f)$ is SU. From now on, we will concentrate on the characterization of SU signed hypergraphs, rather than on the corresponding matrices.

Proposition 8. *A signed hypergraph $H=(X,E^+,E^-)$ is SU if and only if, for every induced sub-hypergraph $H_S=(S,E_S^+,E_S^-)$ of H, there exists a map $w:S\to\mathbb{R}$, not identically zero on S, such that:*

$$\text{(i)}\quad w(e)=0 \quad \textit{for all } e\in E_S^- \textit{ with } |e|\geq 2, \tag{17}$$

$$\text{(ii)}\quad w(i)=w(j) \quad \textit{for all } i,j\in e, \textit{ for all } e\in E_S^+. \tag{18}$$

Proof. (Only if) Let S be a subset of vertices, and consider the matrix $A=A(f)$ associated with H as above. A has its columns indexed by X (like the x_i's, $i\in X$) and by $P\cup N$ (like the y_k's, $k\in P\cup N$). Consider the set of columns J corresponding to $S\cup N$, and let A_J be defined as in Theorem 1.

Now, by Theorem 1, the columns of A_J are linearly dependent. Therefore, there exists an $(|S|+|P|)$-dimensional vector v, different from zero, such that $A_Jv=0$. It is easy to check that the map $w:S\to\mathbb{R}$ defined by $w(i)=v_i$ $(i\in S)$ is not identically zero on S and satisfies (17)–(18).

(If) Given a set J of columns of A, indexed by $S\cup P'\cup N'$, consider $w:S\to\mathbb{R}$, not identically zero on S, and satisfying (17)–(18). Then, define an $(|S|+|P'|+|N'|)$-dimensional vector v by:

$$\begin{array}{ll} v_i=w(i) & \text{for all } i\in S,\\ v_k=w(i) & \text{for all } k\in P' \text{ such that } |S(k)\cap S|\geq 1, \text{ and for any } i\in S(k)\cap S,\\ v_k=w(i) & \text{for all } k\in N' \text{ such that } |S(k)\cap S|=1,\ S(k)\cap S=\{i\},\\ v_k=0 & \text{else} \end{array}$$

(notice that this definition is unambiguous since w satisfies (18)).

It can be checked that $A_Jv=0$. So, by Theorem 1, A is SU. □

This result takes an especially attractive form when the function f in (8) happens to be *quadratic*, i.e. when each term in the polynomial representation of f involves at most two variables. In that case, the signed hypergraph $H(f)$ associated with f turns out to be a signed graph. Recall that a signed graph is called *balanced* if it has no negative cycles, that is no cycles involving an odd number of negative edges. We can state the following corollary to Proposition 8:

Corollary 1. *A signed graph is SU if and only if it is balanced.*

Proof. It follows from a theorem of Harary [7] that a signed graph $G=(X,E^+,E^-)$ is balanced if and only if there exists $w:X\to\{-1,1\}$ such that

(i) $w(e)=0$ for all $e\in E^-$,

(ii) $w(x)=w(y)$ for all edges $e=\{x,y\}\in E^+$.

By Proposition 8, this is equivalent to saying that G is SU. □

In [6], Hansen and Simeone introduce a family L of linear relaxations of the problem (8)-(9), which does not contain our continuous linear form (10)-(12), (15)-(16). They call a pseudo-Boolean function *unimodular* if at least one member of L has a totally unimodular constraint matrix, and *completely unimodular* if all members of L have a totally unimodular matrix. They show that, for a quadratic function f, these two classes coincide: f is unimodular (and completely unimodular) if and only if $H(f)$ is balanced. Combining their result with Corollary 1 above, we get:

Corollary 2. *A quadratic pseudo-Boolean function is unimodular if and only if it is SU.*

Proof. Follows from the previous discussion. □

We give now a characterization of SU signed hypergraphs in terms of forbidden cycles. A *sparse negative* cycle in a signed hypergraph is a cycle $C=(x_1,e_1,x_2,\ldots,e_p,x_1)$ using an odd number of *negative* edges, and such that at most one of its *negative* edges contains three vertices of C, while *all other* edges contain exactly two vertices of C.

It should be noticed that, in a hypergraph having only negative edges, the sparse negative cycles are exactly the sparse odd cycles defined in Section 4. Also, it is clear that positive edges play essentially no role in the definition of a sparse negative cycle. This remark will now be expressed in a more formal way.

Given a signed hypergraph $H=(X,E^+,E^-)$, define an equivalence relation R on its vertex set by: xRy if there is a chain from x to y in H using only positive edges. Then, pick a vertex x_j ($j=1,2,\ldots,k$) in each equivalence class of R, and define $\psi:X\to X$ by

$$\psi(x)=x_j \quad \text{if } xRx_j.$$

Proposition 9. *If $H=(X,E^+,E^-)$ has no sparse negative cycles, then $|e|=|\psi(e)|$ for all $e\in E^-$.*

Proof. Assume $e\in E^-$ and $|e|>|\psi(e)|$. Then, there exist $x,y\in e$ such that $\psi(x)=\psi(y)$, i.e. there exists a chain $C=(x_1,e_1,x_2,\ldots,e_p,x_{p+1})$ using only positive edges, such that $x=x_1$ and $y=x_{p+1}$. Assume that C is the shortest among such chains. Therefore, no e_i $(1\le i\le p)$ contains three vertices in C. But this implies that $(x_1,e_1,x_2,\ldots,e_p,x_{p+1},e,x_1)$ is a sparse negative cycle in H. □

Consider now the hypergraph $H^c=(X^c,E^c)$ with vertex set $X^c=\psi(X)$ and edge set $E^c=\{\psi(e):e\in E^-\}$. Intuitively, H^c is the hypergraph obtained by 'contracting' the positive edges of H. We look at H^c as having only negative edges.

Proposition 10. *If H has no sparse negative cycles, then H^c has no sparse odd cycles.*

Proof. Assume by contradiction that $C^c=(x_1',e_1',x_2',\ldots,e_p',x_1')$ is a shortest sparse odd cycle in H^c. Now, consider a sequence C in H defined as follows:

$$C=((y_1,X_1,z_1),e_1,(y_2,X_2,z_2),e_2,\ldots,(y_p,X_p,z_p),e_p,y_1),$$

where

(i) X_i is a shortest chain of positive edges from y_i to z_i $(1\le i\le p)$;
(ii) $y_i\in e_{i-1}$, $z_i\in e_i$ $(1\le i\le p)$;
(iii) $\psi(y_i)=\psi(z_i)=x_i'$ $(1\le i\le p)$;
(iv) $\psi(e_i)=e_i'$ $(1\le i\le p)$.

Intuitively, C is a sequence whose 'image' by ψ is C^c. Note that C is not necessarily a cycle of H, since it can have repeated vertices or edges.

Nevertheless, assume now that C is the shortest among all sequences satisfying (i)–(iv). Then:

(a) If C has a repeated vertex, one sub-sequence of C between two occurences of that vertex generates in H^c a sparse odd cycle not longer than C^c. So, we may assume there is no repeated vertex in C;

(b) By the same reasoning, we can exclude the case where a positive edge of C contains three vertices in C, or where a positive edge is repeated in C;

By (a) and (b), we can assume that C is a cycle and, because H has no sparse negative cycles, C must have two negative edges containing three vertices of C, say e_i and e_j. But then, e_i' and e_j' contain three vertices of C^c. Therefore, C^c is not sparse: contradiction. □

Now, we are ready to prove our last characterization theorem:

Theorem 7. *A signed hypergraph is SU if and only if it has no sparse negative cycles.*

Proof. (Only if) The proof is essentially the same as for the 'only if' part of Theorem 2, and we omit it here.

(If) Assume H has no sparse negative cycles. Then, all the induced sub-hypergraphs of H are without sparse negative cycles too. Therefore, and using Proposition 8, we only have to show that there exists a map $w: X \to \mathbb{R}$, not identically zero, satisfying conditions (17)-(18).

By Proposition 10, $H^c = (X^c, E^c)$ is SU. So, we can find $\omega: X^c \to \mathbb{R}$, not identically zero on X^c, and such that $\omega(e) = 0$ for all $e \in E^c$ with $|e| \geq 2$. Now, we can define w on X by

$$w(x) = \omega(\psi(x)) \quad \text{for all } x \in X.$$

Using Proposition 9, one checks that w satisfies (17)-(18). □

Remark. The result stated in Corollary 1 is just a special case of Theorem 7.

8. Proof of Lemma 1

Notice that the statement to prove is equivalent to the following one:

Lemma 1′. *If H is strongly balanced hypergraph, there exists a nonempty subset S of vertices such that each edge e of H with $|e| \geq 2$ contains either zero or two vertices of S.*

Indeed, such a subset S obviously induces a graph without odd cycles, since every cycle in a graph in sparse. The fact that H_S is connected follows from the proof.

Proof. Let $H = (X, E)$ be a strongly balanced hypergraph. The proof of Lemma 1′ is in two stages.

(1) We first describe a procedure which yields a certain subset S of vertices; if the procedure halts, then S satisfies the conditions stated in the lemma.

(2) We prove that the procedure halts.

(1) S is obtained as follows (we use freely the notations $A \cup i$ and $A \setminus i$ to denote $A \cup \{i\}$ and $A \setminus \{i\}$ respectively).

Step 1. Pick an arbitrary vertex x in X. Let $S \leftarrow \{x\}$ and label x *not examined.*

Step 2. Choose a vertex i labelled *not examined.* If no such vertex exists, return S and stop.

Step 3. Let $e_1, \ldots, e_p$ denote the edges of H such that $|e_k| > 1$ and $e_k \cap S = \{i\}$ $(k = 1, \ldots, p)$. For $k = 1, \ldots, p$, define j_k as follows:

(i) if e_k contains j_r for some $r < k$, let $j_k \leftarrow j_r$;

(ii) else, if $j \in X$ is such that

$$j \in e_k \setminus i, \tag{19}$$

$$|e \cap (S \cup j)| \leq 2 \quad \text{for all } e \in E, \tag{20}$$

let $j_k \leftarrow j$, $S \leftarrow S \cup j$, and, if j is not labelled *examined*, label it *not examined*;

(iii) else, if there is no j satisfying (19)-(20), apply the procedure described in either Case 1 or Case 2 below.

Step 4. Call i *examined* and go to Step 2.

In Step 3, case (iii) occurs if, for all $j \in e_k \setminus i$, there is $e \in E$ such that:

$$|e \cap (S \cup j)| \geq 3. \tag{21}$$

But, because the procedure always maintains $|e \cap S| \leq 2$ for all e in E, (21) can be rewritten as: for all $j \in e_k \setminus i$, there is $e \in E$ and $r, s \in S$ such that

$$e \cap (S \cup j) = \{j, r, s\}. \tag{22}$$

Assume now that (iii) occurred, and choose $j \in e_k \setminus i$. There are two cases to consider.

Case 1. Assume e, r and s are as in (22) but $r = i$ (see Fig. 2). Then, remove all the labels, let $S \leftarrow \{i, j\}$, label i and j *not examined* and go to Step 2.

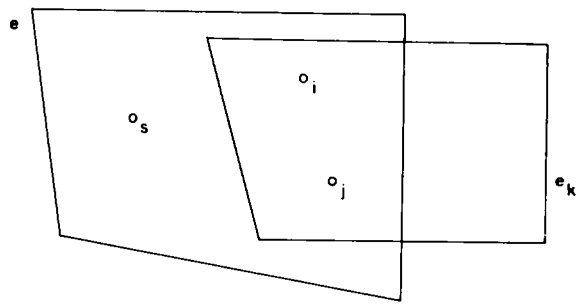


Fig. 2.

Case 2. Assume that, for all e, r and s as in (22), $r \neq i$ and $s \neq i$. Let

$$T = \{(e, r, s) : e \in E,\ r, s \in X,\ e \cap (S \cup j) = \{j, r, s\}\}.$$

Notice that the subgraph induced by S is always connected, by construction. Therefore, we can assume without loss of generality that, for all (e, r, s) in T, there is a chain $P(r)$ from r to i which uses only vertices in S, but does not use s (else, permute the roles of r and s).

Consider now a pair (r, s) such that $(e, r, s) \in T$ for some edge e. We show that, as suggested by Fig. 3, the edge $\{r, s\}$ is an isthmus in H_S (i.e. the deletion of $\{r, s\}$ disconnects H_S). Indeed, suppose by contradiction that there is in H_S a chain Q from s to r which does not use the edge $\{r, s\}$. Then, Q must be odd, since (Q, e, s) is a sparse cycle in H. By a similar argument, $P = (P(r), e_k, j)$ is an odd chain in H. Hence, the symmetric difference $Q \triangle P$ contains an even chain C from s to j, and (C, e, s) is a sparse odd cycle in H: contradiction.

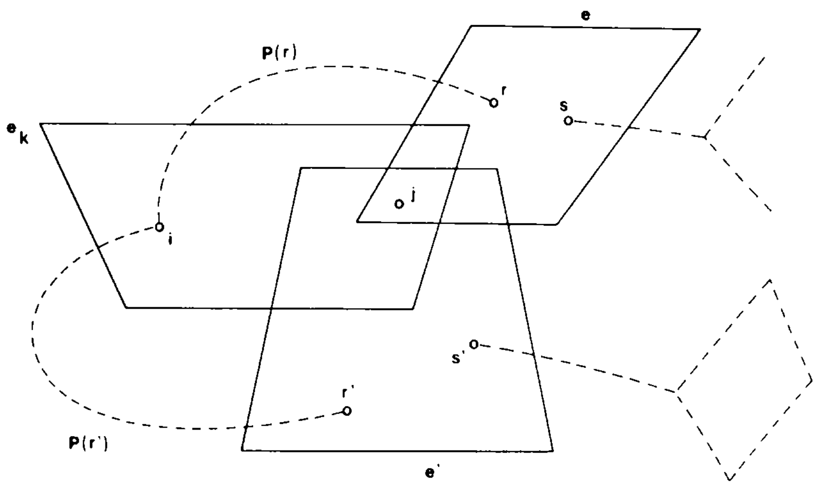


Fig. 3.

So, $\{r,s\}$ is an isthmus in H_S. Call Y the subset of vertices of S which remain connected to i, after deletion of all the edges $\{r,s\}$ such that $(e,r,s)\in T$ for some e. Now, the procedure to apply when Case 2 occurs is as follows: remove all labels, let $S\leftarrow Y\cup j$, label all the vertices in S *not examined*, and go to Step 2.

At this point, it should be clear that, if the procedure described above halts, it does so with a set of vertices S satisfying the conditions in Lemma 1′.

(2) A careful inspection of the procedure reveals that the only step which could possibly cause it to cycle is Step 3(iii). But it should be remarked that, every time Step 3(iii) is performed, one vertex at least is deleted from the current set S (namely, the vertices generically denoted by s are deleted). We prove hereunder that such a deleted vertex will never be put back in S, in any subsequent stage of the procedure. Hence, since the number of vertices is finite, the procedure must halt.

Claim. *Assume vertex v is removed from S during iteration t of the procedure, after an occurrence of either Case 1 or Case 2. Let e_v be any edge of H containing v. Then, no vertices of e_v will ever be added to S, in any subsequent step of the procedure.*

We prove this by contradiction. Consider the first triple (v,e_v,u) causing a contradiction; i.e., v and e_v are as in the claim, $u\in e_v$, and u is added to S at iteration $t+t'$. Denote by Σ the set of all vertices which have been put at least once in S (i.e., which have been labelled at least once *not examined*) during iterations $t,t+1,\dots,t+t'$. Notice that, since u creates the first contradiction, no edges of H contain more than two vertices of Σ, except for the edges which caused an occurrence of Case 1 or Case 2 during iteration $t+t'$.

(a) Assume first there has been an occurrence of Case 1 during one of the iterations $t, t+1, \ldots, t+t'$. Let i, j, s, e_k and e be as in the description of Case 1. We know there are chains $P(v,s)$ from v to s, and $P(j,u)$ from j to u in Σ. Also, these chains can be chosen such that no edge of H contains more than two vertices of $P(v,s)$ and $P(j,u)$. Then, the cycles

$$(P(v,s), e, P(j,u), e_v, v) \text{ and } (P(v,s), e, i, e_k, P(j,u), e_v, v)$$

are sparse (see Fig. 4). But one of these cycles is odd, and this contradicts the assumption that H is strongly balanced.

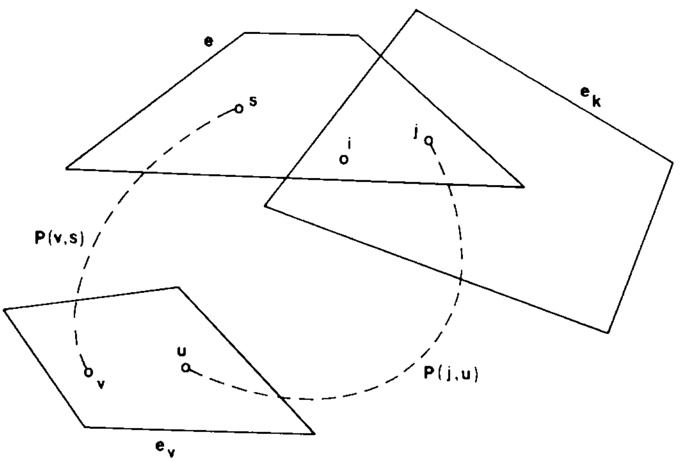


Fig. 4.

(b) Assume now there has been an occurrence of Case 2 during one of the iterations $t, t+1, \ldots, t+t'$. Let i, j, r, s, e, e_k and $P(r)$ be defined as in Case 2. Choose a path $P(v,s)$ and a path $P(j,u)$ as in (a), with $r \notin P(j,u)$ (see Fig. 5). Now, the cycles $(P(v,s), e, P(r), e_k, P(j,u), e_v, v)$ and $(P(v,s), e, P(j,u), e_v, v)$ are sparse, and one of them is odd (since $P(r)$ is even). Again, we reach a contradiction, and this completes the proof. $\square$

Acknowledgements

The results in Section 5 were obtained in collaboration with Svatopluk Poljak. The authors wish to acknowledge his contribution and express their appreciation.

This work was partially supported by the National Science Foundation under grant #DMS-83-05569. The first author was supported by a doctoral fellowship from the Intercollegiate Center for Management Science, Belgium.

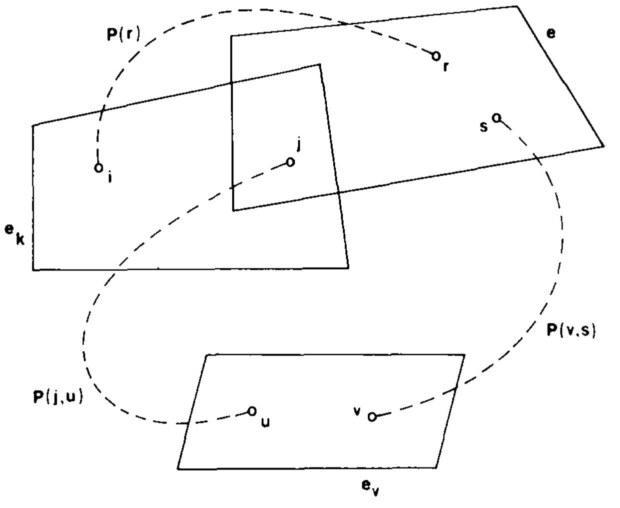


Fig. 5.

References

[1] R.P. Anstee, Hypergraphs with no special cycles, Combinatorica 3 (1983) 141-146.

[2] C. Berge, Balanced matrices, Math. Programming 2 (1972) 19-31.

[3] C. Berge, Graphs and Hypergraphs (North-Holland, Amsterdam, 1973).

[4] M. Conforti and M.R. Rao, Structural properties and recognition of restricted and strongly unimodular matrices, Preprint (New York University, New York, 1985).

[5] P.L. Hammer and S. Rudeanu, Boolean Methods in Operations Research and Related Areas (Springer, Berlin, 1968).

[6] P. Hansen and B. Simeone, Unimodular functions, Discrete Appl. Math. 14 (1986) 269-281.

[7] F. Harary, On the notion of balance of a signed graph, Michigan Math. J. 2 (1953-1954) 143-146.

[8] A.J. Hoffman, On greedy algorithms that succeed, in: I. Anderson, ed., Surveys in Combinatorics 1985, London Math. Soc. Lecture Note Series 103 (Cambridge University Press, Cambridge, 1985) 97-112.

[9] A.J. Hoffman and J.B. Kruskal, Integral boundary points of convex polyhedra, in: H. Kuhn and A. Tucker, eds., Linear Inequalities and Related Systems, Annals of Mathematical Studies 38 (Princeton University Press, Princeton, 1956) 223-246.

[10] J. Rhys, A selection problem of shared fixed costs and network flows, Management Sci. 17 (1970) 200-207.

[11] P.D. Seymour, Decomposition of regular matroids, J. Combin. Theory (Ser. B) 28 (1980) 305-359.