Chapter 12

Almost Regular Matroids

12.1 Overview

So far in this book, we have always used matrices to understand matroids. We have selected a base of the matroid to be investigated. Then we have constructed for that base a representation matrix over some field or even an abstract matrix. Finally, we have analyzed the matrix structure to deduce matroid results. In the third step, we have employed pivots, in particular in the path shortening technique, to modify the matrix in agreement with some change of the matroid base. We have also deleted or added rows and columns to represent matroid reductions or extensions.

Occasionally, we have reversed the just-described roles of matroids and matrices. An example is the test of total unimodularity in Section 11.4 via a test of matroid regularity. A second example is the analysis of the structure of totally unimodular matrices in Section 11.5 via the structure of regular matroids. But generally, it seems to be difficult to answer matrix questions with matroid techniques. In particular, the taking of matroid minors, a most useful matroid operation, has a cumbersome translation into matrix language unless one permits pivots. But the pivot operation almost always changes the matrix structure rather drastically. Thus, pivots often impede the understanding of matrix structure.

The preceding arguments seem to lead to the inescapable conclusion that matrix properties generally cannot be conveniently analyzed with matroids. In particular, one is inclined to accept that conclusion in the following setting. The matroid property in question, say $\mathcal{P}$, is inherited under submatrix-taking. One wants to understand the minimal matrices that do not have $\mathcal{P}$. We call any such matrix a *minimal violation matrix* of $\mathcal{P}$.

In this chapter, we show that the above reasoning, flawless as it may seem, is not valid for the investigation of the minimal violation matrices of certain properties. Indeed, we describe a general matroid technique for a class of such problems where $\mathcal{P}$ observes certain conditions. We demonstrate the technique for the case where $\mathcal{P}$ is the property of regularity. The method deduces a matroid formulation that involves a class of binary matroids called almost regular. An analysis of that class produces the ΔY matroid construction stated earlier in Theorem (4.4.16). Additional analysis leads to a number of matrix constructions of surprising simplicity. Two of the constructions generate the minimal violation matrices for the following two properties: regularity and total unimodularity.

At the outset, the sections of this chapter may appear to be like the pieces of a puzzle: diverse and seemingly unrelated. Only toward the end of the chapter do the relationships and functions of the puzzle pieces emerge. We sketch now the content of each section.

In Section 12.2, we determine for undirected graphs G the minimal subgraphs that rule out success for a certain signing of the edges of G. The signing process is supposed to convert G to a graph G' with edge labels "$+1$" and "-1." The labels are to be such that in every chordless circuit of G', the edge labels sum (in $\mathbb{R}$) to prescribed values specified in a given vector α. If such signing is possible, then we call the resulting graph G' α-balanced. The signing condition is then linked to the set $\mathcal{N}$ of the binary minimal violation matrices of regularity. That way, we determine complete characterizations for two proper subsets of $\mathcal{N}$. We also find an admittedly weak characterization of the remaining matrices of $\mathcal{N}$.

In Section 12.3, we shift our attention from graphs and the set $\mathcal{N}$ to several matrix classes. In particular, we define the class $\mathcal{U}$ of real complement totally unimodular matrices, the classes $\mathcal{A}$ and $\mathcal{B}$ of almost representative matrices over GF(3) and GF(2), respectively, and the class $\mathcal{V}$ of real minimal violation matrices of total unimodularity. The definitions of these classes will give the impression that they are quite unrelated. But in a way, these classes as well as $\mathcal{N}$ are different manifestations of one and the same phenomenon: the absence of matroid regularity plus certain minimality conditions.

Another puzzle piece is introduced in Section 12.4. There we describe the previously mentioned technique for the investigation of the minimal violation matrices of certain matrix properties $\mathcal{P}$. We specialize the general method to the case where $\mathcal{P}$ is the property of regularity. As a result, we define the class of almost regular matroids. We state and sketch a proof of the construction of the almost regular matroids via ΔY extension sequences. The latter result should be familiar since a summary is included in Section 4.4.

The odd and unrelated puzzle pieces of Sections 12.2–12.4 are merged to a rather beautiful picture in Section 12.5. The almost regular matroids are seen to be the matroid manifestation of the matrices of the classes of $\mathcal{A}$, $\mathcal{B}$, $\mathcal{N}$, $\mathcal{U}$, and $\mathcal{V}$. In particular, the construction of the almost regular matroids via ΔY extension sequences produces an elementary construction of the class $\mathcal{U}$ of complement totally unimodular matrices. With $\mathcal{U}$ in hand, we very easily construct the remaining classes $\mathcal{A}$, $\mathcal{B}$, $\mathcal{N}$, and $\mathcal{V}$. We thus have a construction for the binary minimal violation matrices of regularity, and for the real minimal violation matrices of total unimodularity. In the final section, 12.6, we sketch applications and extensions, and cite references.

The chapter assumes knowledge of Chapters 2, 3, 4, 9, and 10.

12.2 Characterization of Alpha-Balanced Graphs

We are given an undirected graph G. For each chordless circuit C of G, we are given an integer number $\alpha_C = 0, 1, 2$, or 3. We collect the α_C in a vector α. To each edge of G, we want to assign the label $+1$ or -1 so that in the resulting graph G' the real sum of the edge labels of each chordless cycle C is congruent (mod 4) to α_C. When such labels can be found, we say that G' is *α-balanced*. Suppose an α-balanced G' cannot be produced. Then G apparently contains a configuration of chordless cycles with conflicting requirements. In this section, we identify the possible sources of such conflicts. Specifically, we pinpoint three subgraph configurations that collectively represent the minimal obstacles to α-balancedness. In the last portion of the section, we specialize that result for the case when α is the zero vector. That specialization leads to a characterization of the binary matrices that may be signed to become so-called balanced $\{0, \pm 1\}$ real matrices. From the latter characterization, we deduce a partition of the class $\mathcal{N}$ of the binary minimal violation matrices of regularity. The partition consist of three subclasses. Two of them are well described, and indeed are readily constructed. But the third subclass has a rather unsatisfactory characterization. One could say that the subsequent sections of this chapter are devoted to replacing that unsatisfactory description with a mathematically appealing and practically useful one.

We begin with the detailed technical discussion. We use a number of terms that we define in the next few paragraphs.

It is convenient for us to consider every graph to be undirected and to have *a priori* a $+1$ or -1 label on each edge. If nothing is said about a graph, then all labels are assumed to be $+1$. Thus, the previously described assignment of $\{\pm 1\}$ labels to G becomes a change of some edge labels. That change we call a *signing* of G. We *scale* a star of G by multiplying the edge labels of that star by -1. We *scale* G by a sequence of scaling steps, each one involving some star of G. The *label sum* of a subgraph $\overline{G}$ is the real sum of the labels of the $\overline{G}$ edges. That sum is denoted by $L(\overline{G})$.

The emphasis on nodes in this section demands that we temporarily abandon our usual viewpoint where nodes are edge subsets. Thus, we view nodes as points, and consider edges to be unordered node pairs. That approach is reasonable since we never have to deal with the contraction operation or with parallel edges. We still view trees, paths, and cycles to be given by their edge sets. Most subgraphs will be induced by some node subset. We use *n-subgraph* to specify that case. Special instances are *n-path* and *n-cycle*. If $\overline{G}$ is a subgraph of G but not an n-subgraph, then G has an edge that is not in $\overline{G}$, but both of whose endpoints are in $\overline{G}$. Such an edge is a *G-chord* for $\overline{G}$.

Let α be an integer vector whose entries are in one-to-one correspondence with the n-cycles of a graph G. Throughout, it is assumed that each entry of α is $0, 1, 2$, or 3. Then G is *α-balanced* if for each n-cycle C the label sum $L(C)$ satisfies $L(C) \equiv \alpha_C (\bmod 4)$. Note that scaling in G changes $L(C)$ by a multiple of 4, and thus does not affect α-balancedness. Suppose a graph G has only $+1$s as edge labels and is not α-balanced. Then possibly an α-balanced graph G' can be deduced from G by signing. The label changes of edges in an n-cycle C, say producing C', modify $L(C)$ for some integer k to $L(C') = L(C) + 2k = |C| + 2k$. Thus, if $L(C') \equiv \alpha_C (\bmod 4)$ is to be achieved at all, then necessarily $|C| \equiv \alpha_C (\bmod 2)$. From now on, we assume that any α satisfies this necessary condition.

The relation "is an n-subgraph of" is transitive. In particular, every n-cycle of an n-subgraph of G is an n-cycle of G. Let α be given for G. By the above observation, it makes sense to apply the term "α-balanced" not just to G, but also to n-subgraphs of G. In particular, α-balancedness of an n-cycle C of G means $L(C) \equiv \alpha_C (\bmod 4)$. We also say in that case that *C agrees with α*.

One may suspect that signing to achieve α-balancedness is essentially unique. The next lemma confirms this notion. The proof is almost identical to that of Lemma (9.2.6).

(12.2.1) Lemma. *Let G and G' be connected α-balanced graphs that are identical up to edge labels. Then G' may be obtained from G by scaling.*

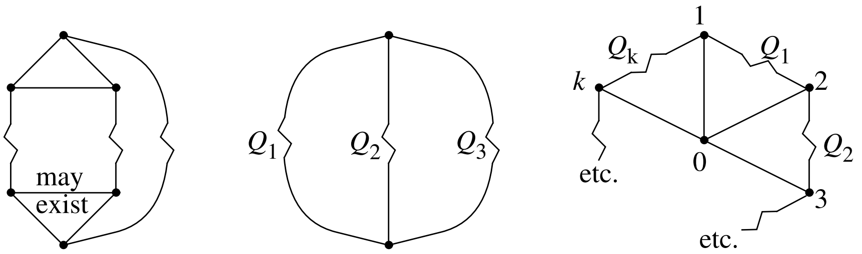
Proof. Let T be a tree of G, and T' be the corresponding tree of G'. Because of scaling, we may suppose that the labels of G and G' agree on the edges of T and T'. Suppose the labels of G and G' differ, say on an edge e of G and on the corresponding edge e' of G'. The edges e and e' form fundamental 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 label of any chord of C must agree with that of the corresponding chord of C'. But then C and C' do not have minimum cardinality, a contradiction. Thus, C and C' are chordless cycles with labels in agreement except for e and e'. But then $L(C)$ and $L(C')$ differ by ± 2, and necessarily $L(C) \not\equiv \alpha_C (\bmod 4)$ or $L(C') \not\equiv \alpha_C (\bmod 4)$, a contradiction. $\square$

In this section, the drawings of graphs follow special rules. A solid straight line connecting two nodes represents an edge, while a solid line with a short zigzag segment indicates a path where all intermediate nodes have degree 2. A broken line represents a path connecting the two endpoints of the broken line, and two or more such paths may have one or more intermediate nodes in common. However, the path of a broken line has no intermediate node in common with any node explicitly shown. The labels on edges are always omitted.

Of particular interest are the following graphs.

(12.2.2)



H_0 $\qquad$ $|Q_i| \geq 3$, odd $\quad H_1$ $\qquad$ $k \geq 3$, odd $\quad H_2$

Graphs of type H_0, H_1, and H_2

With the aid of the following lemma, it is easy to check whether or not a graph G of type H_0, H_1, or H_2 may be signed to become α-balanced for a given α.

(12.2.3) Lemma. *The following statements are equivalent for a graph G of type H_0, H_1, or H_2, and a given α.*

(i) *G can be signed so that an even (resp. odd) number of n-cycles do not agree with α.*
(ii) *G can be signed so that every n-cycle (resp. every n-cycle except a designated one) agrees with α.*

Proof. It is easily seen that one can always sign H_0, H_1, and H_2 such that at most one designated n-cycle does not agree with α. Now every edge of H_0, H_1, and H_2 is part of exactly two n-cycles, and hence every signing of G produces the same number $(\bmod 2)$ of n-cycles that do not agree with α. $\square$

The main theorem of this section follows. It establishes the central role of n-subgraphs of type H_0, H_1, and H_2 when one wants to achieve α-balancedness by signing.

(12.2.4) Theorem. *For a given vector α, a graph G may be signed to become α-balanced if and only if every n-subgraph of type H_0, H_1, and H_2 can be so signed and $\alpha_C \equiv |C| (\text{mod}\, 2)$ for every n-cycle C of G.*

We prove the theorem in a moment. As an aside, we mention a corollary that rephrases the theorem in terms of edge subsets.

(12.2.5) Corollary. *Let β be a $\{0, 1\}$ vector whose entries are in one-to-one correspondence with the n-cycles of a graph G. Then there exists a subset F of the edge set of G such that $|F \cap C| \equiv \beta_C (\text{mod}\, 2)$, for all n-cycles C of G, if and only if the latter condition is true for all n-subgraphs of type H_0, H_1, and H_2 of G.*

Proof. For each n-cycle C of G, define $\alpha_C = 2\beta_C - |C| (\text{mod}\, 4)$. We call the requirement $|F \cap C| \equiv \beta_C (\text{mod}\, 2)$ the *β-condition*. Obviously, we only need to prove the "if" part of the corollary. Thus, we assume that the β-condition can be satisfied for each n-subgraph H of type H_0, H_1, or H_2, say by edge subset F_H of H. Sign H so that precisely the edges of F_H receive a $+1$. By the definition of α and by the β-condition, for any n-cycle C of H there are integral l_C and k_C such that $|F_H \cap C| + 2l = \beta_C = (\alpha_C + 4k_C + |C|)/2$. Because of this equation and the signing rule for H, $L(C) = 2|F_H \cap C| - |C| \equiv \alpha_C (\text{mod}\, 4)$. Thus, H is α-balanced. By Theorem (12.2.4), G can be signed to become α-balanced. Consider G to be so signed. Let F be the subset of edges of G with $+1$ label. Since for any n-cycle C we have $L(C) \equiv \alpha_C (\text{mod}\, 4)$, there are integers k_C and l_C so that $|F \cap C| = (\alpha_C + 4k_C + |C|)/2 = \beta_C + 2l_C$. Thus, F satisfies the β-condition. □

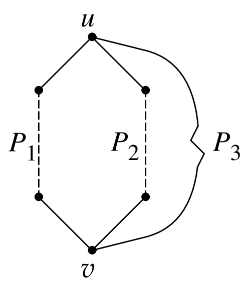
We accomplish the proof of Theorem (12.2.4) in two steps. First we prove a rather technical lemma. Let G be a graph without parallel edges, and P_0 be a path of G where all intermediate nodes have degree 2. For a given vector α, the graph G is *almost α-balanced with respect to P_0* if all n-cycles of G that do not contain P_0 agree with α.

(12.2.6) Lemma. *Let G be a graph without parallel edges, and let vector α be given. Suppose that every n-subgraph of type H_0, H_1, or H_2 of G can be signed to become α-balanced, and that G is almost α-balanced with respect to a given P_0. Assume C_1 and C_2 are n-cycles of G that include P_0, and let $P_3 \subseteq C_1 \cap C_2$ be a path of maximal cardinality satisfying $P_0 \subseteq P_3$. If one of the n-cycles C_1, C_2 agrees with α, then the other n-cycle agrees with α as well, provided one of the following conditions is satisfied:*

(a) *$|P_3| \geq 2$;*
(b) *$|P_3| = 1$ and $(C_1 \cup C_2) - P_3$ is not an n-cycle of $G - P_3$.*

Proof. The proof is by induction on $|P_3|$. The lemma holds trivially for the maximal value that $|P_3|$ may take on since then $C_1 = C_2$. Hence, we will prove validity for $|P_3| = l$ assuming that the lemma holds whenever $|P_3| \geq l+1$. In the nontrivial case, we have $C_1 \neq C_2$. Thus, the endpoints u and v of P_3 have degree 3 in $\overline{G} = C_1 \cup C_2$. The graph $\overline{G}$ is depicted below. P_1, P_2, and P_3 are the three paths u to v, and for $i = 1, 2$, $C_i = P_i \cup P_3$.

(12.2.7)



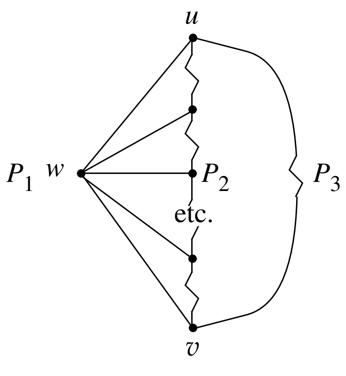
Graph $\overline{G}$

Note that $(C_1 \cup C_2) - P_3$ is the set $P_1 \cup P_2$. Furthermore, $|P_1| \geq 2$ since C_2 is an n-cycle of G. Similarly, $|P_2| \geq 2$. It will be convenient to consider two cases.

(1) $|P_1|$ or $|P_2| = 2$.

Without loss of generality, suppose $|P_1| = 2$. Since C_2 is an n-cycle, the intermediate node of P_1, say w, cannot be a node of P_2. Addition of all edges of G from w to intermediate nodes of P_2 produces the following n-subgraph $\overline{\overline{G}}$ of G.

(12.2.8)



Graph $\overline{\overline{G}}$

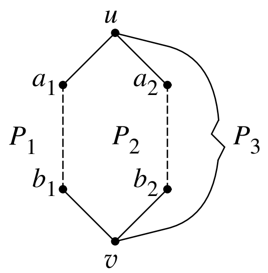
If $|P_3| = 1$, then w of $\overline{\overline{G}}$ has degree of at least 3 since otherwise $(C_1 \cup C_2) - P_3$ is an n-cycle of $G - P_3$. Thus, $\overline{\overline{G}}$ is a graph of type H_1 or H_2, and all n-cycles of $\overline{\overline{G}}$ are α-balanced except possibly for C_1 and C_2, since G is almost α-balanced.

By assumption, $\overline{\overline{G}}$ can be signed to become α-balanced. So by Lemma (12.2.3), C_2 must agree with α if C_1 does, and vice versa.

(2) Both $|P_1|, |P_2| \geq 3$.

For $i = 1, 2$, let a_i, b_i be the vertices of P_i adjacent to u, v, respectively. Thus, $\overline{G}$ is as follows.

(12.2.9)



Graph $\overline{G}$

Clearly, u, v, a_1, a_2, b_1, b_2 are all distinct. Suppose there is a path in $G - P_3$ from u to v using only vertices of $P_1 \cup P_2$ and also avoiding a_2 and b_1. Then there exists an n-path P with these properties, and $P \cup P_3$ is an n-cycle C of G. Now $C \cap C_1$ contains a path that in turn properly contains P_3. Thus, by induction and part (a), C agrees with α if and only if C_1 does. The same conclusion holds for C and C_2, so C_1 and C_2 both agree with α or both do not. Hence, we may suppose that there is no such P. Then P_1 and P_2 have no vertex in common except for u, v. Also, every $(G - P_3)$-chord for the cycle $P_1 \cup P_2$ is incident with a_2 or b_1. If there is no such chord, we must be in case (a) of the lemma, and $C_1 \cup C_2$ is a graph of type H_1. Since $P_1 \cup P_2$ agrees with α, we again have the desired result by Lemma (12.2.3). Hence, suppose there exists at least one such chord. Repeat the above argument, but this time try to find a P avoiding a_1 and b_2. If again we are unsuccessful, then the $(G - P_3)$-chords for $P_1 \cup P_2$ are found only at a_1 or b_2. Thus, there are at most two such chords, one from a_1 to a_2, the other from b_1 to b_2, and $C_1 \cup C_2$ must be a graph of type H_0. All n-cycles agree with α except at most C_1 and C_2, so Lemma (12.2.3) produces the desired conclusion. $\square$

Proof of Theorem (12.2.4). For proof of the nontrivial "if" part, it is sufficient to consider the case where all edges but those incident at some node m have been signed such that $G - \{m\}$ is α-balanced. We want to sign the edges of G incident at m so that G becomes α-balanced. Let C_1 and C_2 be two n-cycles of G, each containing edges (i, m) and (j, m) for some i and j. Derive G_1 from G by deleting all neighbors of m not equal to i or j. Clearly, C_1 and C_2 are contained in G_1, and G_1 is almost α-balanced with respect to $P_0 = \{(i, m), (j, m)\}$. Since G_1 is an n-subgraph of G, every n-subgraph of type H_0, H_1, or H_2 of G_1 can be signed to become α-balanced. So by Lemma (12.2.6), C_1 and C_2 agree with α or they both do not.

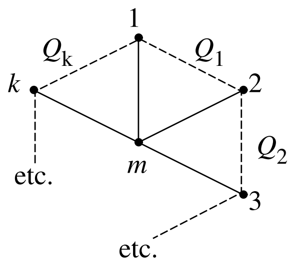
Define a graph J, also with $\{\pm 1\}$ arc labels, from G as follows. Each edge of G incident at node m becomes a node of J. An arc connects two nodes of J if the corresponding edges of G are part of at least one n-cycle C of G. This arc is signed $+1$ if C agrees with α, and -1 otherwise. By the previous argument, the classification of each arc of J is well-defined. For clarity we use "arc" (resp. "edge") in connection with J (resp. G). Suppose we change the sign of edge (i, m) in G. In J, we must change the sign on all arcs incident at node (i, m), so this is a scaling step. Conversely, scaling in J leads to signing of edges in G incident at m. Clearly, it is possible to scale J such that all arcs of a given forest have $+1$ labels. Furthermore, note that a given cycle of J has an even number of -1 arcs if and only if that is true after scaling.

It is claimed that every cycle of J has an even number of -1 arcs. It is sufficient to prove the claim for an n-cycle C_J of J. We will consider two cases depending on $k = |C_J|$.

(1) $k \geq 4$.

By definition of J, the graph G contains the following subgraph $\overline{G}$, where each path Q_i forms an n-cycle C_i with edges (i, m) and $(i+1, m)$ ($k+1$ is interpreted as 1).

(12.2.10)



Graph $\overline{G}$, case (1)

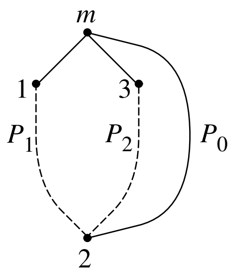
It is claimed that $C_G = \bigcup_{i=1}^{k} Q_i$ is an n-cycle of G. We first show that C_G is a cycle. Suppose Q_1 and Q_j have a node u in common, where $u \neq 2$ if $j = 2$, and where without loss of generality $j \neq k$. Since C_i is an n-cycle, for any i we have $v \notin Q_i$, where $v \neq i, i+1$ is a neighbor of m in G. Hence, u is not equal to any endpoint of Q_1 or Q_j. If $j \neq 2$, define $\overline{P}$ to be composed of the paths from 1 to u on Q_1 and u to j on Q_j. Clearly, $\overline{P}$ contains no neighbor v of m except for nodes 1 and j, and we may replace it by an n-path P observing the same condition. But P and the edges $(1, m)$, (j, m) form an n-cycle of G. But then, since $j \neq k$, C_J cannot be an n-cycle of J. If $j = 2$, $\overline{P}$ is composed of paths 1 to u on Q_1 and u to 3 on Q_2 ($= Q_j$). Again, we conclude that C_J is not an n-cycle of J. Hence, C_G must be a cycle of G. Similar arguments prove that C_G is indeed an n-cycle of G. Since $\overline{G}$ is an n-subgraph of G of type H_2, it can be signed to become α-balanced. By Lemma (12.2.3), an even number of n-cycles of

$\overline{G}$ do not agree with α. C_G cannot be one of these, since it is an n-cycle of $G - \{m\}$. The C_i, $i = 1, 2, \ldots, k$, constitute the remaining n-cycles of $\overline{G}$. The fact that an even number of the C_i do not agree with α, results in an even number of -1 arcs in the n-cycle C_J of J.

(2) $k = 3$.

Again, G has $\overline{G}$ of (12.2.10) as a subgraph. If $C_G = \bigcup_{i=1}^{3} Q_i$ is an n-cycle of G, then arguments as for the case $k \geq 4$ yield the desired conclusion. So suppose Q_1 and Q_2 prevent C_G from being an n-cycle, i.e., Q_1 and Q_2 have a node $i \neq 1, 2, 3$ in common, or there exists a G-chord for $Q_1 \cup Q_2$. Define $P_0 = \{(2, m)\}$, $P_1 = Q_1 \cup \{(1, m)\}$, and $P_2 = Q_2 \cup \{(3, m)\}$. Graph $\overline{G} = \bigcup_{i=0}^{2} P_i$ is shown below.

(12.2.11)



Graph $\overline{G}$, case (2)

Derive G_1 from G by deleting all neighbors of m not equal to 1, 2, or 3. Clearly, $C_i = P_0 \cup P_i$, $i = 1, 2$, is contained in G_1. Because of scaling in J, we may suppose that the arc of J connecting nodes $(1, m)$ and $(3, m)$ has a $+1$ label. This implies that every n-cycle C of G containing edges $(1, m)$ and $(3, m)$ agrees with α, and G_1 is therefore almost α-balanced with respect to P_0. Furthermore, every n-subgraph of G_1 of type H_0, H_1, or H_2 can be signed to become α-balanced. Since $P_1 \cup P_2$ is not an n-cycle of $G_1 - P_0$, either part (a) or part (b) of Lemma (12.2.6) holds, and C_1 agrees with α if and only if C_2 does. But this implies that C_J has an even number of -1 arcs.

The remainder is simple. We scale J so that all arcs of an arbitrarily selected forest T of J receive $+1$ labels. Then all out-of-forest arcs must also have $+1$ labels, since otherwise a cycle with an odd number of -1 arcs has been found. Related signing in G results in an α-balanced graph. $\square$

One application of Theorem (12.2.4) is as follows. Call a binary matrix B *balancedness-inducing* if its 1s can be replaced by ± 1s so that the resulting real matrix A satisfies the following requirement. For each $k \geq 2$, each $k \times k$ submatrix of A of the form given by (12.2.12) below must have real determinant 0. By Lemma (9.2.4), the determinant condition is equivalent to the demand that the entries of the submatrix sum to $0 (\mathrm{mod}\, 4)$.

(12.2.12)

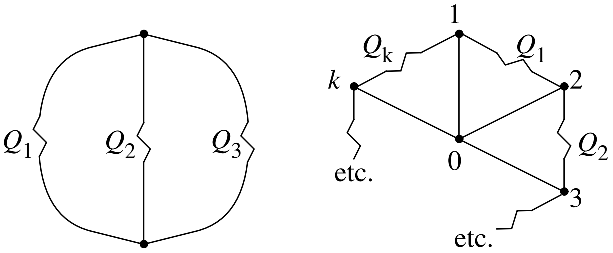
$$\boxed{\begin{array}{ccccc} \pm1 & & & & \pm1 \\ \pm1 & \pm1 & & & \\ & \pm1 & \ddots & & \\ & & \ddots & \pm1 & \\ & & & \pm1 & \pm1 \end{array}}$$

Matrix whose bipartite graph
is a chordless cycle

The next theorem characterizes the class of balancedness-inducing binary matrices. As we shall see, the proof requires one easy application of Theorem (12.2.4).

(12.2.13) Theorem. *A binary matrix B is balancedness-inducing if and only if B does not have a submatrix $\overline{B}$ such that* $\mathrm{BG}(\overline{B})$ *is one of the graphs H_1 or H_2 below.*

(12.2.14)



$|Q_i| \geq 3$, odd $\qquad\qquad$ $k \geq 3$, odd

H_1 $\qquad\qquad$ H_2

Graphs of type H_1 and H_2
for balancedness-inducing matrix case

Proof. Let B be a binary matrix. Define G to be the bipartite graph $\mathrm{BG}(B)$ with additional $+1$ labels on the edges. Each n-cycle of G corresponds precisely to a $\{0, 1\}$ submatrix of B that looks like the matrix of (12.2.12) except for the signs of the entries. Signing B to produce an A with the desired property is equivalent to signing G so that each n-cycle C of the resulting graph has $L(C) \equiv 0 (\mathrm{mod}\, 4)$. Put differently, an appropriate signing of B is possible if and only if G can be signed to become a 0-balanced graph, say G'. We call G' as well as the related $\{0, \pm1\}$ signed real version of B *balanced*. By Lemma (12.2.1), G' is unique up to scaling. The latter operation corresponds to scaling of some rows and columns of A by -1 factors. Thus, up to such scaling, the matrix A is unique. The proof of Lemma (12.2.1) implies a simple scheme to effect the signing if it is possible at all.

By Theorem (12.2.4), G' and thus A exist if and only if every n-subgraph of G of type H_0, H_1, or H_2 (see (12.2.2)) can be signed to become balanced. The graph H_0 is not bipartite, so it cannot be present in the bipartite G. For the same reason, the paths Q_1, Q_2, and Q_3 of any H_1 n-subgraph must have the same parity. One readily sees that H_1 with a $+1$ label on each edge has an odd number of circuits C with $L(C) \equiv 0(\bmod 4)$ if and only if for $i = 1, 2, 3$, $|Q_i|$ is odd and at least 3. By Lemma (12.2.3), the latter condition characterizes the H_1 graphs that cannot be signed to become balanced. Corresponding arguments for H_2 show that H_2 cannot be signed to become balanced if and only if k, which is the number of spokes of the wheel-like graph, is odd and at least 3. $\square$

Balancedness-inducing matrices are related to regular ones as follows.

(12.2.15) Lemma. *Let B be a regular matrix. Then B is balancedness-inducing. Furthermore, any balanced $\{0, \pm 1\}$ real matrix derived from B by signing is totally unimodular.*

Proof. Regularity implies that a totally unimodular matrix A can be deduced from B by signing. By Lemma (9.2.4), every $k \times k$, $k \geq 2$, submatrix A equal to the matrix of (12.2.12) must have its entries sum to $0(\bmod 4)$. Thus, A is balanced. As observed above, any balanced matrix A' derived from B by signing must be obtainable from A by scaling. Thus, A' is totally unimodular. $\square$

Suppose we want to characterize the class $\mathcal{N}$ of binary minimal violation matrices of regularity. Theorem (12.2.13) and Lemma (12.2.15) bring us quite close to that goal, as follows.

(12.2.16) Theorem. *Let $\mathcal{N}$ be the class of binary minimal violation matrices of regularity. Then $\mathcal{N}$ has a partition into the following three subclasses $\mathcal{N}_1$, $\mathcal{N}_2$, $\mathcal{N}_3$.*

(a) *$\mathcal{N}_1$ (resp. $\mathcal{N}_2$) is the set of binary matrices B for which $\mathrm{BG}(B)$ is a graph of type H_1 (resp. H_2) of (12.2.14).*
(b) *$\mathcal{N}_3$ is the set of binary balancedness-inducing matrices B satisfying the following condition. Any balanced $\{0, \pm 1\}$ real matrix A derived from B by signing is a minimal violation matrix of total unimodularity with at least three ± 1s in some row and in some column.*

Proof. Let B be any matrix in $\mathcal{N}$. Thus, B is not regular, but every proper submatrix of B is regular. Suppose B is not balancedness-inducing. By Theorem (12.2.13), $\mathrm{BG}(B)$ has as n-subgraph a graph H of type H_1 or H_2 of (12.2.14). Since H cannot be signed to become balanced, by Lemma (12.2.15) we must have $\mathrm{BG}(B) = H$. Thus, $B \in (\mathcal{N}_1 \cup \mathcal{N}_2)$.

Now suppose B to be balancedness-inducing. Let A be any real $\{0, \pm 1\}$ balanced matrix obtained from B by signing. Since B is not regular, A cannot be totally unimodular. Since every proper submatrix of B is regular, by Lemma (12.2.15) every proper submatrix of A is totally unimodular. Thus, A is a minimal violation matrix of total unimodularity. The nonregularity of B implies that B is not graphic or cographic. Thus, B has a row and a column with at least three 1s. Then A has at least three $\pm$1s in some row or column. We conclude that $B \in \mathcal{N}_3$.

So far we have shown that $\mathcal{N} \subseteq \mathcal{N}_1 \cup \mathcal{N}_2 \cup \mathcal{N}_3$. It is easy to see by (12.2.14) that any $B \in (\mathcal{N}_1 \cup \mathcal{N}_2)$ is a minimal nonregular matrix. Thus, $\mathcal{N}_1 \cup \mathcal{N}_2 \subseteq \mathcal{N}$. Let $B \in \mathcal{N}_3$. By definition, every proper submatrix of B is regular. If B is regular, then by Lemma (12.2.15) any balanced matrix A induced by B is totally unimodular. But this violates the definition of $\mathcal{N}_3$. Thus, $\mathcal{N}_3 \subseteq \mathcal{N}$.

At this point, we know $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2 \cup \mathcal{N}_3$. By Theorem (12.2.13), $\mathcal{N}_1 \cup \mathcal{N}_2$ consist of the minimal binary matrices that are not balancedness-inducing. The matrices of $\mathcal{N}_3$ are by definition balancedness-inducing. Thus, $\mathcal{N}_1 \cup \mathcal{N}_2$ and $\mathcal{N}_3$ are disjoint. By (12.2.14), $\mathcal{N}_1$ and $\mathcal{N}_2$ are disjoint as well. Thus, $\mathcal{N}_1$, $\mathcal{N}_2$, and $\mathcal{N}_3$ partition $\mathcal{N}$ as claimed. $\square$

Theorem (12.2.16) is a substantial step toward the goal of understanding the class $\mathcal{N}$ of binary minimal violation matrices of regularity. Indeed, the subclasses $\mathcal{N}_1$ and $\mathcal{N}_2$ of such matrices have a simple and appealing description. But the characterization of the remaining class $\mathcal{N}_3$ in terms of minimal violation matrices of total unimodularity has little value unless we understand the latter matrices. In the next section, we take small but nevertheless useful steps toward understanding those matrices.

12.3 Several Matrix Classes

We define four classes of matrices. Each of them is connected in some way with the binary minimal violation matrices of regularity. We point out the relationships as the classes are introduced one by one. Their significance will become apparent in Sections 12.4 and 12.5.

Complement Totally Unimodular Matrices

We begin with $\{0,1\}$ real matrices. For such a matrix U, we may define the following complement operations. Let k index a row of U. Then the *row k complement* of U is the real $\{0,1\}$ matrix U' derived from U as follows. For all column indices j with $U_{kj} = 1$ and for all row indices $i \neq k$, replace the entry U_{ij} by its complement, i.e., 1 by 0, and 0 by 1. Let l index a column of U. Then the *column l complement* of U is the transpose of the row l complement of U^t.

For example, if U is the 4×4 identity with indices k and l as shown,

$$U = \begin{array}{c|cccc|} & l & & & \\ \cline{2-5} k & 1 & 0 & 0 & 0 \\ & 0 & 1 & 0 & 0 \\ & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 1 \\ \cline{2-5} \end{array} \tag{12.3.1}$$

Matrix U

then the row k complement of U is the matrix U' below.

$$U' = \begin{array}{|cccc|} \hline 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \hline \end{array} \tag{12.3.2}$$

Row k complement of U

The column l complement of U' is the following matrix U''.

$$U'' = \begin{array}{|cccc|} \hline 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ \hline \end{array} \tag{12.3.3}$$

Column l complement of U'

How many different matrices, up to indices, may be derived from an $m \times n$ $\{0,1\}$ matrix U in a sequence of complement operations? The answer relies on two readily verified claims. For $k \neq l$, let U' be the row k complement of U, and let U'' be the row l complement of U. Then up to a change of indices, the row l complement of U can be seen to be U''. Suppose one obtains a matrix from U by a row complement step followed by a column complement step. Then the same matrix results if one reverses the order of the two steps. By these two claims, any matrix obtainable from a given $m \times n$ $\{0,1\}$ matrix U in a sequence of complement operations may up to indices be produced by at most one row complement step and/or one column complement step. Thus, at most $(m+1)(n+1)$ numerically different matrices may be deduced from U by repeated complement steps.

With respect to total unimodularity, the complement operation can be very destructive. For example, U of (12.3.1) is the 4×4 identity, and thus totally unimodular. Consider the matrix U'' of (12.3.3) deduced from U by two complement steps. In the right-hand corner, U'' has a 3×3 submatrix with real determinant 2. Thus, U'' is not totally unimodular.

We define a real $\{0,1\}$ matrix U to be *complement totally unimodular* if U and all matrices derivable from U by possibly repeated complement operations are totally unimodular. Clearly, complement total unimodularity is maintained under submatrix-taking and complement operations. We collect the complement totally unimodular matrices in a set $\mathcal{U}$. By the above observation and examination of a few additional examples, one quickly sees that complement total unimodularity is a very demanding property. In fact, one is inclined to believe that there are only a few structurally different complement totally unimodular matrices. But that notion turns out to be mistaken, as we shall see in Section 12.5.

Complement total unimodularity of U implies that certain extensions of U are totally unimodular as follows.

(12.3.4) Lemma. *Let U be a real $\{0,1\}$ matrix that is complement totally unimodular. Enlarge U to a matrix U' by adjoining a row or column containing only 1s. Then U' is totally unimodular.*

Proof. Let $\overline{U}$ be any square submatrix of U'. We must show that $\det_{\mathbb{R}} = 0$ or ± 1. In the nontrivial case, $\overline{U}$ intersects the vector of 1s adjoined to U. A real pivot on any 1 of that vector followed by deletion of the pivot row and column produces a matrix $\overline{\overline{U}}$. One readily confirms that, up to scaling by $\{\pm 1\}$ factors, $\overline{\overline{U}}$ is a submatrix of a row or column complement of U. Thus, $\overline{\overline{U}}$ is totally unimodular, and hence $\det_{\mathbb{R}} \overline{U} = 0$ or ± 1 as desired. □

Almost Representative Matrices

We change fields and consider matrices over GF(3) or GF(2). Let A be a matrix over GF(3), and B its support matrix. View B to be over GF(2). Let M be the ternary matroid represented by A over GF(3), and N be the binary matroid represented by B over GF(2). By the derivation of B from A, certain sets are bases of both M and N. In fact, the two matroids may be identical. Suppose they are not. If there is just one set that is a base of one of the matroids and not of the other one, then we say that A *almost represents* N over GF(3), and that B *almost represents* M over GF(2). One may re-express the assumption as follows. The GF(3)-determinant of each square submatrix of A is nonzero if and only if the GF(2)-determinant of the corresponding square submatrix of B is nonzero, except for one square submatrix $\overline{A}$ of A and for the corresponding submatrix $\overline{B}$ of B. By Corollary (3.5.3), we know that $\det_3 \overline{A} \neq 0$ and $\det_2 \overline{B} = 0$. The arguments to come will confirm this fact.

Suppose we perform a GF(3)-pivot on a ± 1 of $\overline{A}$ in A. We also carry out the corresponding GF(2)-pivot on the related 1 of $\overline{B}$ in B. In the matrices so deduced from A and B, the difference between M and N manifests itself by submatrices $\overline{\overline{A}}$ and $\overline{\overline{B}}$ that are smaller than $\overline{A}$ and $\overline{B}$. Indeed, had we carried out the respective pivots just in $\overline{A}$ and $\overline{B}$ and deleted the pivot row and column, we would have obtained $\overline{\overline{A}}$ and $\overline{\overline{B}}$. Because of this reduction possibility, we may assume $\overline{A}$ and $\overline{B}$ to be of order 2×2. A simple case analysis of the 2×2 matrices proves that the difference in determinants between $\overline{A}$ and $\overline{B}$ can be produced in essentially one way. That is, we must have, up to scaling in $\overline{A}$,

$$\overline{A} = \boxed{\begin{matrix} -1 & 1 \\ 1 & 1 \end{matrix}}\ ; \qquad \overline{B} = \boxed{\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}} \tag{12.3.5}$$

Submatrices $\overline{A}$ and $\overline{B}$ proving $M \neq N$

Thus, $\det_3 \overline{A} \neq 0$ and $\det_2 \overline{B} = 0$, as predicted by Corollary (3.5.3). Evidently, the same relationship must have held prior to any pivots. The reader who has covered Section 3.5 surely recognizes the similarity of the above arguments to those of the proof of Theorem (3.5.2) and Corollary (3.5.3). Here we have a GF(3)-matrix A instead of the abstract matrix of Section 3.5.

We continue with $\overline{A}$ and $\overline{B}$ given by (12.3.5). We perform one more GF(3)-pivot on a ± 1 of $\overline{A}$ in A, and also carry out the related GF(2)-pivot in B. This change effectively reduces $\overline{A}$ to a 1×1 matrix $\overline{\overline{A}}$, and $\overline{B}$ to a 1×1 matrix $\overline{\overline{B}}$, for which $\det_3 \overline{\overline{A}} \neq 0$ and $\det_2 \overline{\overline{B}} = 0$. Thus, $\overline{\overline{B}} = [\,0\,]$. Because of scaling of A, we may assume $\overline{\overline{A}} = [-1]$.

Before going on, we record the insight attained so far in the following lemma.

(12.3.6) Lemma. *Let A be a matrix over* GF(3). *View the support matrix B of A to be over* GF(2). *Assume that the* GF(3)*-determinant of each square submatrix of A is nonzero if and only if the* GF(2)*-determinant of the corresponding submatrix of B is nonzero, with the exception of just one submatrix $\overline{A}$ in A and the related submatrix $\overline{B}$ in B, both of order $k \geq 2$. Then by* GF(3)*-pivots within the submatrix $\overline{A}$ and scaling, and by corresponding* GF(2)*-pivots in B, the matrices A and B can be transformed to matrices with determinants agreeing analogously to A and B, except for a submatrix $\overline{\overline{A}} = [-1]$ and $\overline{\overline{B}} = [\,0\,]$.*

For notational convenience, we now redefine A and B to be the matrices produced by the pivots. Thus, $\overline{\overline{A}} = [-1]$ is a submatrix of A, and $\overline{\overline{B}} = [\,0\,]$ is the related submatrix of B. We want to analyze the structure of A and B. To this end, let y be the row index of $\overline{\overline{A}}$ and of $\overline{\overline{B}}$, and let x be the column index. Assume that X (resp. Y) is the index set of the remaining rows (resp. columns) of A and B. Also assume that A has no zero rows or columns. Recall that $\underline{1}$ denotes a column vector containing only 1s. We claim that up to row and column scaling of A, the matrices A and B are of the form given by (12.3.7) below, where U is a $\{0, 1\}$ matrix viewed to be over GF(3) or GF(2) as needed.

(12.3.7)
$$A = \begin{array}{c|c|c|} & x & Y \\ \hline y & -1 & \mathbf{1}^t \\ \hline X & \mathbf{1} & U \\ \hline \end{array}\ ; \qquad B = \begin{array}{c|c|c|} & x & Y \\ \hline y & 0 & \mathbf{1}^t \\ \hline X & \mathbf{1} & U \\ \hline \end{array}$$

Matrix A over GF(3) for M,
and matrix B over GF(2) for N

The claim plus additional facts make up the next theorem.

(12.3.8) Theorem.

(a) *Let A be a matrix over* GF(3) *without zero rows or columns, and let B be a matrix over* GF(2) *of the same size and with the same row and column indices. Assume that the* GF(3)*-determinant of each square submatrix of A is nonzero if and only if the* GF(2)*-determinant of the corresponding submatrix of B is nonzero, except for one 1×1 submatrix $\overline{\overline{A}} = [-1]$ of A and the corresponding submatrix $\overline{\overline{B}} = [\,0\,]$ of B, say with row index y and column index x. Let the remaining rows of A be indexed by X, and the remaining columns by Y. Then up to a scaling of rows and columns of A, the matrices A and B are given by* (12.3.7)*, where U is a $\{0,1\}$ matrix to be viewed over* GF(3) *or* GF(2) *as needed. When U is considered to be real, then it is complement totally unimodular.*

(b) *Let A and B be the matrices of* (12.3.7)*. The column x and the row y must be present, but X or Y may be empty. Assume that the submatrix U of either matrix is a $\{0,1\}$ matrix that is complement totally unimodular when considered real. View A to be over* GF(3)*, and B to be over* GF(2)*. Then the* GF(3)*-determinant of each square submatrix of A is nonzero if and only if the* GF(2)*-determinant of the corresponding submatrix of B is nonzero, except for the 1×1 submatrix $\overline{\overline{A}} = [-1]$ of A indexed by x and y, and the corresponding submatrix $\overline{\overline{B}} = [\,0\,]$ of B.*

Proof. We establish part (a). A and B of (12.3.7) correctly display $\overline{\overline{A}}$ and $\overline{\overline{B}}$. Suppose column x of A contains a 0, say in row $i \in X$. Since A has no zero rows, there is a $j \in Y$ with $A_{ij} = \pm 1$. From the rows x, i and from the columns y, j of A and B, we extract the submatrices

(12.3.9)
$$\overline{A} = \begin{array}{c} \\ x \\ i \end{array}\!\!\begin{array}{c} \begin{array}{cc} y & j \end{array} \\ \begin{bmatrix} -1 & \gamma \\ 0 & \pm 1 \end{bmatrix} \end{array}; \qquad \overline{B} = \begin{array}{c} \\ x \\ i \end{array}\!\!\begin{array}{c} \begin{array}{cc} y & j \end{array} \\ \begin{bmatrix} 0 & \delta \\ 0 & 1 \end{bmatrix} \end{array}$$

Submatrices $\overline{A}$ and $\overline{B}$ of counterexample

We ignore the entries γ and δ. Indeed, for any γ and δ, we have $\det_3 \overline{A} \neq 0$ and $\det_2 \overline{B} = 0$, which contradicts the presumed agreement of determinants. Thus, by scaling in A, we may assume that column x of A and column x of B contain only 1s except for the entry of $\overline{\overline{A}}$ or $\overline{\overline{B}}$ in row y. By symmetry, we also may assume that the row y of A and the row y of B contain only 1s except for the entry in column x.

It remains for us to prove that U is a $\{0,1\}$ matrix that occurs in both A and B, and that U, when viewed as real, is complement totally unimodular. If U of A contains a -1, say in row i and column j, then rows x, i and columns y, j of A define a 2×2 GF(3)-singular submatrix of A. But in B, these rows and columns specify a GF(2)-nonsingular submatrix, a contradiction. Thus, U in A is a $\{0,1\}$ matrix. Because of the agreement of determinants of A and B, the matrix U must also occur in B, this time considered over GF(2), of course. By Theorem (9.2.9), the agreement of determinants on U when viewed over GF(3) and GF(2) implies that U as real matrix is totally unimodular.

Suppose we perform a GF(2)-pivot in column x, row i of B of (12.3.7). The latter matrix is like B except that the indices x and i have traded places and U has been replaced by its row i complement U'. We perform the analogous GF(3)-pivot in A. That pivot plus some scaling with $\{\pm 1\}$ factors produces a matrix A' that is like A except that the indices x and i have switched and U has become U'. The determinants of A' and B' agree analogously to A and B, except for the $[-1]$ submatrix in row y and column i of A' and the corresponding $[\,0\,]$ submatrix of B. By the preceding discussion, U' must be totally unimodular. Using additional pivots, we see that U as real matrix is complement totally unimodular.

We turn to part (b). Let C be a square submatrix of A, and let D be the corresponding submatrix of B. We must show that $\det_3 C$ is nonzero if and only if $\det_2 D$ is nonzero, with the single exception of $C = \overline{\overline{A}}$ and $D = \overline{\overline{B}}$. Suppose C intersects at most one of column x and row y of A. By the complement total unimodularity of U and Lemma (12.3.4), the matrix C is totally unimodular. By Theorem (9.2.9), the determinants of C and D agree as desired. This leaves the case where C and D properly include $\overline{\overline{A}}$ and $\overline{\overline{B}}$, respectively. Carry out a GF(3)-pivot on any 1 in column x of C. Then delete the pivot row and column. Let a matrix C' result. Correspondingly, reduce D by a GF(2)-pivot to D'. Up to scaling by $\{\pm 1\}$ factors in C', both reduction steps produce a submatrix of a row complement of U with an additional row of 1s adjoined. By the above discussion, the determinants of C' and D', and hence of C and D, agree. □

The above proof contains an observation that we want to record as a lemma for future reference.

(12.3.10) Lemma. *Let B be the matrix over* GF(2) *of* (12.3.7). *Then* *a* GF(2)*-pivot in column x or row y of B transforms B to a matrix B' structured like B, except that a row index and a column index have traded places and U has been replaced by a row or column complement of U.*

Collect in sets $\mathcal{A}$ and $\mathcal{B}$ the possible cases of A and B, respectively, of part (b) of Theorem (12.3.8). That is, each $A \in \mathcal{A}$ and each $B \in \mathcal{B}$ is given by (12.3.7), and the $\{0,1\}$ submatrix U of either matrix when viewed as real is complement totally unimodular. We permit the extreme cases where U is trivial or empty. Thus, one of the index sets X and Y, or even both of them, may be empty.

Let N be the matroid represented by some $B \in \mathcal{B}$ over GF(2). By part (b) of Theorem (12.3.8), the corresponding $A \in \mathcal{A}$ almost represents N over GF(3). For this reason, we call $\mathcal{A}$ a collection of *almost representative matrices over* GF(3). Analogously, the matroid M represented by an $A \in \mathcal{A}$ over GF(3) is almost represented by the corresponding $B \in \mathcal{B}$ over GF(2). Thus, $\mathcal{B}$ is a collection of *almost representative matrices over* GF(2).

Minimal Violation Matrices of Total Unimodularity

So far in this section, we have defined three matrix classes: $\mathcal{U}$, $\mathcal{A}$, and $\mathcal{B}$. We need one additional class $\mathcal{V}$, which contains the real minimal violation matrices of total unimodularity. Thus, every $V \in \mathcal{V}$ is not totally unimodular, but this is so for every proper submatrix of V. Clearly, each $V \in \mathcal{V}$ is square. To avoid uninteresting instances, we exclude from $\mathcal{V}$ the cases V of order 1. Thus, each $V \in \mathcal{V}$ is for some $k \geq 2$, a $k \times k$ $\{0, \pm 1\}$ matrix. Let W be the support matrix of V. Consider V to be real or over GF(3) as needed below, and W to be over GF(2). We have the following result.

(12.3.11) Theorem. *Let V be any matrix of $\mathcal{V}$, and let W be its binary support matrix. Then by* GF(3)*-pivots in V and scaling with $\{\pm 1\}$ factors, and by corresponding* GF(2)*-pivots in W, the matrices V and W can be transformed to matrices $A \in \mathcal{A}$ and $B \in \mathcal{B}$, respectively, of order at least 2.*

Conversely, suppose in $A \in \mathcal{A}$ and $B \in \mathcal{B}$ of order at least 2, we perform GF(3)*-pivots and related* GF(2)*-pivots, respectively, so that A' and B' result that satisfy the following condition. Let $\overline{X}$ be the subset of $X \cup \{x\}$ indexing columns of A' and B', and $\overline{Y}$ be the subset of $Y \cup \{y\}$ indexing rows. The condition is that $|\overline{X}| \geq 2$ or $|\overline{Y}| \geq 2$. Then the submatrix $\overline{A}$ of A' indexed by $\overline{X}$ and $\overline{Y}$ is in $\mathcal{V}$, and the corresponding submatrix $\overline{B}$ of B' is the support of $\overline{A}$.*

Proof. We start with the first part, where V and W are given. Let $\overline{V}$ be a proper submatrix of V. Then $\overline{V}$ as real matrix is totally unimodular. Thus, by Theorem (9.2.9), $\overline{V}$ as GF(3) matrix has $\det_3 \overline{V}$ nonzero if and only if the corresponding $\overline{W}$ of W has $\det_2 \overline{W}$ nonzero. By analogous arguments, exactly one of $\det_3 V$ and $\det_2 W$ is nonzero. Thus, V and W constitute a pair of matrices to which Lemma (12.3.6) can be applied. Accordingly, V and W can by pivots and scaling be transformed to A over GF(3) and B over GF(2), respectively, with agreeing determinants, except for a submatrix $\overline{\overline{A}} = [-1]$ of A and $\overline{\overline{B}} = [\,0\,]$ of B. By part (a) of Theorem (12.3.8), we have $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

Reversal of the above arguments essentially proves the converse part. We only need to show that the matrices $\overline{A}$ of A' and $\overline{B}$ of B' deduced from $A \in \mathcal{A}$ and $B \in \mathcal{B}$ are the ones with disagreeing determinants. Let M (resp. N) be the matroid represented by A over GF(3) (resp. B over GF(2)). In the matroid M, the set $X \cup \{x\}$ is a base. But that set is not a base in N. By assumption, the set $\overline{X}$ with $|\overline{X}| \geq 2$ is the subset of $X \cup \{x\}$ indexing columns of A', and $\overline{Y}$ is the subset of $Y \cup \{y\}$ indexing rows. Since $X \cup \{x\}$ is a base of M but not of N, the submatrix $\overline{A}$ of A' indexed by $\overline{X}$ and $\overline{Y}$ has $\det_3 \overline{A}$ nonzero, while the corresponding submatrix $\overline{B}$ of B' has $\det_2 \overline{B} = 0$. □

Several simple but useful results follow from Theorem (12.3.11). Recall from Section 2.3 that a $\{0, \pm 1\}$ matrix is *Eulerian* if in every row and every column the entries sum to 0(mod 2). Equivalently, each row and column must have an even number of nonzeros.

(12.3.12) Corollary.

(a) *A square $\{0, \pm 1\}$ matrix of order at least 2 is in $\mathcal{V}$ if and only if the following holds. V can be scaled with $\{\pm 1\}$ factors to become, for some square, nonsingular, complement totally unimodular matrix U, the matrix*

$$\begin{array}{|c|c|}\hline \alpha & a \\ \hline b & U^{-1} \\ \hline \end{array}\ ; \quad \begin{array}{l} a = \underline{1}^t \cdot U^{-1} \\ b = U^{-1} \cdot \underline{1} \\ \alpha = \underline{1}^t \cdot U^{-1} \cdot \underline{1} - 2 \end{array} \tag{12.3.13}$$

Matrix V up to scaling

(b) *For every matrix $V \in \mathcal{V}$, the real inverse of V contains only $\frac{1}{2}$ entries, $|\det_{\mathbb{R}} V| = 2$, V is Eulerian, and the real sum of the entries of V in $\mathbb{R}$ is congruent to* 2(mod 4).

Proof. We start with part (a). By Theorem (12.3.11), we may deduce V, up to scaling with $\{\pm 1\}$ factors, from a square GF(3)-nonsingular $A \in \mathcal{A}$ by GF(3)-pivots. Redefine V so that it is the appropriately scaled version of the original V. The matrix A is given by (12.3.7). We may recreate A from V by performing the GF(3)-pivots in reverse order. Suppose we use real pivots instead. All intermediate real matrices must be numerically identical to the GF(3)-matrices in the original GF(3)-pivot sequence, by virtue of the fact that each 2×2 submatrix of each such real matrix is totally unimodular. A different conclusion applies to the last pivot. If performed in GF(3), it would produce A. But carried out in $\mathbb{R}$, it produces a matrix $\tilde{A}$ containing $\det_{\mathbb{R}} V$ as entry where A has the -1. All other entries of $\tilde{A}$ must agree numerically with those of A. Now $\det_{\mathbb{R}} V$ is the real determinant of a 2×2 submatrix of the predecessor matrix of the real pivot sequence. Since $\det_{\mathbb{R}} V \neq 0, \pm 1$, we must have $|\det_{\mathbb{R}} V| = 2$. We know that the final pivot, when done in GF(3), produces the -1 of A instead of $\det_{\mathbb{R}} V$. Thus, $\det_{\mathbb{R}} V$ is congruent to $-1 (\bmod 3)$, and accordingly $\det_{\mathbb{R}} V = 2$.

We just have proved that $\tilde{A}$ is the matrix

(12.3.14)
$$\tilde{A} = \begin{array}{c|c|c|} & x & Y \\ \hline y & 2 & \underline{1}^t \\ \hline X & \underline{1} & U \\ \hline \end{array}$$

Matrix $\tilde{A}$ derived by real pivots from V

The submatrix U is complement totally unimodular. By the existence of the real pivot sequence, $\det_{\mathbb{R}} U$ must be nonzero. Thus, $|\det_{\mathbb{R}} U| = 1$.

Suppose we employ the following well-known method for computing the inverse of V. We begin with the real matrix $[I \mid V]$. Then we carry out elementary row operations until the submatrix V of $[I \mid V]$ has become an identity matrix. At that time, the submatrix I of $[I \mid V]$ has become V^{-1}. We claim that the matrix $\tilde{A}$ of (12.3.14) contains in compact form the results of most of these row operations. For a proof, we first note that each one of the real pivots deducing $\tilde{A}$ of (12.3.14) from a scaled version of V involves a ± 1 as pivot element. Then by the relationship between elementary row operations and pivots described in Section 2.3, the pivots producing $\tilde{A}$ correspond to row operations and scaling steps with $\{\pm 1\}$ factors in $[I \mid V]$ that convert the submatrix I of $[I \mid V]$ to the matrix

(12.3.15)
$$\tilde{V}^1 = \begin{array}{|c|c|} \hline 1 & \underline{1}^t \\ \hline 0 & U \\ \hline \end{array}$$

Matrix $\tilde{V}^1$ derived from $[I \mid V]$
by row operations and scaling

and that change the submatrix V of $[I \mid V]$ to the matrix

(12.3.16)
$$\tilde{V}^2 = \begin{array}{|c|c|} \hline 2 & 0 \\ \hline \underline{1} & \begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix} \\ \hline \end{array}$$

Matrix $\tilde{V}^2$ derived from V
by row operations and scaling

Observe that the columns of $\tilde{A}$ save the first one occur in $\tilde{V}^1$, and that the first column of $\tilde{A}$ is also the first column of $\tilde{V}^2$, as is implied by the discussion of Section 2.3.

We now perform row operations in $[\tilde{V}^1 \mid \tilde{V}^2]$ that convert $\tilde{V}^2$ to an identity matrix. Correspondingly, $\tilde{V}^1$ becomes the inverse matrix of a scaled version of V. Up to scaling, that inverse matrix is

$$\tilde{V} = \frac{1}{2}\begin{array}{|c|c|}\hline -1 & \underline{1}^t \\ \hline \underline{1} & \tilde{U} \\ \hline \end{array}\ ; \qquad \tilde{U} = 2U - \underline{1} \cdot \underline{1}^t \tag{12.3.17}$$

Inverse matrix $\tilde{V}$ of scaled version of V

A multiplication check verifies that the matrix of (12.3.13) is $\tilde{V}^{-1}$, and thus is V up to scaling. This fact proves part (a).

Part (b) is now easily shown. We already know $|\det_{\mathbb{R}} V| = 2$. That V^{-1} is a $\{\pm\frac{1}{2}\}$ matrix is evident from the scaled version $\tilde{V}$ of V^{-1} given by (12.3.17). The matrix V is Eulerian, since this is clearly so for the scaled version given by (12.3.13). The latter matrix has its entries sum to $2 (\mathrm{mod}\, 4)$. Scaling does not affect that result, so the same conclusion applies to V. $\square$

Let us summarize the main results of this section for the matrix classes $\mathcal{U}$, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{V}$. The class $\mathcal{U}$ contains the real $\{0,1\}$ complement totally unimodular matrices. $\mathcal{A}$ and $\mathcal{B}$ contain the almost representative matrices over GF(3) and GF(2), respectively, as depicted by (12.3.7). From any one of the three classes, the remaining two are obtained by trivial operations, as is evident from (12.3.7). The class $\mathcal{V}$ contains the real minimal violation matrices of total unimodularity of order at least 2. By (12.3.13), each $V \in \mathcal{V}$ can be readily computed from a square $\mathbb{R}$-nonsingular $U \in \mathcal{U}$. With equal ease, we can derive $\mathcal{V}$ from $\mathcal{A}$ or $\mathcal{B}$. But note that we do not know how to deduce the entire class $\mathcal{U}$, or $\mathcal{A}$ or $\mathcal{B}$, from $\mathcal{V}$. It turns out that this is not possible by the matrix operations of scaling by $\{\pm 1\}$ factors, pivots, submatrix-taking, and change of fields from $\mathbb{R}$ to GF(2) or GF(3). We prove this fact in Section 12.5. There we construct $\mathcal{U}$, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{V}$, taking the following viewpoint. First we construct $\mathcal{U}$ by a process yet to be described. Then we deduce $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{V}$ as just mentioned.

Recall that Theorem (12.2.16) establishes a partition of the class $\mathcal{N}$ of binary minimal violation matrices of regularity. Two subclasses labeled $\mathcal{N}_1$ and $\mathcal{N}_2$ are well described by that theorem. But the third class $\mathcal{N}_3$ is not well explained. Indeed, each $B \in \mathcal{N}_3$ is the support matrix of a minimal violation matrix V of total unimodularity with at least three nonzeros in some row and some column. Thus, $V \in \mathcal{V}$. As an aside, the just-mentioned bound of 3 on the number of nonzeros in some row and some column of V can now be strengthened to 4 by part (b) of Corollary (12.3.12). At any rate, the construction of $\mathcal{V}$ via $\mathcal{U}$ gives a construction of $\mathcal{N}_3$. So the yet-to-be-described construction process for $\mathcal{U}$ effectively brings to an end the quest for an understanding of the structure of $\mathcal{N}$.

The promised construction of $\mathcal{U}$ relies on some matroid results that we introduce in the next section.

12.4 Definition and Construction of Almost Regular Matroids

As argued in the introductory section of this chapter, one is tempted to claim that matroids are not suitable for investigations of minimal violation matrices. There it is also claimed that this argument is flawed. Here we show why, by providing a general method for a matroid-based investigation of the minimal violation matrices of certain matrix properties. We specialize the method to a particular instance. In doing so, we define and analyze the matroids that we called almost regular in Section 4.4. In particular, we establish the construction for almost regular matroids that was already listed in Section 4.4. We use that construction to obtain a construction for the class $\mathcal{U}$ of the complement totally unimodular matrices.

We begin with a general discussion about matrix properties and matroids. Let $\mathcal{P}$ be a property defined for the matrices over a field $\mathcal{F}$, where $\mathcal{F}$ must be GF(2) or GF(3). Technically, one may consider $\mathcal{P}$ to be a subset of the matrices over $\mathcal{F}$. The property is to be maintained under submatrix-taking, row and column permutations, scaling with $\{\pm 1\}$ factors, and $\mathcal{F}$-pivots, and when a row or column unit vector is adjoined.

Examples for $\mathcal{P}$ are regularity, graphicness, cographicness, graphic-or-cographicness, and planarity, all defined for $\mathcal{F} = \mathrm{GF}(2)$. Also qualifying is the following property for $\mathcal{F} = \mathrm{GF}(3)$. A $\{0, \pm 1\}$ matrix has the property when over $\mathbb{R}$ it is totally unimodular.

Suppose we want to understand the matrices over $\mathcal{F}$ that are minimal violation matrices of $\mathcal{P}$. For any matrix A over $\mathcal{F}$, define $M(A)$ to be the matroid represented by A over $\mathcal{F}$. $M(A)$ may be representable over $\mathcal{F}$ by a number of different matrices A'. If $\mathcal{F} = \mathrm{GF}(2)$, then all such A' are obtainable from A by GF(2)-pivots. By assumption, both A' and A have $\mathcal{P}$ or they do not. Thus, it is well defined when we declare $M(A)$ to have $\mathcal{P}$ if A has $\mathcal{P}$. The same conclusion applies when $\mathcal{F} = \mathrm{GF}(3)$. This time any A' representing $M(A)$ over GF(3) may, by a slightly modified proof of Lemma (9.2.6), be obtained from A by GF(3)-pivots and scaling with $\{\pm 1\}$ factors. By the assumptions on $\mathcal{P}$, thus A' and A have $\mathcal{P}$ or they both do not.

Note that series or parallel extensions of $M(A)$ maintain $\mathcal{P}$, for such an extension corresponds to at most one $\mathcal{F}$-pivot in A followed by adjoining of a row or column unit vector. By assumption, the latter operations maintain $\mathcal{P}$.

We describe a five-step process that leads to insight into the minimal violation matrices of $\mathcal{P}$ over $\mathcal{F}$.

Step 1. Let $\mathcal{W}$ be the class of minimal matrices over $\mathcal{F}$ that do not have $\mathcal{P}$. Assign to the indices of each matrix $W \in \mathcal{W}$ the following additional labels. If an index z labels a row (resp. column) of W, then assign the label "*con*" (resp. "*del*"). For example, if the rows of W are indexed by X and the columns by Y, then W with the labels is the matrix of (12.4.1) below. From now on, we assume each $W \in \mathcal{W}$ to be so labeled. We also assume that the elements of the matroid $M(W)$ are labeled correspondingly.

(12.4.1)
$$\begin{array}{c|c|} & \overset{Y}{\begin{array}{ccc} d & & d \\ e & \cdots & e \\ l & & l \end{array}} \\ \hline X \begin{array}{c} con \\ \vdots \\ con \end{array} & W \\ \hline \end{array}$$

Minimal violation matrix W with labels

The unusual "*con*" and "*del*" labels of W and $M(W)$ tell the following. If element z of the matroid $M(W)$ has a "*con*" label, then $z \in X$. By the minimality of W, deletion of row $z \in X$ from W results in a matrix W' with $\mathcal{P}$. Correspondingly, $M(W')$, which is $M(W)/z$, has $\mathcal{P}$. The "*con*" label on z allows us to predict this outcome for $M(W)/z$. That is, "*con*"traction of an element of $M(W)$ with a "*con*" label produces a minor having $\mathcal{P}$. Similarly, "*del*"etion of an element with a "*del*" label results in a minor having $\mathcal{P}$. Note that we cannot tell from the labels whether for an element z with "*con*" label the minor $M\backslash z$ has $\mathcal{P}$. Similarly, when z has a "*del*" label, we are ignorant about M/z having or not having $\mathcal{P}$.

Collect in a set $\mathcal{M}_1$ the matroids $M(W)$ with $W \in \mathcal{W}$. For brevity, we call an element with "*con*" (resp. "*del*") label simply a "*con*" (resp. "*del*") element.

Step 2. Establish elementary facts about the matrices $W \in \mathcal{W}$. If possible, translate some of these facts into matroid language so that they apply to the matroids of $\mathcal{M}_1$. Let $\mathcal{E}$ be the collection of such matroid facts.

Step 3. At this time, we reverse the sequence of arguments. We use certain necessary conditions satisfied by the matroids of $\mathcal{M}_1$ to define a class $\mathcal{M}_2$ of matroids representable over $\mathcal{F}$. The conditions for membership in $\mathcal{M}_2$ are as follows. Each $M \in \mathcal{M}_2$ must not have $\mathcal{P}$. Each one of its elements must be labeled "*con*" or "*del*" in such a way that a "*con*" (resp. "*del*") label on an element z of M implies M/z (resp. $M\backslash z$) to have $\mathcal{P}$. Finally, M must satisfy the conditions of $\mathcal{E}$. Clearly, $\mathcal{M}_1$ is a subset of $\mathcal{M}_2$.

Step 4. Enlarge $\mathcal{M}_2$ by adding for each member M all proper minors. The elements of these minors are labeled in agreement with the labels of M. By definition, $\mathcal{M}_2$ is now closed under minor-taking.

Step 5. Analyze the structure of the matroids of $\mathcal{M}_2$. Specialize the conclusions for $\mathcal{M}_2$ to $\mathcal{M}_1$. Finally, translate the latter results to statements about the matrices $W \in \mathcal{W}$.

The above procedure is of course more of a recipe than an algorithm. Let us demonstrate its use by applying it to the case where $\mathcal{F}$ is GF(2) and $\mathcal{P}$ is regularity.

Step 1. By Theorem (12.2.16), we have a good understanding of a portion of the class $\mathcal{N}$ of minimal violation matrices of regularity. Indeed, at this point, only the subclass $\mathcal{N}_3$ is poorly characterized. That class consists of the binary support matrices of the minimal violation matrices of total unimodularity with at least three (by Corollary (12.3.12), at least four) nonzeros in some row and some column. Thus, we take $\mathcal{W}$ to be $\mathcal{N}_3$. We assign "*con*" and "*del*" labels to the rows and columns, respectively, of the matrices of $\mathcal{W}$. Then we define $\mathcal{M}_1$ to be the set of matroids $M(W)$ with $W \in \mathcal{W}$.

Step 2. By Corollary (12.3.12), each matrix of $\mathcal{N}_3$ is Eulerian. That is, each $W \in \mathcal{W}$ as given by (12.4.1) has an even number of 1s in each row and each column. This latter fact has a convenient translation into matroid language: Each $M \in \mathcal{M}_1$ has a base such that each fundamental circuit (resp. cocircuit) has an even number of "*con*" (resp. "*del*") elements. Now each circuit (resp. cocircuit) of M is the symmetric difference of some of these fundamental circuits (resp. cocircuits). We conclude that each circuit (resp. cocircuit) has an even number of "*con*" (resp. "*del*") elements. Another fact, trivial yet important, is that M has at least one "*con*" element and at least one "*del*" element. The preceding conditions on circuits, cocircuits, and labels we declare to be the collection $\mathcal{E}$ of matroid facts about $\mathcal{M}_1$.

Step 3. We reverse the arguments and define $\mathcal{M}_2$ from facts known for $\mathcal{M}_1$. Specifically, $\mathcal{M}_2$ is the class of binary matroids M satisfying the following conditions. First, M must be nonregular. Second, each element z of M must be labeled "*con*" or "*del*." The "*con*" (resp. "*del*") label must imply that M/z (resp. $M\backslash z$) is regular. Third, the circuits and cocircuits of M must obey the following *parity condition*: Each circuit (resp. cocircuit) is to have an even number of "*con*" (resp. "*del*") elements. Fourth and last, the following *existence condition* must be satisfied: There is to be at least one "*con*" element and at least one "*del*" element. The matroids of $\mathcal{M}_2$ so defined we call *almost regular*.

Step 4. We enlarge $\mathcal{M}_2$ by adding all possible minors. Each minor assumes the labels of the matroid producing it. We claim that each minor so added to $\mathcal{M}_2$ is regular or almost regular. By duality and induction, we only need to consider a 1-element deletion of an element z in an almost regular M. If $M\backslash z$ is regular, we are done. So assume $M\backslash z$ to be nonregular. A "*con*" (resp. "*del*") label on an element $w \neq z$ of M implies M/w (resp. $M\backslash w$) to be regular. Thus, for such w the minor $(M\backslash z)/w$ (resp. $(M\backslash z)\backslash w$) of $M\backslash z$ is regular as well. Since $M\backslash z$ is nonregular, z must have a "*con*" label. By (3.3.11), each circuit of $M\backslash z$ is a circuit of M, and each cocircuit of $M\backslash z$ is a minimal member of the collection $\{C^* - \{z\} \mid C^* =$ cocircuit of $M\}$. By the parity condition, each circuit (resp. cocircuit) of M has an even number of "*con*" (resp. "*del*") elements. Since z has a "*con*" label, the same conclusion applies to the circuits and cocircuits of $M\backslash z$. The parity and existence conditions for M imply that M has at least two "*con*" elements and at least two "*del*" elements. Thus, $M\backslash z$ satisfies the existence condition. These arguments establish $M\backslash z$ to be almost regular. They also prove that the enlarged $\mathcal{M}_2$ is the class of almost regular matroids plus their regular minors. As desired, $\mathcal{M}_2$ is now closed under minor-taking.

Step 5. We must analyze $\mathcal{M}_2$. Then we must specialize the results to $\mathcal{M}_1$. Finally, we must express the latter results in matrix terminology to obtain conclusions about $\mathcal{W}$, and hence about $\mathcal{N}_3$.

In the remainder of this section, we carry out the tasks mandated by Step 5. We begin with some elementary facts about almost regular matroids. Let M be such a matroid. We dualize M in the usual way, but also switch "*con*" (resp. "*del*") labels to "*del*" (resp. "*con*") labels. As usual, we denote that dual matroid of M by M^*. The next lemma establishes M^* to be almost regular and provides some additional facts about M. A *hyperplane* of a matroid M is a maximal set of M with rank equal to the rank of M minus 1. A *cohyperplane* is a hyperplane in the dual matroid M^* of M.

(12.4.2) Lemma. *Let M be an almost regular matroid. Then the following holds.*

(a) *The dual matroid M^* of M is almost regular.*
(b) *Every nonregular minor of M is almost regular.*
(c) *The set of "con" (resp. "del") elements of M is a cocircuit and a cohyperplane (resp. a circuit and a hyperplane) of M.*

Proof. (a) This is proved by routine checking of the definition of almost regular matroids given in Step 3 above.
(b) The proof is given under Step 4 above.
(c) Since M is nonregular, it has by Theorem (9.3.2) an F_7 or F_7^* minor. By duality and part (a), we may assume presence of F_7. The "*con*" and "*del*" labels for F_7 are unique up to isomorphism, as may be checked by a simple case analysis. Indeed, the matrix B^7 given by (12.4.3) below is a representation matrix of F_7 with correct "*con*" and "*del*" labels.

(12.4.3)

$$B^7 = \begin{array}{r|cccc|} & del & con & con & con \\ \hline con & 0 & 1 & 1 & 1 \\ del & 1 & 1 & 0 & 1 \\ del & 1 & 0 & 1 & 1 \\ \hline \end{array}$$

Matrix B^7 for matroid F_7
with "*con*" and "*del*" labels

Take B to be any representation matrix of M that displays an F_7 minor in agreement with B^7 of (12.4.3). We claim that B can be partitioned as

(12.4.4)

$$B = \begin{array}{cr|cccc|c|} & & & Y_1 & & & Y_2 \\ & & del & con & con & con & con \\ \hline & con & 0 & 1 & 1 & 1 & a \\ X_1 & del & 1 & 1 & 0 & 1 & \\ & del & 1 & 0 & 1 & 1 & \\ \hline X_2 & del & b & & & & 0/1 \\ \hline \end{array}$$

Matrix B displaying F_7 minor

Note that up to indices, B^7 is displayed in the upper left corner of B. For each $x \in X_2$, the minor M/x is nonregular. Thus, x must have a "*del*" label. Similarly, for each $y \in Y_2$, nonregularity of $M \backslash y$ implies a "*con*" label for y. So far, we have justified all labels of B.

We claim that the subvectors a and b of B contain only 1s. If the subvector b has a 0, say in row $x \in X_2$, then the 1s in that row of B correspond to a cocircuit of M with exactly one "*del*" element, in violation of the parity condition. Similarly, a 0 in the subvector a contradicts the parity condition for a fundamental circuit of M.

At this point, we have shown that the first column and the first row of B contain only 1s except for the 0 in the $(1, 1)$ position. This fact, plus the given assignments of labels to B, implies that the elements with "*con*" (resp. "*del*") labels form a cocircuit and cohyperplane (resp. circuit and hyperplane) of M, as is easily confirmed by direct checking. □

We now review the construction of the almost regular matroids as described in Section 4.4. The starting point for the construction of an almost regular matroid is either F_7 or a 1-element extension of R_{10}, both with appropriate "*con*" and "*del*" labels. The representation matrix of F_7 we have already seen. It is B^7 of (12.4.3). The matrix for the extension of R_{10} we call B^{11}. It is derived from the matrix of (10.2.8) as follows. We permute the rows of the matrix of (10.2.8) so that the last row becomes the first one. To the resulting matrix, we adjoin a new leftmost column. Then we assign appropriate "*con*" and "*del*" labels. The matrices B^7 and B^{11} have already been listed under (4.4.13). We repeat them here for ready reference.

(12.4.5)

$$
B^7 = \begin{array}{c|cccc|}
 & del & con & con & con \\ \hline
con & 0 & 1 & 1 & 1 \\
del & 1 & 1 & 0 & 1 \\
del & 1 & 0 & 1 & 1 \\ \hline
\end{array}\ ; \qquad
B^{11} = \begin{array}{c|cccccc|}
 & del & con & con & con & con & con \\ \hline
con & 0 & 1 & 1 & 1 & 1 & 1 \\
del & 1 & 1 & 0 & 0 & 1 & 1 \\
del & 1 & 1 & 1 & 0 & 0 & 1 \\
del & 1 & 0 & 1 & 1 & 0 & 1 \\
del & 1 & 0 & 0 & 1 & 1 & 1 \\ \hline
\end{array}
$$

Labeled matrices B^7 and B^{11}

By now, the reader has acquired enough machinery, in particular that of graphs plus T sets of Section 10.2, that he/she can quickly verify the matroid of B^{11} to be almost regular. Thus, we omit details of that check.

As stated above, one of the two matroids represented by B^7 or B^{11} is the starting point for the construction of an almost regular matroid. The construction itself consists of a sequence of series or parallel extension steps and of triangle-to-triad and triad-to-triangle exchanges. The extension steps we call SP *steps*, and the exchanges steps, ΔY *exchanges*. These operations are controlled by rules that may be summarized as follows. A parallel (resp. series) extension is permitted only if the involved element z has a "*con*" (resp. "*del*") label. The new element receives the same label as z. The ΔY exchanges are depicted in terms of representation matrices by (4.4.5)–(4.4.7). The conditions on labels are summarized by (4.4.14). Instead of covering all possibilities for the ΔY exchange as done in Section 4.4, here we show just one instance based on (4.4.6). All other cases can be reduced by GF(2)-pivots to the one displayed. In (12.4.6) below, the matrix on the left represents the matroid with the triangle $\{e, f, g\}$, and the one on the right the matroid with the triad $\{x, y, z\}$.

(12.4.6)

$$
\begin{array}{cc|c|c|c|}
 & & & e & g \\
 & & & con & con \\ \hline
 & & \overline{B} & b & b \\ \hline
f & del & a & 1 & 0 \\ \hline
\end{array}
\quad
\begin{array}{c}
\xrightarrow{\text{replace triangle by triad}} \\
\xleftarrow{\text{replace triad by triangle}}
\end{array}
\quad
\begin{array}{cc|c|c|}
 & & & x \\
 & & & con \\ \hline
 & & \overline{B} & b \\ \hline
z & del & a & 1 \\ \hline
y & del & a & 0 \\ \hline
\end{array}
$$

ΔY exchange rule for almost regular matroids

In Section 4.4, we define a sequence of SP extensions and ΔY exchanges under the preceding rules to be a *restricted* ΔY *extension sequence.* The next result justifies our reliance on such sequences.

(12.4.7) Lemma. *Let a matroid M' be created from an almost regular matroid by a restricted ΔY extension sequence. Then M' is almost regular.*

Proof. By induction, we may assume M' to be derived from M in a single SP extension or ΔY exchange step. Consider the first case. Clearly, M' has a minor isomorphic to M. Thus, M' is nonregular since M is almost regular. The parity condition for M' is readily verified because of the restriction that a parallel (resp. series) extension is only permitted for a "*con*" (resp. "*del*") element of M. For the same reason, for each "*con*" (resp. "*del*") element of M', we have M'/z (resp. $M'\backslash z$) regular. Clearly, both "*con*" and "*del*" labels occur in M'. Thus, M' is almost regular.

The ΔY exchange is almost as easily proved. By Theorem (11.2.11), such an exchange maintains regularity. Thus, by contradiction, M' is nonregular. The remaining conditions are readily verified with the matrices of (12.4.6). Thus, M' is almost regular. □

The surprising fact is that restricted ΔY extension sequences create all almost regular matroids from the two matroids given by B^7 and B^{11} of (12.4.5). The precise statement is given in Theorem (4.4.16), which we repeat here.

(12.4.8) Theorem. *The class of almost regular matroids has a partition into two subclasses. One of the subclasses consists of the almost regular matroids producible by ΔY extension sequences from the matroid represented by B^7 of (12.4.5). The other subclass is analogously generated by B^{11} of (12.4.5). There is a polynomial algorithm that obtains an appropriate ΔY extension sequence for creating any almost regular matroid from the matroid of B^7 or B^{11}, whichever applies.*

Proof. The existing proof is so long that we cannot include it here. Nevertheless, we sketch the main arguments, since they involve interesting matroid decomposition ideas involving the matroids R_{10} and R_{12} of Sections 10.2 and 11.3.

We take M to be a minimal almost regular matroid that cannot be produced by restricted ΔY extension sequences from any one of the two matroids given by (12.4.5). Put differently, M cannot be reduced to one of the two matroids by restricted ΔY extension sequences. For short, we say that M is not reducible.

By the minimality, M obviously cannot have series or parallel elements. In addition, it turns out that M cannot have a 3-separation with at least four elements on each side. Indeed, such a 3-separation implies a 3-sum decomposition of M where one component is a wheel. That 3-sum is readily proved to be reducible, a contradiction of the minimality of M.

Suppose M does not have R_{10} or R_{12} minors. First, one can show that M must have a "*del*" element z such that $M\backslash z$ is graphic. Thus, M can be represented by a graph plus T set. Rather complicated arguments then prove M to be isomorphic to F_7 or F_7^* with appropriate labels. A single ΔY exchange transforms the F_7^* case to F_7.

Next, we assume that M has an R_{10} minor. Rather easily, we reach the conclusion that M is isomorphic to the matroid of B^{11} or its dual. The second case can be transformed to the first one by one ΔY exchange. The proof relies on the graph plus T set representation of Section 10.2. That approach also produces the insight, at present irrelevant but later useful, that any almost regular matroid with an R_{10} minor must be labeled in such a way that the "*con*" elements do not form a base and the "*del*" elements do not form a cobase.

One case remains, where M has an R_{12} minor but no R_{10} minors. By duality and the symmetry of R_{12}, we may assume that M has a "*del*" element z such that $M\backslash z$ has an R_{12} minor. To investigate this case, we apply the recursive decomposition algorithm of Section 10.5, starting with $\mathcal{H} = \{R_{12}\}$. The class of matroids under consideration is $\mathcal{M}_2$, defined in Steps 3 and 4 at the beginning of this section. In the first iteration of the decomposition algorithm, we use the 3-separation of R_{12} given by (11.3.11). We find exactly two matroids that prevent induced 3-separations. They are duals of each other. Let V_{13} be one of them. Thus, after the first iteration, we have $\mathcal{H}' = \{V_{13}, V_{13}^*\}$, which is the set $\mathcal{H}$ for the next iteration. There we see that both V_{13} and V_{13}^* induce certain 4-sum decompositions. We conclude the second iteration with $\mathcal{H}' = \emptyset$, and stop the decomposition algorithm.

We utilize the output of the decomposition algorithm as follows. As argued at the beginning of the proof, M cannot have a 3-separation with at least four elements on each side. In particular, the 3-separation of any R_{12} minor, as given by (11.3.11), cannot induce a 3-separation of M. By the results of the decomposition algorithm, M then has a V_{13} or V_{13}^* minor.

By duality, we only need to pursue the case where M has a V_{13} minor. We prove that one component of a certain 4-sum induced by V_{13} is almost regular and is represented by a graph plus T set with special structure. In fact, that graph can be created by a particular ΔY extension sequence. For that sequence, the last SP extension step plus the subsequent ΔY exchanges can be viewed as final steps of a construction of M, thus contradicting the minimality of M. □

Later we need the following observation made in the preceding proof.

(12.4.9) Lemma. *Let M be an almost regular matroid. Then M has an R_{10} minor if and only if it is produced from the almost regular matroid of B^{11} by some restricted ΔY extension sequence. Furthermore, if M has an R_{10} minor, then the set of "con" elements does not form a base of M, and the set of "del" elements does not form a cobase.*

At this point, it probably is not apparent how the matrix classes $\mathcal{U}$,

$\mathcal{A}$, $\mathcal{B}$, $\mathcal{V}$ of Section 12.3 are related to the class of almost regular matroids. The next section will quickly change that situation.

12.5 Matrix Constructions

Let us pause for a moment to assess our position. In Section 12.3, we have defined the matrix classes $\mathcal{U}$, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{V}$. We have seen that knowledge of just one of the classes $\mathcal{U}$, $\mathcal{A}$, or $\mathcal{B}$ permits easy construction of all others. In Section 12.4, we have established an elementary procedure for creating the almost regular matroids from two initial matroids given by B^7 and B^{11} of (12.4.5). In this section, we use these results to determine elementary constructions for $\mathcal{U}$, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{V}$. From $\mathcal{B}$ or $\mathcal{V}$, we obtain the class $\mathcal{N}_3$ of Theorem (12.2.16). Thus, we solve the characterization problem of the minimal binary nonregular matrices.

We could proceed in several ways. Particularly appealing appears to be the following route. We first tie a subclass of $\mathcal{U}$ to the almost regular matroids. Then we identify the remaining members of $\mathcal{U}$. Finally, we construct $\mathcal{U}$, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{V}$, in that order. The first step is accomplished by the following lemma.

(12.5.1) Lemma. *A binary matroid M is almost regular if and only if M is represented by a binary nonregular matrix of the form*

(12.5.2)

$$\tilde{B} = \begin{array}{c|c|c|} & \begin{matrix} x \\ del \end{matrix} & \begin{matrix} Y \\ con \end{matrix} \\ \hline y\ \ con & 0 & \underline{1}^t \\ \hline X\ \ del & \underline{1} & U \\ \hline \end{array}$$

Matrix $\tilde{B}$ for almost regular matroid M

where U when viewed over $\mathbb{R}$ *is complement totally unimodular.*

Proof. We start with the "only if" part. Let M be almost regular. According to some arbitrary choice, partition the set of "*del*" elements of M into a singleton set $\{x\}$ and a remainder X. Similarly, partition the set of "*con*" elements into $\{y\}$ and Y. By Lemma (12.4.2), $X \cup \{x\}$ is a circuit and hyperplane of M, and $Y \cup \{y\}$ is a cocircuit and cohyperplane. These facts imply $X \cup \{y\}$ to be a base of M, and the corresponding representation matrix to be $\tilde{B}$ of (12.5.2) for some $\{0, 1\}$ matrix U.

We must show U to be complement totally unimodular. Since element x of M has a "*del*" label, $M \backslash x$ is regular. Thus, the column submatrix of $\tilde{B}$ indexed by Y is regular. Indeed, because of the 1s in row y of that submatrix, the matrix U when viewed as real must be totally unimodular.

By GF(2)-pivots in column x or row y of $\tilde{B}$, we can transform U to all matrices obtainable from U by complement operations. As just argued, each such matrix when viewed over $\mathbb{R}$ must be totally unimodular. Thus, U is complement totally unimodular.

For proof of the "if" part, we reverse the above arguments. Thus, according to the nonregular matrix $\tilde{B}$ and the complement totally unimodular matrix U, the matroid M is nonregular, and each of its "*con*" (resp. "*del*") labels indicates regularity upon a contraction (resp. deletion). The parity condition is easily confirmed for the fundamental circuits and cocircuits displayed by $\tilde{B}$. Thus, that condition holds for all circuits and cocircuits. Since $\tilde{B}$ is nonregular, it must have at least three rows and at least three columns. Thus, the existence condition on "*con*" and "*del*" labels is satisfied. We conclude that M is almost regular. □

Recall from Section 12.3 that the class $\mathcal{B}$ consists of the binary matrices B of the form

(12.5.3)

$$B = \begin{array}{c|c|c|} & x & Y \\ \hline y & 0 & \underline{1}^t \\ \hline X & \underline{1} & U \\ \hline \end{array}$$

Matrix B of class $\mathcal{B}$

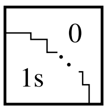
where U when viewed over $\mathbb{R}$ is complement totally unimodular. Lemma (12.5.1) thus implies the following result for $\mathcal{B}$.

(12.5.4) Corollary. *The representation matrices $\tilde{B}$ of* (12.5.2) *of the almost regular matroids become upon removal of labels precisely the nonregular matrices of $\mathcal{B}$.*

Proof. This follows directly from Lemma (12.5.1) and a comparison of $\tilde{B}$ of (12.5.2) with B of (12.5.3). □

By Lemma (12.5.1) and Corollary (12.5.4), we have a characterization of the subclass of $\mathcal{U}$ whose members produce the nonregular matrices of $\mathcal{B}$. The next lemma establishes the structure of the remaining members of $\mathcal{U}$, i.e., those generating the regular matrices of $\mathcal{B}$. To this end, we define any nonempty square triangular matrix U satisfying for all $i \geq j$, $U_{ij} = 1$, to be *solid triangular*. When we add parallel or zero vectors any number of times to such a matrix, we get a *solid staircase* matrix. A typical example of such a matrix is given below.

(12.5.5)



Solid staircase matrix

We need an auxiliary result about solid staircase matrices.

(12.5.6) Lemma. *A $\{0,1\}$ matrix is a solid staircase matrix if and only if it has no 2×2 identity as submatrix.*

Proof. The "only if" part is elementary. The "if" part is proved by a straightforward inductive argument. One removes a row with maximum number of 1s, invokes induction, then adds that row again for the desired conclusion. □

Here is the promised characterization of the regular matrices of $\mathcal{B}$.

(12.5.7) Lemma. *A matrix B of $\mathcal{B}$ as given by (12.5.3) is regular if and only if the submatrix U of B is a zero matrix or solid staircase matrix, or becomes a matrix of the latter type by some complement steps.*

Proof. We start with the "if" part. We must show that any B of (12.5.3) with U as specified is regular. Evidently, this is so when U is a zero matrix. So assume that by some complement steps, U becomes a solid staircase matrix. According to the proof of Lemma (12.5.1), any such steps correspond to GF(2)-pivots in B. Thus, they maintain regularity. Series and parallel extensions of a binary matroid also retain regularity. Thus, we may further assume that B has no parallel or unit vectors. By (12.5.3), U then has no parallel or zero vectors. Thus, U is solid triangular, and B is

(12.5.8)
$$B = \begin{array}{c|c|c|} & x & Y \\ \hline y & 0 & \underline{1}^t \\ \hline X & \underline{1} & \begin{matrix} 1 & \ddots & \\ 1\text{s} & & 1 \end{matrix} \\ \hline \end{array}$$

Matrix B with solid triangular U

We claim that B represents the graphic matroid of a wheel, with $X \cup \{x\}$ as set of rim edges, and $Y \cup \{y\}$ as set of spokes. The edges x and y are such that $X \cup \{y\}$ is a path. The claim is easily verified. One only checks that the fundamental cycles for $X \cup \{y\}$ are displayed by B of (12.5.8). We conclude that B is regular.

For proof of the "only if" part, we assume B of (12.5.3) to be regular. Since B is in $\mathcal{B}$, the submatrix U when viewed over $\mathbb{R}$ is complement totally unimodular. We are done if U is a zero matrix. So assume U to be nonzero. We may assume that U has no zero or parallel vectors. Thus, we must show U or a matrix obtained from U by complement steps, to be solid triangular.

If U has a 3×3 identity submatrix, say indexed by $\overline{X} \subseteq X$ and $\overline{Y} \subseteq Y$, then B contains the matrix of (12.5.9) below. A pivot on any 1 of the 3×3 identity submatrix, plus a pivot in the resulting matrix on the 1 in the (x, y) position, produces a matrix that displays an F_7 minor. But then B

is not regular, a contradiction.

(12.5.9)
$$\begin{array}{c|c|ccc|}
 & x & & \overline{Y} & \\
\hline
y & 0 & 1 & 1 & 1 \\
\hline
 & 1 & 1 & 0 & 0 \\
\overline{X} & 1 & 0 & 1 & 0 \\
 & 1 & 0 & 0 & 1 \\
\hline
\end{array}$$

Nonregular submatrix of B induced
by a 3×3 identity submatrix of U

Suppose U has a 2×2 identity submatrix. If U is connected, then U or its transpose contains the submatrix $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, say indexed by $\overline{X} \subseteq X$ and $\overline{Y} \subseteq Y$. By duality, we may assume the former case. Then B contains the matrix

(12.5.10)
$$\begin{array}{c|c|ccc|}
 & x & & \overline{Y} & \\
\hline
y & 0 & 1 & 1 & 1 \\
\hline
\overline{X} & 1 & 1 & 0 & 1 \\
 & 1 & 0 & 1 & 1 \\
\hline
\end{array}$$

Nonregular submatrix of B induced
by a certain 2×3 submatrix of U

which represents an F^7 minor, and once more we have a contradiction.

Still assume that U has a 2×2 identity submatrix. As argued above, U does not contain a 3×3 identity submatrix and is not connected. Thus, U has exactly two connected blocks, neither of which contains a 2×2 identity submatrix. By Lemma (12.5.6) and by the absence of zero or parallel vectors from U, each block must be solid triangular. Let z index a row of U with maximum number of 1s. The row z complement of U is then solid triangular, as desired.

Finally, assume that U has no 2×2 identity submatrix. By Lemma (12.5.6) and by the absence of zero or parallel vectors, U is then solid triangular. □

We are ready to state and validate the following construction of the complement totally unimodular matrices. The construction relies on four seemingly strange matrices. Their origin will become clear shortly.

(12.5.11) Construction of $\mathcal{U}$. *Define real $\{0,1\}$ matrices U^0, U^1, U^7, and U^{11} as follows.*

(12.5.12)
$$U^0 = \boxed{0}\,; \quad U^1 = \boxed{1}\,; \quad U^7 = \begin{array}{|ccc|} \hline 1 & 0 & 1 \\ 0 & 1 & 1 \\ \hline \end{array}\,; \quad U^{11} = \begin{array}{|ccccc|} \hline 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ \hline \end{array}$$

Complement totally unimodular
matrices U^0, U^1, U^7, and U^{11}

Starting with $U = U^0, U^1, U^7$, *or* U^{11}, *apply a sequence of operations each of which is one of* (i), (ii), *or* (iii) *below.*

(i) *Perform a row or column complement operation.*
(ii) *Add a zero or parallel row or column vector.*
(iii) (ΔY exchange) *If* U *is one of the two matrices below, replace* U *by the other matrix.*

(12.5.13)
$$\begin{array}{|c|c|c|} \hline \bar{U} & b & b \\ \hline a & 1 & 0 \\ \hline \end{array} \quad \longleftrightarrow \quad \begin{array}{|c|c|} \hline \bar{U} & b \\ \hline a & 1 \\ \hline a & 0 \\ \hline \end{array}$$

ΔY exchange for complement
totally unimodular matrices

Collect in sets $\mathcal{U}_0$, $\mathcal{U}_1$, $\mathcal{U}_7$, *and* $\mathcal{U}_{11}$ *the matrices that can be so deduced from* U^0, U^1, U^7, *and* U^{11}, *respectively. These sets form a partition of* $\mathcal{U}$.

An example matrix constructed from U^7 is given by (12.5.25) below.

Proof of Construction (12.5.11). It is easy to see that the steps (i)–(iii) produce from $U^0 = [\,0\,]$ the class of zero matrices, which thus constitutes $\mathcal{U}_0$. Indeed, steps (i) and (iii) never apply. For validation of the remaining cases, we assign a "*con*" (resp. "*del*") label to each column (resp. row) of any U produced by Construction (12.5.11) from U^1, U^7, or U^{11}. For the moment, these labels are purely formal. Next, we embed U into the matrix $\tilde{B}$ of (12.5.2), which we repeat here.

(12.5.14)
$$\tilde{B} = \begin{array}{cc|c|c|} & & \begin{array}{c} x \\ del \end{array} & \begin{array}{c} Y \\ con \end{array} \\ \hline y & con & 0 & \underline{1}^t \\ \hline X & del & \underline{1} & U \\ \hline \end{array}$$

Matrix $\tilde{B}$ derived from matrix U

Steps (i)–(iii) of Construction (12.5.11) are then equivalent to the following operations on $\tilde{B}$. The complement-taking of step (i) corresponds to a GF(2)-pivot in row y or column x of $\tilde{B}$. The addition of zero or parallel vectors of step (ii) is the addition of parallel vectors or of unit vectors with 1 in row y or column x. Finally, the exchange of step (iii) is the ΔY exchange depicted by (12.4.6).

We take these observations one step further. Let M be the labeled matroid represented by $\tilde{B}$. We claim that the operations just specified for $\tilde{B}$ are equivalent to the operations of restricted ΔY extension sequences in M. Indeed, the addition of parallel vectors or of unit vectors with 1 in row y or column x of $\tilde{B}$ becomes the restricted SP extension for M. The GF(2)-pivot in $\tilde{B}$ plus the exchange given by (12.4.6) are equivalent to a restricted ΔY exchange in M.

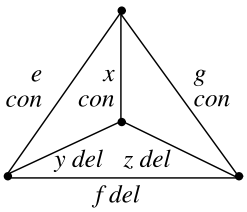
We interpret the matrices U of $\mathcal{U}_1$, $\mathcal{U}_7$, and $\mathcal{U}_{11}$ in terms of the related matroid M. We start with $\mathcal{U}_1$. The matrix $\tilde{B}$ of (12.5.14) produced by $U = U^1 = [\,1\,]$ is

(12.5.15)
$$\begin{array}{r|cc|}
 & \begin{smallmatrix} x \\ d \\ e \\ l \end{smallmatrix} & \begin{smallmatrix} Y \\ c \\ o \\ n \end{smallmatrix} \\ \hline
y \ con & 0 & 1 \\
X \ del & 1 & 1 \\ \hline
\end{array}$$

Matrix $\tilde{B}$ derived from $U = U^1 = [\,1\,]$

The corresponding M is the graphic matroid of the wheel with two spokes. Indeed, the spokes are labeled "*con*" and form the set $Y \cup \{y\}$. The rim edges are labeled "*del*" and constitute the set $X \cup \{x\}$. We interpret restricted ΔY extension sequences for M as sequences of graph operations applied to the preceding graph. Each such operation is either a subdivision of a "*del*" edge into two "*del*" edges, or an addition of a "*con*" edge parallel to a "*con*" edge, or an exchange of a triangle by a 3-star, or an exchange of a 3-star by a triangle. In the latter two operations, the labels are assigned according to (4.4.14), which we repeat below.

(12.5.16)



Triangle and 3-star with labels

We leave it to the reader to verify that any graph produced by these operations is obtainable from some wheel graph with at least two spokes by subdivision of rim edges and addition of edges parallel to spokes. In terms of M, the rim edges of such an extended wheel form the set $X \cup \{x\}$, and the spokes constitute $Y \cup \{y\}$. We interpret this result in terms of U as in the proof of Lemma (12.5.7). Thus, we see that U or some matrix obtained from U by complement operations, is a solid staircase matrix.

We turn to the cases of U^7 and U^{11}. The matrix U^7 (resp. U^{11}) produces as $\tilde{B}$ the matrix B^7 (resp. B^{11}) of (12.4.5). In either case, $\tilde{B}$ is nonregular, and M is an almost regular matroid. By Theorem (12.4.8), restricted ΔY extension sequences produce all almost regular matroids from the two matroids represented by B^7 and B^{11}. By Lemma (12.5.1) and Corollary (12.5.4), the sets $\mathcal{U}_7$ and $\mathcal{U}_{11}$ thus contain the complement totally unimodular matrices that produce the nonregular members of the class $\mathcal{B}$.

By Lemma (12.5.7) and the above characterization of $\mathcal{U}_0$ and $\mathcal{U}_1$, the class $\mathcal{U}_0 \cup \mathcal{U}_1$ contains precisely the complement totally unimodular matrices U that produce the regular matrices B of $\mathcal{B}$. We have also seen that the class $\mathcal{U}_7 \cup \mathcal{U}_{11}$ generates the nonregular matrices B of $\mathcal{B}$. Thus, $\mathcal{U}_0 \cup \mathcal{U}_1 \cup \mathcal{U}_7 \cup \mathcal{U}_{11}$ is equal to $\mathcal{U}$. The sets $\mathcal{U}_0 \cup \mathcal{U}_1$ and $\mathcal{U}_7 \cup \mathcal{U}_{11}$ are necessarily disjoint. Evidently, this is also so for $\mathcal{U}_0$ and $\mathcal{U}_1$. Theorem (12.4.8) says that any almost regular matroid can be generated from exactly one of the matroids represented by B^7 and B^{11}. Thus, $\mathcal{U}_7$ and $\mathcal{U}_{11}$ are disjoint. We conclude that $\mathcal{U}_0$, $\mathcal{U}_1$, $\mathcal{U}_7$, and $\mathcal{U}_{11}$ form a partition of $\mathcal{U}$ as claimed by Construction (12.5.11). □

Construction (12.5.11) has two interesting corollaries.

(12.5.17) Corollary. *Let U be a nonempty complement totally unimodular matrix without zero or parallel vectors. Then U, or some matrix obtainable from U by complement operations, contains a unit vector.*

Proof. We use the notation of the proof of Construction (12.5.11). By the assumptions, U is in $\mathcal{U}_1 \cup \mathcal{U}_7 \cup \mathcal{U}_{11}$. Let U define $\tilde{B}$ of (12.5.14), and M be the labeled matroid represented by $\tilde{B}$. Now U has no zero or parallel vectors. Thus, M has no series or parallel elements. M is obtained from the graphic matroid of (12.5.15) or from the almost regular matroid of B^7 or B^{11} by a restricted ΔY extension sequence. In all three cases, the initial matroid has a triangle with two "*con*" labels and one "*del*" label, and the final matroid has a triangle with two "*con*" labels or triad with two "*del*" labels. By duality, we may assume M to have a triangle C with two "*con*" labels. In the notation for $\tilde{B}$ of (12.5.14), assume $x \notin C$ and $y \in C$. Thus, y is one of the two "*con*" elements of C. The second "*con*" element of C must be in Y. The third element of C, with "*del*" label, must be in X. Then column $z \in Y$ of $\tilde{B}$ has exactly two 1s, one of which is in row y. Thus, column z of U is a unit vector. If $x \in C$ or $y \notin C$, we can achieve the desired configuration by GF(2)-pivots in column x or row y of $\tilde{B}$. The pivots correspond to complement operations for U. Thus, the corollary holds in all cases. □

(12.5.18) Corollary. *If $U \in \mathcal{U}$ is $\mathbb{R}$-nonsingular, then U is in $\mathcal{U}_1$ or $\mathcal{U}_7$, but not in $\mathcal{U}_0$ or $\mathcal{U}_{11}$.*

Proof. Candidate classes are the claimed ones and $\mathcal{U}_{11}$. Let $U \in \mathcal{U}_{11}$, and let M be the matroid defined via $\tilde{B}$ of (12.5.14). By the proof of Construction (12.5.11), M is an almost regular matroid produced by some restricted ΔY extension sequence from the matroid represented by B^{11}. Lemma (12.4.9) says that the set $Y \cup \{y\}$ of "*con*" elements of M does not form a base, and the set $X \cup \{x\}$ of "*del*" elements does not form a cobase. By assumption, U is $\mathbb{R}$-nonsingular. Since U is complement totally unimodular, U when viewed as binary is GF(2)-nonsingular as well. But then, by $\tilde{B}$ of (12.5.14), the set $Y \cup \{y\}$ is a base of M, a contradiction. □

At this point, it is a simple matter to construct the classes $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{V}$. For completeness, we include details. We start with $\mathcal{A}$ and $\mathcal{B}$. Recall that $\mathcal{A}$ (resp. $\mathcal{B}$) is the class of almost representative matrices over GF(3) (resp. GF(2)).

(12.5.19) Construction of $\mathcal{A}$ and $\mathcal{B}$. *The classes $\mathcal{A}$ and $\mathcal{B}$ are deduced from $\mathcal{U}$ as follows. Each matrix of $\mathcal{A}$ (resp. $\mathcal{B}$) is precisely a matrix of the form*

(12.5.20)
$$\begin{array}{|c|c|} \hline \alpha & \underline{1}^t \\ \hline \underline{1} & U \\ \hline \end{array}$$

Matrix of $\mathcal{A}$ or $\mathcal{B}$

For $\mathcal{A}$ (resp. $\mathcal{B}$), the matrix is over GF(3) *(resp.* GF(2)*), with $\alpha = -1$ (resp. $\alpha = 0$). The submatrix U is a $\{0,1\}$ matrix that, when considered over $\mathbb{R}$, is in $\mathcal{U}$.*

Proof. Validity of the construction follows directly from the definition of $\mathcal{A}$ and $\mathcal{B}$ following Lemma (12.3.10). □

Next we give a construction for $\mathcal{V}$, the class of minimal violation matrices of total unimodularity.

(12.5.21) Construction of $\mathcal{V}$. *The class $\mathcal{V}$ is deduced from $\mathcal{U}$ as follows. Each matrix V of $\mathcal{V}$ is up to scaling by $\{\pm 1\}$ factors of the form*

(12.5.22)
$$\begin{array}{|c|c|} \hline \alpha & a \\ \hline b & U^{-1} \\ \hline \end{array} \; ; \quad \begin{aligned} a &= \underline{1}^t \cdot U^{-1} \\ b &= U^{-1} \cdot \underline{1} \\ \alpha &= \underline{1}^t \cdot U^{-1} \cdot \underline{1} - 2 \end{aligned}$$

Matrix V up to scaling

The matrix U is a square $\mathbb{R}$-nonsingular matrix of $\mathcal{U}_1$ or $\mathcal{U}_7$. If $U \in \mathcal{U}_1$, then V has exactly two nonzeros in each row and in each column; indeed, the bipartite graph BG(V) *is a cycle. If $U \in \mathcal{U}_7$, then V has at least four nonzeros in some row and in some column.*

Proof. Validity of the construction follows from Corollaries (12.3.12) and (12.5.18). □

Corollary (12.5.17) and Construction (12.5.21) yield an interesting result.

(12.5.23) Corollary. *Every matrix $V \in \mathcal{V}$ has a row or column with exactly two nonzeros.*

Proof. The inverse of V is, up to scaling by $\{\pm 1\}$ factors, given by $\tilde{V}$ of (12.3.17). The U^{-1} occurring in V of (12.5.22) is the inverse of the U defining $\tilde{V}$ of (12.3.17). A simple check confirms that replacement of U by a matrix deduced from U by complement operations effectively may be viewed as scaling in $\tilde{V}$. Thus, up to scaling, U and all matrices obtainable from U by complement operations produce the same V. Since U is nonsingular, it cannot contain zero or parallel vectors. By Corollary (12.5.17), U, or some matrix deduced by complement operations from U, contains a unit vector. Let U' be that matrix. The inverse of U' contains a unit vector as well. By (12.5.22) and the above discussion, V must have a row or column with exactly two nonzeros. $\square$

At long last, we can fill the gap left by Theorem (12.2.16) and complete the characterization of minimal nonregular submatrices. The desired theorem is as follows.

(12.5.24) Theorem. *Let $\mathcal{N}$ be the class of binary minimal violation matrices of regularity. Then $\mathcal{N}$ has a partition into three subclasses $\mathcal{N}_1$, $\mathcal{N}_2$, $\mathcal{N}_3$ as follows.*

(a) *$\mathcal{N}_1$ (resp. $\mathcal{N}_2$) is the set of binary matrices B for which $\mathrm{BG}(B)$ is a graph of type H_1 (resp. H_2) of* (12.2.14).
(b) *$\mathcal{N}_3$ is the set of binary support matrices of $V \in \mathcal{V}$ produced via* (12.5.22) *from the $\mathbb{R}$-nonsingular matrices $U \in \mathcal{U}_7$.*

Proof. Part (a) is taken from Theorem (12.2.16). Part (b) follows from that theorem and Construction (12.5.21) of $\mathcal{V}$. $\square$

The above constructions are readily performed by hand. That way one can rapidly produce structurally interesting matrices. An example of a complement totally unimodular matrix U and a minimal violation matrix V of total unimodularity generated that way is as follows. We first obtain by Construction (12.5.11) the following complement totally unimodular matrix U.

$$(12.5.25) \qquad U = \begin{array}{c|ccc|ccc|} & z & i & b & f & e & c \\ \hline a & 1 & 1 & 1 & 1 & 1 & 1 \\ g & 0 & 1 & 0 & 1 & 1 & 1 \\ h & 0 & 0 & 1 & 1 & 1 & 1 \\ \hline j & 1 & 0 & 1 & 1 & 0 & 1 \\ k & 0 & 0 & 0 & 0 & 1 & 1 \\ d & 1 & 0 & 1 & 0 & 0 & 1 \\ \hline \end{array}$$

Complement totally unimodular matrix U
representing R_{12}

The labels may be used to verify that U represents the regular matroid R_{12} of (10.2.9). One only needs to check via the graph plus T set of (10.2.9) the fundamental circuits for the base $\{a, d, g, h, j, k\}$, which is the index set of the rows of U. The indicated partition of U corresponds to the 3-separation $(\{a, b, g, h, i, z\}, \{c, d, e, f, j, k\})$ of R_{12}. Straightforward computations confirm that U is $\mathbb{R}$-nonsingular, and that up to scaling by $\{\pm 1\}$ factors and row and column exchanges, U is its own inverse. Thus, for this special case, we are tempted to use U instead of U^{-1} in the formula for V of (12.5.22). But U^{-1} is a scaled and permuted version of U. Because of the scaling, the formula (12.5.22) cannot be used. But we may rely on part (b) of Corollary (12.3.12). According to that result, V is Eulerian and its entries sum in $\mathbb{R}$ to $2 (\mathrm{mod}\, 4)$. Using these two facts, we determine the following V from U.

(12.5.26)

$$V = \begin{array}{c|c|cccccc|} & & z & i & b & f & e & c \\ \hline & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline a & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ g & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ h & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ j & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ k & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ d & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ \hline \end{array}$$

Matrix V deduced from U of (12.5.25)

The above example supports the following result.

(12.5.27) Lemma. *The regular matroid R_{12} has a real nonsingular $\{0, 1\}$ representation matrix that is complement totally unimodular.*

Construction (12.5.11) and the matrix V of (12.5.26) permit us to relate $\mathcal{V}$ to the regular matroids R_{10} and R_{12} of (10.2.8) and (10.2.9), as follows.

(12.5.28) Lemma. *Let M be the binary matroid represented by the binary support matrix W of a minimal non-totally unimodular matrix $V \in \mathcal{V}$. Then M does not have R_{10} minors, but may have an R_{12} minor.*

Proof. By Theorem (12.3.11), M is represented not just by W as stated, but also by some $B \in \mathcal{B}$ given by (12.5.20). Indeed, the submatrix U^{-1} of V of (12.5.22) is the inverse of the matrix U defining B. By Construction (12.5.21) for $\mathcal{V}$, we have $U \in \mathcal{U}_1 \cup \mathcal{U}_7$. Thus, $U \notin \mathcal{U}_{11}$. By the proof of Construction (12.5.11) for $\mathcal{U}$, $U \notin \mathcal{U}_{11}$ implies that M has no R_{10} minor. The matrix V of (12.5.26) demonstrates that M may have an R_{12} minor. $\square$

Lemma (12.5.28) implies the claim made toward the end of Section 12.3 that none of the classes $\mathcal{U}$, $\mathcal{A}$, and $\mathcal{B}$ can be produced from $\mathcal{V}$ by the matrix operations of scaling by $\{\pm 1\}$ factors, pivots, submatrix taking, and change of fields from $\mathbb{R}$ to GF(2) or GF(3). In particular, the matrix U^{11} of (12.5.12) cannot be obtained that way.

In the last section, we discuss applications and extensions, and include references.

12.6 Applications, Extensions, and References

The entire Section 12.2 is taken from Truemper (1982b). Variations of the graph signing problem of Section 12.2 are treated in Zaslavsky (1981a), (1981b), (1982), (1987), (1989), (1991). The notion of α-balancedness extends the concept of $\{0,1\}$ balanced matrices covered extensively elsewhere (see Berge (1972), (1973), (1980), Fulkerson, Hoffman, and Oppenheim (1974), and Anstee and Farber (1984)). Profound decomposition results for balanced matrices are proved in Conforti and Rao (1989), (1992a)–(1992d), Conforti and Cornuéjols (1990), and Conforti, Cornuéjols, and Rao (1997). In Truemper and Chandrasekaran (1978), balancedness of $\{0,1\}$ matrices is tied to total unimodularity by the exclusion of certain minimal violation matrices of total unimodularity. That result motivated the search for a construction of $\mathcal{V}$.

A $\{0,1\}$ matrix is *perfect* if the polyhedron $\{x \mid A \cdot x \leq \underline{1}, x \geq 0\}$ has only integer vertices. A graph is *perfect* if its clique/node incidence matrix is perfect. Any $\{0,1\}$ balanced matrix is perfect, but the converse does not hold. The so-called strong perfect graph conjecture says that the minimal nonperfect graphs, and thus the minimal nonperfect matrices, have a certain simple structure (see, e.g., Padberg (1974)). There are numerous partial results with regard to that conjecture. Approaches based on graph decomposition have been cited in Section 10.7. Padberg (1974) contains a very interesting matrix-based attack on the problem.

The concept of complement total unimodularity of Section 12.3 is defined in Truemper (1980b). Almost representative matrices are characterized in Truemper (1982b). A number of papers explicitly or implicitly include properties of the class $\mathcal{V}$ (see Ghouila-Houri (1962), Camion (1963a), (1963b), (1965), Chandrasekaran (1969), Gondran (1973), Padberg (1975), (1976), Tamir (1976), Kress and Tamir (1980), Truemper (1977), (1978), (1980b), (1982b), and de Werra (1981)). But none of the results captures the complexity of $\mathcal{V}$. Indeed, the cited results do not contain a single clue to how one might construct even a small subset of structurally different matrices of $\mathcal{V}$.

Section 12.4 relies on Truemper (1992a), (1992b). The analysis technique applies not just to the cited properties, but also to certain representability questions. An important case is treated in Truemper (1982b), where the minimal violation matrices of abstract matrices not representable over GF(2) or GF(3) are characterized.

Section 12.5 is entirely based on Truemper (1992b). That reference also contains alternate constructions using pivots. The constructions do not have descriptions as brief as the ones included here, but they are well suited for hand calculations. Truemper (1992b) also includes a proof that every 3-connected almost regular matroid different from F_7 and F_7^* has a binary representation matrix that is balanced. An example is the matrix V of (12.5.26). The proof relies on an extension of an efficient algorithm of Fonlupt and Raco (1984) that proves existence of a balanced binary representation matrix for any regular matroid. The algorithm is a modification of a scheme due to Camion (1968) where the latter existence result was first established.