Chapter 9

Matrix Total Unimodularity and Matroid Regularity

9.1 Overview

At this point, we have assembled the basic matroid results and tools for the remaining developments. We now begin the investigation into the two main technical subjects, which are matrix total unimodularity and matroid regularity.

Total unimodularity and its many variants are very important for combinatorial optimization. For a long time, matrix techniques evidently did not permit any profound insight into total unimodularity. That fact motivated the translation of the matrix property of total unimodularity into the matroid property of regularity. The translation opened the way for the application of powerful matroid techniques. After an effort spanning three decades, matroid regularity and thus total unimodularity were at last understood. The main results, in historical order, are due to Tutte and Seymour. The subsequent chapters contain these results as well as closely related material.

We proceed as follows. In Section 9.2, we define total unimodularity, sketch important applications, and translate total unimodularity into matroid language, thus getting matroid regularity. Section 9.3 contains the first major result for regularity, which is Tutte's characterization of the regular binary matroids. Recall that F_7 is the Fano matroid and F_7^* its dual, and that U_n^m is the uniform matroid on n elements where any subset with at most m elements is independent. The characterization says that a matroid is regular if and only if it has no U_4^2, F_7, or F_7^* minors. The original proof of that characterization is long and complicated. Here we rely on a proof of amazing brevity and clarity by Gerards. A minor modification of the proof produces Reid's characterization of the ternary matroids, i.e., the matroids representable over GF(3). The characterization says that a matroid is ternary if and only if it has no U_5^2, U_5^3, F_7, or F_7^* minors. This result is presented in Section 9.4. The final section, 9.5, contains extensions and references.

The chapter requires knowledge of the material of Chapters 2 and 3.

9.2 Basic Results and Applications of Total Unimodularity

In this section, we define matrix total unimodularity and matroid regularity and establish elementary results about these two properties. We also point out representative applications of total unimodularity.

A real matrix is *totally unimodular* if every square submatrix has 0 or ± 1 as determinant. Thus, the entries of a totally unimodular matrix must be 0 or ± 1. For example, the zero matrices and the identity matrices are totally unimodular. Nontrivial examples are given by the next lemma.

(9.2.1) Lemma. *Let A be a real $\{0, \pm 1\}$ matrix where every column contains exactly one $+1$ and one -1. Then A is totally unimodular.*

Proof. Let D be a square submatrix of A. We induct on the order of D. In the nontrivial case, D has order $k \geq 2$ and has no zero columns. If D has a column with exactly one ± 1, we use cofactor expansion and induction to calculate the determinant as 0 or ± 1. Otherwise, D has exactly one $+1$ and one -1 in each column, and thus the determinant is 0. □

Define F to be the support matrix of the matrix A of Lemma (9.2.1). View F to be binary. Since A has exactly one $+1$ and one -1 in each column, the matrix F has exactly two 1s in each column. Then according to the definitions of Section 3.2, F is the node/edge incidence matrix of some graph G. Specifically, each row of F defines a node of G, and each column g of F, say with 1s in rows i and j, defines an edge connecting nodes i and j of G. We now declare the matrix A to represent a certain directed version of G, as follows. Assume that column g of A has a $+1$ in row i and a -1 in row j. Then we replace the undirected edge of G connecting the nodes i and j by a directed arc from i to j. Let H be the resulting directed graph. We call A the *node/arc incidence matrix* of H. By Lemma (9.2.1), every node/arc incidence matrix is totally unimodular.

The next lemma summarizes elementary operations that maintain total unimodularity.

(9.2.2) Lemma. *Total unimodularity is maintained under the taking of submatrices, transposition, pivots, and the adjoining of zero or unit vectors or of parallel* $\{0, \pm 1\}$ *vectors.*

Proof. The pivot is the only nontrivial operation. Let a pivot convert a totally unimodular matrix A to a matrix A'. Adjoin to A an identity matrix to get $[I \mid A]$. Now A is totally unimodular if and only if every basis matrix of $[I \mid A]$ has ± 1 as determinant. The latter property is maintained under the elementary row operations in $[I \mid A]$ that correspond to the pivot in A, and that, together with scaling by $\{\pm 1\}$ factors and a column exchange, convert $[I \mid A]$ to $[I \mid A']$. Thus, A' is totally unimodular. □

According to the next lemma, counting may be used to check total unimodularity for the $\{0, \pm 1\}$ matrices whose bipartite graph $\mathrm{BG}(\cdot)$ is a chordless cycle, i.e., for the $k \times k$ real matrices, $k \geq 2$, of the form

$$\text{(9.2.3)} \qquad \begin{array}{|ccccc|} \hline \pm 1 & & & & \pm 1 \\ \pm 1 & \pm 1 & & & \\ & \pm 1 & \ddots & & \\ & & \ddots & \pm 1 & \\ & & & \pm 1 & \pm 1 \\ \hline \end{array}$$

Matrix whose bipartite graph
is a chordless cycle

(9.2.4) Lemma. *The real matrix of* (9.2.3) *is totally unimodular if and only if its entries sum to* $0 (\mathrm{mod}\, 4)$.

Proof. Let A be the matrix of (9.2.3). Scaling of a row or column of A by -1 does not affect total unimodularity and changes the sum of the entries of A by a multiple of 4. Thus, for the proof of the lemma, we scale A to the matrix A' given by

$$\text{(9.2.5)} \qquad \begin{array}{|ccccc|} \hline 1 & & & & \alpha \\ -1 & 1 & & & \\ & -1 & \ddots & & \\ & & \ddots & 1 & \\ & & & -1 & 1 \\ \hline \end{array}$$

Particular matrix whose bipartite graph
is a chordless cycle

By cofactor expansion and counting, we confirm that $\det_{\mathbb{R}} A'$ is 2 (resp. 0) if and only if $\alpha = 1$ ($\alpha = -1$), which holds if and only if the entries of A' sum to 2 (resp. 0). The "only if" part of the proof is then immediate. The "if" part follows from Lemma (9.2.1). □

A matrix B is *regular* if it is binary and if its 1s can be replaced by ± 1s so that a real matrix results that is totally unimodular. By the above discussion, every node/edge incidence matrix is regular. The signing of a regular matrix to achieve a real totally unimodular matrix is essentially unique, a fact proved next.

(9.2.6) Lemma. *Let A and A' be two totally unimodular matrices with the same support matrix. Then A' may be obtained from A by a scaling of the rows and columns by $\{\pm 1\}$ factors.*

Proof. We may assume A to be connected. Let T be a tree of $\mathrm{BG}(A)$, and let T' be the corresponding tree of $\mathrm{BG}(A')$. Because of scaling, we may suppose that A and A' agree on the entries corresponding to the edges of T and T'. Suppose A and A' differ, say on the (i, j) element. The corresponding edges e and e' of $\mathrm{BG}(A)$ and $\mathrm{BG}(A')$ form cycles C and C' with T and T', respectively. Select T and e, and thus T' and e', so that the cardinality of the cycles is minimum. Suppose C and C' have chords. By the minimality condition, the entries of A and A' corresponding to such chords must agree. But then C and C' do not have minimum cardinality, a contradiction. Thus, C and C' are chordless cycles of $\mathrm{BG}(A)$ and $\mathrm{BG}(A')$, and the submatrices of A and A' corresponding to C and C' are of the form (9.2.3). The sums of the entries of the two submatrices differ by 2, since they differ on just one entry. But then one of the sums is $0 \pmod 4$, and the other one is $2 \pmod 4$. By Lemma (9.2.4), one of A and A' is not totally unimodular, a contradiction. $\square$

Lemma (9.2.6) and its proof imply the following result.

(9.2.7) Corollary. *There is a polynomial algorithm that by signing converts any regular matrix B to a real matrix A that is totally unimodular. The signing can be carried out as follows. First some submatrix is signed. Then, iteratively, the method signs an arbitrarily selected additional row or column that is adjoined to the submatrix on hand.*

Proof. Use Lemma (9.2.6) and the arguments of its proof to establish the correct signs for the additional row and column. $\square$

A matroid M is *regular* if it has a regular representation matrix B. The essentially unique totally unimodular matrix A deduced from the regular B may be used to represent M as follows.

(9.2.8) Lemma. *Let M be a matroid represented by a real totally unimodular matrix A. Define A' to be a numerically identical copy of A, but view A' to be a matrix over an arbitrary field $\mathcal{F}$. Then M is represented by A' over $\mathcal{F}$.*

Proof. Since all square submatrices of A have real determinant 0 or ± 1, any such submatrix of A must be $\mathbb{R}$-nonsingular if and only if the corresponding submatrix of A' is $\mathcal{F}$-nonsingular. Thus, A over $\mathbb{R}$ and A' over $\mathcal{F}$ define the same matroid. $\square$

Lemma (9.2.8) leads to a short proof of the following theorem.

(9.2.9) Theorem. *The following statements are equivalent for a matroid M.*

(i) *M is regular.*
(ii) *M has a real representation matrix that is totally unimodular.*
(iii) *M is representable over every field.*
(iv) *M is representable over* GF(2) *and* GF(3).
(v) *M representable over* GF(3), *and every representation matrix over* GF(3), *when viewed as real, is totally unimodular.*

Proof. By definition, (i) $\iff$ (ii), and by Lemma (9.2.8), (ii) $\Longrightarrow$ (iii). Trivially, (iii) $\Longrightarrow$ (iv) and (v) $\Longrightarrow$ (ii). We show (iv) $\Longrightarrow$ (v). By (iv), some matrix C over GF(3) and the binary support matrix B of C represent M. Declare A to be a copy of C, but view A to be over the reals. We have shown (v) once we have proved A to be totally unimodular. Suppose A has a submatrix with real determinant different from 0, ± 1. We may assume that A itself is such a matrix, and that every proper submatrix of A is totally unimodular. If A is a 2×2 matrix, then by a trivial case analysis we must have up to scaling by $\{\pm 1\}$ factors in A and C,

$$A = \boxed{\begin{matrix} 1 & 1 \\ 1 & -1 \end{matrix}}\ ; \quad B = \boxed{\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}}\ ; \quad C = \boxed{\begin{matrix} 1 & 1 \\ 1 & -1 \end{matrix}} \tag{9.2.10}$$

A over $\mathbb{R}$, B over GF(2), and C over GF(3)

Then $\det_2 B = 0$ and $\det_3 C = 1$. We conclude that B and C represent different matroids, a contradiction. Suppose the order of A is greater than 2. We produce the 2×2 case by pivots in A, B, and C, as follows. We perform an $\mathbb{R}$-pivot on any nonzero entry of A and delete the pivot row and pivot column. In B and C, we carry out the corresponding operations. Let A', B', C' be so obtained from A, B, and C. It is easy to see that A' is not totally unimodular, and that every proper submatrix of A' is totally unimodular. Furthermore, B' is the binary support matrix of A', and A' and C' have their 0s, +1s, and −1s in the same positions. By induction, the contradictory 2×2 case applies. □

The preceding results and arguments yield the following corollaries.

(9.2.11) Corollary. *For every regular matroid M, the following holds.*

(a) *Every binary representation matrix of M is regular.*
(b) *Every minor of M is regular.*
(c) *The dual of M is regular.*

Proof. Lemma (9.2.2) and Theorem (9.2.9) imply (a), (b), and (c). □

(9.2.12) Corollary. *The graphic matroids as well as the cographic ones are regular.*

Proof. By Corollary (9.2.11), we only need to consider the case of graphicness. So let G be a graph producing a graphic matroid M. Add a new node to G and connect it to all other nodes. Let G' be the resulting graph. The added edges constitute a spanning tree T' of G'. The representation matrix of the graphic matroid M' of G' is nothing but the node/edge incidence matrix of G, with rows indexed by T'. By Lemma (9.2.1), the latter matrix is regular. Thus, M' is regular. By Corollary (9.2.11), the minor $M' \backslash T'$, which is M, is regular as well. □

We know that the transpose of a graphic matrix need not be graphic. Examples are the representation matrices of $M(K_5)$ and $M(K_{3,3})$ given by (3.2.38) and (3.2.41). By 1-sum composition, we thus can create regular matrices B that are not graphic and not cographic. For example, we may choose as one block of B the matrix B' of (3.2.38), and as the second block of B the transpose of B'. There are less trivial ways of creating regular nongraphic and noncographic matrices. We present them in Chapter 11. At this time, we introduce two regular nongraphic and noncographic matrices, called B^{10} and B^{12}, that play a very special role in Chapters 10 and 11. The matrix B^{10} is

(9.2.13)
$$B^{10} = \begin{array}{|ccccc|} \hline 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ \hline \end{array}$$

Matrix B^{10} of regular matroid R_{10}

To prove regularity, we declare B^{10} to be over $\mathbb{R}$ and check that all square submatrices have real determinant equal to 0 or ± 1. Brute-force checking turns out to be quite tedious. But in Chapter 12, we learn much about minimal $\{0, \pm 1\}$ matrices that are not totally unimodular. The conditions presented there almost immediately prove B^{10} to be totally unimodular. The regular matroid represented by B^{10} is called R_{10}. We prove nongraphicness and noncographicness of R_{10} as follows. Delete the last column from B^{10}. Up to indices, the matrix of (3.2.41) results. Hence, R_{10} has an $M(K_{3,3})$ minor. The matrix B^{10} is also its transpose. Thus, R_{10} also has an $M(K_{3,3})^*$ minor. We conclude that R_{10} is not graphic and not cographic.

We have met the matrix B^{12} already in (4.4.9). Below we include that matrix in (9.2.14) in partitioned form, for reasons explained shortly. Following that matrix, we display a signed version of B^{12} in (9.2.15). We claim that the matrix of (9.2.15) is totally unimodular, which implies that B^{12} is regular. Analogously to the B^{10} case, one may verify the claim using the results of Chapter 12.

(9.2.14)
$$B^{12} = \begin{array}{|cccc|cc|}\hline 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \hline \end{array}$$

Matrix B^{12} of regular matroid R_{12}

(9.2.15)
$$\begin{array}{|cccc|cc|}\hline 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & -1 & 0 & -1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \\ \hline \end{array}$$

Totally unimodular version of B^{12}

An easier method for verifying total unimodularity of the matrix of (9.2.15) uses decomposition, as follows. By (8.3.10) and (8.3.11), the partition of the matrix B^{12} of (9.2.14) induces a 3-sum decomposition with component matrices B^1 and B^2 given by (9.2.16) below. A GF(2)-pivot on the 1 in the lower right corner of B^2 converts that matrix, up to indices, to the matrix of (3.2.41). The latter matrix represents up to indices $M(K_{3,3})$, and so does B^2. The matrix B^1 is the transpose of B^2. Thus, the matroid of B^1 is isomorphic to $M(K_{3,3})^*$.

(9.2.16)
$$B^1 = \begin{array}{|cccc|c|}\hline 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ \hline \end{array} \quad ; \quad B^2 = \begin{array}{|cc|cc|}\hline 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \hline \end{array}$$

Component matrices B^1 and B^2 of B^{12}

These facts imply that R_{12} is not graphic and not cographic. They also establish that R_{12} is a 3-sum of a copy of the regular $M(K_{3,3})$ with one of the regular $M(K_{3,3})^*$. In Chapter 11, we see that any 3-sum of two regular matroids is also regular. This fact proves R_{12} to be regular. The simple signing algorithm implicit in the proof of Corollary (9.2.7) confirms that the signed version of B^{12} given by (9.2.15) is indeed totally unimodular.

Total unimodularity is an important property for combinatorial optimization. Numerous problems of that area of mathematics can be expressed as an *integer program* of the form

$$
\begin{aligned}
\text{min} \quad & d^t \cdot x \\
\text{s. t.} \quad & A \cdot x \leq b \\
& 0 \leq x \leq c \\
& x \text{ integer}
\end{aligned}
\tag{9.2.17}
$$

where A is a given integral matrix, b, c, and d are given integral column vectors, and x is a column vector representing the solution. The dimensions of the arrays are such that the indicated multiplications and inequalities make sense. The abbreviation "s. t." stands for "subject to." In general, (9.2.17) is not easy. But if A is a totally unimodular matrix, then one can effectively drop the integrality requirement from (9.2.17) and solve the resulting *linear program.* In this book, we do not dwell on details and implications of this approach, which has been treated extensively elsewhere. In Section 9.5, we provide appropriate references.

A special class of the problems (9.2.17) involves as A the node/arc incidence matrices of directed graphs, or matrices derived from node/arc incidence matrices by pivots and deletion of rows and columns. Any problem of (9.2.17) in that class is called a *network flow problem.* That class has also been treated extensively, and numerous special algorithms and famous inequalities exist. That these algorithms work and that the inequalities are valid can almost always be traced back to the total unimodularity of node/arc incidence matrices. Again, we must resist an even cursory treatment and must point to Section 9.5 for references. However, in Section 10.6 of the next chapter, we do present an efficient algorithm for testing whether or not a given $\{0, \pm 1\}$ matrix A is the matrix of some network flow problem.

The importance of total unimodularity naturally leads us to ask a number of questions. How can we recognize a totally unimodular matrix? Is there a simple construction of the entire class of totally unimodular matrices? What are the characteristics of the matrices that are not totally unimodular, but all of whose proper submatrices are totally unimodular? Are there other well-behaved problem classes of type (9.2.17) where A is not totally unimodular, but where A is related to some totally unimodular matrix?

Theorem (9.2.9) links matrix total unimodularity and matroid regularity. Thus, one may pose the following related matroid questions. Is there a simple construction of the entire class of regular matroids? Which are the nonregular binary matroids all of whose proper minors are regular? Which are the nongraphic binary matroids all of whose proper minors are graphic? The latter two questions make sense since regularity as well as graphicness are maintained under minor-taking. Are there other classes of binary matroids that are not regular, but that are closely related to the regular matroids?

The above questions have complete answers, all of which are provided in the sequel. The answer to the question concerning the minimal nonregular binary matroids is given in the next section.

9.3 Characterization of Regular Matroids

By Theorem (9.2.9), every regular matroid is binary. We already have a characterization of the binary matroids: Theorem (3.5.2) says that a matroid is binary if and only if it has no U_4^2 minors. Thus, a regular matroid has no U_4^2 minors. By Corollary (9.2.11), regularity is maintained under minor-taking and dualizing. For a complete characterization of regularity, we thus must identify the minimal binary matroids M that are not regular. In this section, we prove the famous theorem of Tutte according to which there is just one such M up to isomorphism and dualizing. That matroid is the *Fano matroid* F_7 represented by the matrix B^7 of (9.3.1) below. The name is due to the fact that the matroid is the Fano plane, which is the projective geometry PG$(2,2)$. The seven elements of the matroid are the points of the geometry. We saw the Fano matroid earlier in Sections 3.3 and 4.4, under (3.3.22) and (4.4.13). In the latter case, "*con*" and "*del*" labels were assigned to its elements. We have no such labels here.

(9.3.1)

$$B^7 = \boxed{\begin{array}{cccc} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{array}}$$

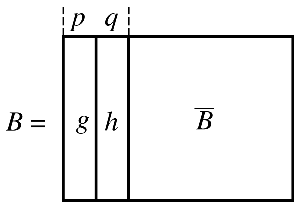
Matrix B^7 of Fano matroid F_7

The matrix B^7 cannot be signed to become totally unimodular, as follows. The last three columns of B^7 give a totally unimodular submatrix. If B^7 is regular, then by Corollary (9.2.7), we can sign the first column of B^7 to achieve a totally unimodular matrix. But for all such ways, a 2×2 or 3×3 matrix with real determinant equal to $+2$ of -2 is created. Thus, B^7 is not regular. Tutte's theorem is as follows. The amazingly simple proof is due to Gerards.

(9.3.2) Theorem. *A binary matroid is regular if and only if it has no F_7 or F_7^* minors.*

Proof. For proof of the nontrivial "if" part, let M be a nonregular binary matroid all of whose proper minors are regular. Evidently, any binary representation matrix B of M must be connected. BG(B) cannot be a single cycle or a path, for the first case corresponds to a wheel matroid, and the second one to a minor of a wheel matroid; both matroids are regular. Thus, BG(B) is connected and has a node of degree 3, and hence has a spanning tree with a degree 3 node. Such a tree has at least three tip nodes, and hence has at least two tip nodes that correspond to two rows or to two columns of B. Thus, deletion of two columns or of two rows, say indexed by p and q, reduces B to a connected matrix $\overline{B}$. Because of dualizing of M, we may assume the former case. Thus,

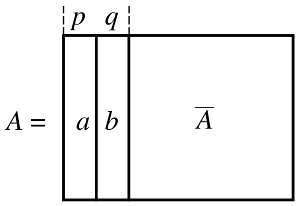
(9.3.3)



Binary Matrix B of minimal nonregular matroid M

where the submatrices $\overline{B}$, $[g \mid \overline{B}]$, and $[h \mid \overline{B}]$ are connected. By Corollary (9.2.7), we may sign B so that a real matrix A of the form

(9.3.4)



Matrix A derived from B of (9.3.3) by signing

results where both $[a \mid \overline{A}]$ and $[b \mid \overline{A}]$ are totally unimodular. Since M is not regular, A is not totally unimodular. By the construction, any submatrix of A proving the latter fact must intersect both columns p and q. Let D be a minimal such submatrix. If D is not a 2×2 matrix, we may convert it to one by real pivots in $\overline{A}$. The corresponding binary pivots in B must lead to a matrix that is the support of the one deduced from A, for otherwise, at least one of the matrices $[a \mid \overline{A}]$, $[b \mid \overline{A}]$ is not totally unimodular. Furthermore, by Lemma (9.2.2), the pivots convert $[a \mid \overline{A}]$ and $[b \mid \overline{A}]$ to some other totally unimodular matrices, and do not affect connectedness of $\overline{A}$. Thus, we may assume the 2×2 D to be already present in A of (9.3.4). That is, the columns a and b of A contain a 2×2 matrix with determinant equal to $+2$ or -2, say in rows v and w.

Up to scaling, A can thus be further partitioned as shown in (9.3.5) below. The submatrix $\overline{A}$ has been subdivided into $\overline{\overline{A}}$, $\overline{c}$, and $\overline{e}$. The vectors a and b have become $\overline{a}$ and $\overline{b}$ plus the explicitly shown ± 1s. The latter entries make up D.

(9.3.5)
$$A = \begin{array}{c|c|c|c|}
 & p & q & \\
\hline
 & \overline{a} & \overline{b} & \overline{\overline{A}} \\
\hline
v & 1 & 1 & \overline{c} \\
\hline
w & 1 & -1 & \overline{e} \\
\hline
\end{array}$$

Partition of matrix A displaying
non-totally unimodular 2×2 submatrix

Examine the row subvectors $\overline{c}$ and $\overline{e}$ in rows v and w. Suppose both subvectors have ± 1 entries in some column $y \neq p, q$. Then those two entries plus the two ± 1s of column p or column q must form a non-totally unimodular 2×2 matrix, a contradiction that both $[a \mid \overline{A}]$ and $[b \mid \overline{A}]$ are totally unimodular. None of the subvectors $\overline{c}$ and $\overline{e}$ can be zero, since otherwise $\overline{A}$ is not connected. By scaling, we thus can assume $\overline{c}$ and $\overline{e}$ to be as indicated below.

(9.3.6)
$$A = \begin{array}{c|c|c|c|c|}
 & p & q & \overline{Y}_1 & \overline{Y}_2 \\
\hline
 & \overline{a} & \overline{b} & \multicolumn{2}{c|}{\overline{\overline{A}}} \\
\hline
v & 1 & 1 & 1 \cdots 1 & 0 \\
\hline
w & 1 & -1 & 0 & 1 \cdots 1 \\
\hline
\end{array}$$

Further partitioning of matrix A

Since $\overline{A}$ is connected, there is a path from some $r \in \overline{Y}_1$ to some $s \in \overline{Y}_2$ in the bipartite graph of the submatrix $\overline{\overline{A}}$. By the path shortening technique of Chapter 5, we may assume the path to have exactly two edges. Thus, we can refine A of (9.3.6) to

(9.3.7)
$$A = \begin{array}{c|c|c|c|c|c|}
 & p & q & r & s & \\
\hline
 & \overline{\overline{a}} & \overline{\overline{b}} & \multicolumn{3}{c|}{\overline{\overline{A}}} \\
\hline
u & \alpha & \beta & \pm 1 & \pm 1 & \\
\hline
v & 1 & 1 & 1 & 0 & \overline{\overline{c}} \\
\hline
w & 1 & -1 & 0 & 1 & \overline{\overline{e}} \\
\hline
\end{array}$$

Final partitioning of matrix A

By scaling in row u, we may presume the entry in row u and column r to be a $+1$. We now concentrate on the submatrix A' of A indexed by $\{u, v, w\}$ and $\{p, q, r, s\}$. We show that submatrix below. The entries α and β of the submatrix are yet to be determined.

(9.3.8)
$$A' = \begin{array}{c|cccc|} & p & q & r & s \\ \hline u & \alpha & \beta & 1 & \pm 1 \\ v & 1 & 1 & 1 & 0 \\ w & 1 & -1 & 0 & 1 \\ \hline \end{array}$$

Submatrix A' of A displaying F_7 minor

If $\alpha = \beta = 0$, then by Lemma (9.2.4), the 3×3 column submatrix of A' indexed by $\{p, r, s\}$ or the one indexed by $\{q, r, s\}$ is not totally unimodular. But then one of $[a \mid \overline{A}]$ and $[b \mid \overline{A}]$ is not totally unimodular, a contradiction. Suppose both α and β are nonzero. Due to column r and total unimodularity of $[a \mid \overline{A}]$ and $[b \mid \overline{A}]$, we have $\alpha = \beta = 1$. But due to column s and rows u and w, the entries α and β must have opposite sign, a contradiction. Thus, exactly one of α and β is zero. Then A' of (9.3.8) has up to index sets the matrix B^7 as support. We conclude that M or M^* is isomorphic to F_7. □

During a first reading of the book, the next section may be skipped. There we characterize the ternary matroids.

9.4 Characterization of Ternary Matroids

The proof of Theorem (9.3.2) of the preceding section can be extended to establish the following characterization by R. Reid of the ternary matroids, i.e., the matroids representable over GF(3). Recall from Section 3.4 that representability over a given field is maintained under minor-taking and dualizing.

(9.4.1) Theorem. *A matroid is representable over* GF(3) *if and only if it has no F_7, F_7^*, U_5^2, or U_5^3 minors.*

Proof. We assume that the reader is quite familiar with the material on abstract representations of Section 3.4. It is easy to check that the excluded minors are not representable over GF(3). Indeed, it suffices to verify this for F_7 and U_5^2, since F_7^* and U_5^3 are the duals of these matroids. Thus, the "if" part holds. For proof of the converse, we argue as in the proof of Theorem (9.3.2), except that B is an abstract matrix representing M. The graph BG(B) cannot be a cycle or path, since otherwise M is a GF(3)-representable wheel or whirl matroid or a minor of such a matroid. Thus, up to dualizing of M, B can be partitioned as in (9.3.3), where $\overline{B}$ is connected.

We need the following auxiliary result: Two matrices over GF(3) and with the same support represent the same matroid if and only if either one of the two matrices may be obtained from the other one by scaling of rows and columns by $\{\pm 1\}$ factors. The proof is identical to that of Lemma (9.2.6), except that the final sentence of the proof is replaced by the observation that a matrix of (9.2.3) has the GF(3)-determinant equal to 0 if and only if its entries sum (in $\mathbb{R}$) to 0(mod 4).

Because of the auxiliary result and its proof, the signing of an abstract matrix to achieve a representation matrix over GF(3) can be done column by column, provided of course the matroid is GF(3)-representable. This observation is the GF(3)-analogue of Corollary (9.2.7). Thus, for the case at hand, we may deduce from B, which is partitioned as in (9.3.3), a matrix A of (9.3.4) over GF(3) where $[a \mid \overline{A}]$ represents $M\backslash q$ and where $[b \mid \overline{A}]$ represents $M\backslash p$. Since M is not representable over GF(3), A does not represent M.

Due to abstract pivots in $\overline{B}$ of B and corresponding GF(3)-pivots in $\overline{A}$ of A, we may assume that B and A satisfy the following two additional conditions. First, A is the matrix of (9.3.5), except that the explicitly shown -1 in row w and column q is a $+1$ or -1. We denote that entry by γ. Second, the GF(3)-determinant of the 2×2 submatrix of rows v, w and columns p, q is not correct for M. That is, the GF(3)-determinant of that submatrix is zero (resp. nonzero), i.e., $\gamma = 1$ (resp. $\gamma = -1$), if and only if the related 2×2 submatrix of B has abstract determinant 1 (resp. 0). Let us include the just-described matrix A for easy reference.

(9.4.2)
$$A = \begin{array}{c|c|c|c|} & p & q & \\ \hline & \overline{a} & \overline{b} & \overline{\overline{A}} \\ \hline v & 1 & 1 & \overline{c} \\ \hline w & 1 & \gamma & \overline{e} \\ \hline \end{array}$$

Partition of matrix A displaying 2×2 submatrix
with incorrect determinant

Due to the γ entry, the ensuing arguments are a bit more subtle than those proving Theorem (9.3.2). Assume that both row subvectors $\overline{c}$ and $\overline{e}$ in rows v and w of A have ± 1s in a column $y \neq p, q$. Extract from A of (9.4.2) the submatrix A' indexed by $\{v, w\}$ and $\{p, q, y\}$, i.e.,

(9.4.3)
$$A' = \begin{array}{c|ccc|} & p & q & y \\ \hline v & 1 & 1 & \pm 1 \\ w & 1 & \gamma & \pm 1 \\ \hline \end{array}$$

Submatrix A' of A displaying U_5^2 minor

That submatrix corresponds to a minor M' of M. According to the previous discussion, p and q are parallel in M' if and only if $\gamma = -1$. Assume

$\gamma = -1$. By (9.4.3), one of $\{p, y\}$, $\{q, y\}$, say $\{p, y\}$, contains two parallel elements of M', and the other one is a base of M'. Since $\gamma = -1$, p and q are parallel in M'. But "is parallel to" is an equivalence relation, so y and q must also be parallel in M', a contradiction. Thus, $\gamma = +1$. Arguing by contradiction as in the previous case, both $\{p, y\}$ and $\{q, y\}$ must be bases of M'. Then M' is isomorphic to U_5^2, and we are done.

We return to A of (9.4.2), knowing now that it can be further partitioned as in (9.3.6), except that the explicitly shown -1 in (9.3.6) must be replaced by γ. By path-shortening GF(3)-pivots in $\overline{\overline{A}}$ and subsequent deletion of rows and columns, we get the following analogue of the small case (9.3.8).

(9.4.4)
$$A' = \begin{array}{c} \\ u \\ v \\ w \end{array} \begin{array}{c} \begin{array}{cccc} p & q & r & s \end{array} \\ \boxed{\begin{array}{cccc} \alpha & \beta & \delta & \varepsilon \\ 1 & 1 & 1 & 0 \\ 1 & \gamma & 0 & 1 \end{array}} \end{array}$$

Submatrix A' of A displaying U_5^2, U_5^3, or F_7 minor

Here $\delta, \epsilon = \pm 1$, while α and β are as-yet-undetermined $\{0, \pm 1\}$ entries. We examine the possible cases for α and β. Let M' be the minor of M corresponding to A'.

Case 1: $\alpha, \beta = \pm 1$. Suppose the 2×2 submatrix $\overline{A}'$ of A' indexed by u, v and p, q has incorrect GF(3)-determinant for M'. Then the rows u, v and the columns p, q, r are up to indices an instance of (9.4.3). Thus, M' has a U_5^2 minor. Hence, we may assume the determinant of $\overline{A}'$ to be correct for M'. Similarly, the GF(3)-determinant of the 2×2 matrix indexed by u, w and p, q may be assumed to be correct for M'. But then up to indices, the columns p, q constitute the transpose of the case (9.4.3), thus giving an U_5^3 minor.

Case 2: $\alpha = \beta = 0$. We GF(3)-pivot in A' on the 1 in row w and column s to obtain case 1.

Case 3: $\alpha = 0$, $\beta \neq 0$: This is the last case, since $\alpha \neq 0$, $\beta = 0$ is symmetric to it by scaling. We may assume $\beta = 1$ because of scaling of row u. If $\delta = -1$ (resp. $\epsilon \neq \gamma$), we GF(3)-pivot on the 1 in row v (resp. w) and column r (resp. s) of A' to get case 1. Hence, assume $\delta = 1$ and $\epsilon = \gamma$. If $\gamma = 1$, a GF(3)-pivot on the 1 in row v and column r of A', followed by a GF(3)-pivot on the 1 in row w and column s, produces case 1. If $\gamma = -1$, then a simple analysis of the GF(3)-determinants proves M' to be isomorphic to F_7.

The conclusions drawn via the preceding GF(3)-pivot arguments are valid for the following reasons. Let one of the above GF(3)-pivots transform A' of (9.4.4) to A''. Let B' be the abstract matrix that represents M' and that corresponds to A'. We claim that the related abstract pivot in B' produces a B'' that is linked to A'' as follows. First, the 2×2 submatrix in the lower left corner has determinant 0 in B'' if and only if the corresponding submatrix of A'' has a nonzero GF(3)-determinant. Thus, that submatrix of A'' has incorrect GF(3)-determinant for M' as does that of A'. Second, let deletion of the first or second column reduce A'' and B'' to A''' and B''', respectively. Then any square submatrix of B''' has determinant 0 if and only if the corresponding square submatrix of A''' has GF(3)-determinant 0. Thus, that submatrix of A''' has a correct GF(3)-determinant for M'. In particular, A''' and B''' have the same support. The preceding claims follow directly from the pivot rules for abstract determinants as described in Section 3.4. □

Theorem (9.4.1) implies Theorem (9.3.2) by elementary arguments. By Theorem (9.2.9), a binary matroid M is regular if and only if it is representable over GF(3). By Theorem (9.4.1), a matroid is representable over GF(3) if and only if it has no U_5^2, U_5^3, F_7, F_7^* minors. But U_5^2 and U_5^3 are not binary, and thus Theorem (9.3.2) follows.

In the final section, we indicate extensions and cite references.

9.5 Extensions and References

The prior work on total unimodularity can be roughly divided into two categories. The first one contains papers investigating the structure of totally unimodular matrices, of minimal $\{0, \pm 1\}$ matrices that are not totally unimodular, and of closely related matrix classes. Relevant references are Cederbaum (1957), Ghouila-Houri (1962), Camion (1963a), (1963b), (1965), (1968), Heller (1963), Veinott and Dantzig (1968), Chandrasekaran (1969), Commoner (1973), Gondran (1973), Padberg (1975), (1976), (1988), Brown (1976), (1977), Tamir (1976), Truemper (1977), (1978), (1980b), (1982b), (1990), (1992b), Kress and Tamir (1980), de Werra (1981), Cunningham (1982a), Chandrasekaran and Shirali (1984), Fonlupt and Raco (1984), Seymour (1985a), Yannakakis (1985), Crama, Hammer, and Ibaraki (1986), Conforti and Rao (1987), and Hempel, Herrmann, Hölzer, and Wetzel (1989).

The second category concerns applications of total unimodularity. Pertinent references are Dantzig and Fulkerson (1956), Hoffman and Kruskal (1956), Hoffman and Kuhn (1956), Motzkin (1956), Heller and Tompkins (1956), Heller (1957), Hoffman (1960), (1974), (1976), (1979), Ford and Fulkerson (1962), Heller and Hoffman (1962), Rebman (1974), Lawler (1976), Baum and Trotter (1978), Hoffman and Oppenheim (1978), Truemper and Chandrasekaran (1978), Truemper and Soun (1979), Soun and Truemper (1980), Maurras, Truemper, and Akgül (1981), Bixby (1982a), (1984a), Schrijver (1983), Tamir (1987), Recski (1989), and Ahuja, Magnanti, and Orlin (1989). A comprehensive treatment is given in Schrijver (1986).

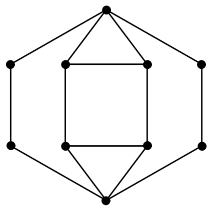
Lemma (9.2.1) is well known, while Lemmas (9.2.2) and (9.2.4) are implicit in most references of the first category. Lemma (9.2.6) is taken from Camion (1963b). It also follows from Brylawski and Lucas (1973). Lemma (9.2.8) and Theorem (9.2.9), though with a quite different proof, are from Tutte (1958), (1965), (1971).

References given above for the second category contain numerous results about problems of the form (9.2.17).

The matroids R_{10} and R_{12} are the two central matroids in the proof of the regular matroid decomposition theorem in Seymour (1980b). The matroid R_{10} had appeared earlier in Hoffman (1960). See also Bixby (1977). Contrary to sometimes-voiced claims, the matroids R_{10} and R_{12} do arise from well-known combinatorial problems. We present two examples. The first one involves the real constraint matrices of two-commodity flow problems on directed graphs. In Soun and Truemper (1980) it is shown that an infinite subclass of such problems has constraint matrices that give rise to regular matroids that are nongraphic and noncographic. Indeed, it can be shown that these matroids have R_{12} minors.

The second example is due to Chvátal (1986). It involves the graph G below.

(9.5.1)



Graph G

Construct the following real matrix B from G. Each column of B corresponds to a node of G, and each row to a clique (= maximal complete subgraph) of G. Each row i of B is then the incidence vector of the nodes in clique i. The matrix B is the *clique/node incidence matrix* of G. The clique/node incidence matrices of graphs are very useful for the solution of the so-called independent vertex set problem. For the graph G at hand, the clique/node incidence matrix B can be proved to be totally unimodular. Indeed, it is not difficult to verify that the regular matroid represented by B is nongraphic and noncographic and has an R_{12} minor.

Theorem (9.3.2) is Tutte's famous characterization of the regular matroids (Tutte (1958), (1965), (1971)). The beautiful proof is due to Gerards (1989a). Variations of the theorem appear in Bixby (1976) and Truemper (1982b), (1992a). Related is work on extremal matroids in Murty (1976).

Theorem (9.4.1) is due to R. Reid, who never published his proof. Other proofs of that result appear in Bixby (1979), Seymour (1979a), Truemper (1982b), Kahn (1984), and Kahn and Seymour (1988). As remarked in Section 9.4, any proof of Theorem (9.4.1) essentially constitutes a proof of Theorem (9.3.2) as well. A characterization of the matroids representable over GF(3) in terms of circuit signatures is given in Roudneff and Wagowski (1989).

Representability over GF(3) and GF(q) is treated in Kung (1990b), and Kung and Oxley (1988). Ternary matroids without $M(K_4)$ minors are covered in Oxley (1987c). Partial results for representability over GF(4) are given in Oxley (1986), (1990a) and Kahn (1988). The difficult characterization of representability over GF(4) in terms of the excluded minors is provided by Geelen, Gerards, and Kapoor (1998). Early examples of nonrepresentable matroids are in MacLane (1936), Lazarson (1958), Ingleton (1959), (1971), and Vamos (1968). It has been conjectured (Rota (1970)) that for any finite field $\mathcal{F}$, the number of matroids that are not representable over $\mathcal{F}$, and that are minimal with respect to that property, is finite.

General questions concerning representability are discussed in Vamos (1978), Kahn (1982), and Ziegler (1990).

The complexity of representability tests using various oracles is examined in Robinson and Welsh (1980), Seymour (1981c), Seymour and Walton (1981), Jensen and Korte (1982), and Truemper (1982a). Virtually every such test requires exponential time when the matroid is given by an oracle that decides independence of sets. An exception is the test of representability over every field. A polynomial algorithm for that problem is given in Truemper (1982a). The algorithm relies on the regular matroid decomposition theorem of Seymour (1980b), which is covered in Chapter 11.