A Short Proof of Tutte's Characterization of Totally Unimodular Matrices

A. M. H. Gerards

Department of Econometrics
Tilburg University
P.O. Box 90153
5000 LE Tilburg, The Netherlands

Dedicated to Alan J. Hoffman on the occasion of his 65th birthday.

Submitted by Robert Bixby

ABSTRACT

We give, in terms of totally unimodular matrices, a short and easy proof of Tutte's characterization of regular matroids.

1. INTRODUCTION

We give a short and easy proof of the following well-known result of Tutte (1958, 1965, 1971):

TUTTE'S THEOREM. Let $A$ be a $\{0,1\}$-matrix. Then the following are equivalent:

(i) $A$ has a totally unimodular signing,

(ii) $A$ cannot be transformed to

$$M(F_\gamma) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

by applying (repeatedly) the following operations:

- deleting rows or columns,
- permuting rows or columns,
- taking the transposed matrix,
- pivoting over $\text{GF}(2)$.

The notions used here are: Signing a \((0,1)\)-matrix \(A\) means replacing some of the 1’s in \(A\) by \(-1\)’s. A matrix is called \textit{totally unimodular} if each of its subdeterminants is 0, 1, or \(-1\). (In particular, all entries of a totally unimodular matrix are 0, 1, or \(-1\).) Pivoting a matrix \(A\) (on an entry \(e = \pm 1\) of \(A\)) means replacing the matrix

\[
A = \begin{pmatrix} \varepsilon & y^T \\ x & D \end{pmatrix} \quad \text{by} \quad B = \begin{pmatrix} -\varepsilon & y^T \\ x & D - \varepsilon xy^T \end{pmatrix}.
\]

We consider pivoting over \(\mathbb{GF}(2)\) as well as over the field \(\mathbb{R}\).

Our proof of Tutte’s theorem is in Section 3. Section 2 contains a few well-known and easy-to-prove preliminary results on graphs, pivoting, and total unimodularity.

Remarks. Tutte formulated his result in terms of regular matroids (= regular chain groups): A binary matroid is regular if and only if it has neither the Fano matroid nor its dual as a minor. It is not hard to establish the equivalence of both formulations [e.g. see Bixby (1982), Schrijver (1986, Sections 21.1, 21.2), Tutte (1958, 1965, 1971), or Welsh (1976)].

Tutte’s original proof is very complicated. Shorter and more transparent proofs are given by Seymour (1979) and Truemper (1978, 1982). In fact they prove a generalization of Tutte’s theorem: Reid’s characterization of \(\mathbb{GF}(3)\)-representable matroids. Lovasz and Schrijver observed that Reid’s theorem can be proved also along the lines of the proof we present in Section 3.

Reid never published his result. The first proofs that appeared in print are due to Seymour (1979) and to Bixby (1979). Bixby’s proof goes along the lines of Tutte’s original proof of his theorem stated above. Other proofs of Reid’s theorem are given by Kahn (1984) and Kahn and Seymour (1986).

An extension of Tutte’s theorem is given by Bixby (1976).

2. PRELIMINARIES

2.1. The Bipartite Graph of a Matrix

The proof of Tutte’s theorem in Section 3 uses the following easy graph-theoretic result.

\textbf{Lemma 1.} Let \(G\) be a connected bipartite graph, with no parallel edges. If deleting any two nodes in the same color class yields a disconnected graph, then \(G\) is a path or a circuit.
Proof. Suppose $G$ is neither a path nor a circuit. Then $G$ has a spanning tree $T$ that is not a path. Hence $T$ has least three end nodes. At least two of them are in the same color class of $G$. Deleting these two nodes from $G$ results in a connected graph.

Let $A$ be a matrix. Denote the index set of the rows (columns) of $A$ by $R$ ($C$ respectively). The bipartite graph, $G(A)$, associated with $A$ has color classes $R$ and $C$. There is an edge from $i \in R$ to $j \in C$ in $G(A)$ if the entry in row $i$ and column $j$ of $A$ is nonzero.

2.2. Total Unimodularity

We recall two well-known facts on totally unimodular matrices.

**Lemma 2.** Let $A$ be an $n \times n$ matrix with entries from $\{0, 1, -1\}$. If $G(A)$ a circuit, then $A$ is totally unimodular if and only if the number of $-1$'s in $A$ is congruent to $n$ modulo 2.

**Proof.** Directly from the fact that $\det(A) \in \{0, 1, -1\}$.

**Lemma 3** (Camion 1963). Let $M_1$ and $M_2$ be two totally unimodular matrices. If $M_1$ and $M_2$ are congruent modulo 2, then $M_1$ can be obtained from $M_2$ by multiplying some rows and columns by $-1$.

**Proof** (Paul Seymour). Call an edge in $G(M_1) \cup G(M_2)$ even if the corresponding entries in $M_1$ and $M_2$ are the same. Call the other edges odd. By Lemma 2, each chordless circuit, and hence each circuit, in $G(M_1)$ has an even number of odd edges. Therefore, the nodes of $G(M_1)$ can be partitioned into two classes, say $V_1$ and $V_2$, such that any edge $e$ is odd if and only if $e$ connects $V_1$ and $V_2$. Now multiply by $-1$ all rows and all columns of $M_1$ corresponding to the nodes in $V_1$. The resulting matrix is $M_2$.

2.3. Pivoting

The following properties of the pivoting operation (2) are easy to prove:

(i) Pivoting $B$ on $-e$ yields $A$.
(ii) If $A$ is square, then $\det(A) = \pm \det(B - ey^T)$.
(iii) If $A$ is totally unimodular, then $B$ is totally unimodular.
(iv) If $G(A)$ is connected, then $G(B)$ is connected [since if $G(B)$ is disconnected, then (i) implies that $G(A)$ is disconnected too].
3. PROOF OF TUTTE’S THEOREM

The existence of a totally unimodular signing is invariant under the operations (1) [by (3) (iii)]. Moreover $M(F_r)$ has no totally unimodular signing. Hence (i) implies (ii). So it remains to prove the reverse implication.

Suppose $A$ is a $\{0,1\}$-matrix, satisfying (ii), with no totally unimodular signing. We may assume that each proper submatrix of $A$ has a totally unimodular signing. So the bipartite graph $G(A)$ is connected. [If not,]

$$A = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}$$

for certain matrices $B$ and $C$ (up to permutation of rows and columns), implying that at least one of $B$ and $C$ has no totally unimodular signing.] $G(A)$ is not a path or circuit (as otherwise $A$ has trivially a totally unimodular signing). Hence, by Lemma 1, $A$ or $A^T$ is equal to $[x|y|N]$ (up to permutation of columns), where $x$ and $y$ are two column vectors and where $G(N)$ is connected. By assumption, both $[x|N]$ and $[y|N]$ have a totally unimodular signing. Moreover, by Lemma 3, these two signings can be chosen such that $N$ is signed in the same way in both cases. Hence $A$ or $A^T$ has a signing $A' = [x'|y'|N']$ satisfying

(i) $G(N')$ is connected.

(ii) Both $[x'|N']$ and $[y'|N]$ are totally unimodular.

**Claim.** *We may assume that the matrix $[x'|y']$ has a submatrix of the form*

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$ 

**Proof.** By (3) (iii) and (3) (iv) of Section 2.3, pivoting $A'$ on an entry in $N'$ does not influence the property (4). Now, pivot $A'$ on an entry in $N'$ such that the smallest submatrix $M$ with determinant not equal to 0, 1, or $-1$ is as small as possible. Then $M$ is a $2 \times 2$ matrix. [If not, pivot on an entry lying both in $M$ and $N'$; cf. (3) (ii).] So $M$ is of the form in the claim (if necessary multiply $x'$, $y'$, or a row by $-1$). Moreover, by (4) $M$ has to be a submatrix of $[x'|y']$.  

Denote the row indices of the two rows of $A'$ in which the submatrix of the claim occurs by $\alpha$ and $\beta$. Since $G(N')$ is connected, there exists a path in $G(N')$ from $\alpha$ to $\beta$. This path cannot have length 2 (as such a path would
correspond with a column of \(N'\) with two \(\pm 1\)'s in the rows \(\alpha\) and \(\beta\), contradicting the fact that both \([x'|N']\) and \([y'|N']\) are totally unimodular. From this it follows that \(A'\) has a submatrix of the form depicted below:

\[
\begin{pmatrix}
\alpha & 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\
\beta & 1 & -1 & 0 & 0 & \cdots & 0 & 1 \\
* & 1 & 1 & \cdots & 0 \\
* & 1 & \cdots & 0 \\
0 & \cdots & 1 \\
1 & 1 & 1 & 0 \\
1 & -1 & 0 & 1 \\
\end{pmatrix}
\]

(If necessary permute rows of \(A'\) and columns of \(N'\), multiply them by \(-1\), or exchange \(x'\) and \(y'\)). By pivoting on the underlined entries, deleting the rows and columns containing these pivoting elements, and multiplying some rows and columns by \(-1\) (and if necessary exchanging \(x'\) and \(y'\)), we get a submatrix of the form

\[
\begin{pmatrix}
1 & 1 & 1 & 0 \\
1 & -1 & 0 & 1 \\
a & b & 1 & 1 \\
\end{pmatrix}
\]

It is still the case that deleting either column \(x'\) or column \(y'\) yields a totally unimodular matrix. This implies that \(a = 1\) and \(b = 0\). Hence \(A\) can be transformed to \(M(F_r)\), contradicting our assumption.

\[\text{\textit{I thank Alexander Schrijver. Our discussions on the first version of the proof presented here stimulated me to find some shortcuts.}}\]

REFERENCES

Camion, P., 1963. Caractérisation des matrices unimodulaires, Cahiers Centre Études
Schrijver, A., 1986. Theory of Linear and Integer Programming, Wiley, Chichester,
1986.
26:159–173.
Truemper, K., 1982. Alpha-balanced graphs and matrices and GF(3)-representability
69:1–47.

Received 30 September 1987; final manuscript accepted 25 May 1988