A Simple Algorithm that Proves Half-Integrality of Bidirected Network Programming

Ethan D. Bolker
Departments of Mathematics and Computer Science,
University of Massachusetts at Boston,
Boston, MA 02125-3393, U.S.A.

and

Thomas Zaslavsky
Department of Mathematical Sciences,
Binghamton University (SUNY),
Binghamton, NY 13902-6000, U.S.A.

Version of April 22, 2006.

Abstract. In a bidirected graph, each end of each edge is independently oriented. We show how to express any column of the incidence matrix as a half-integral linear combination of any column basis, through a simplification, based on an idea of Bolker, of a combinatorial algorithm of Appa and Kotnyek. Corollaries are that the inverse of each nonsingular square submatrix has entries 0, $\pm\frac{1}{2}$, and ± 1, and that a bidirected integral linear program has half-integral solutions.

Mathematics Subject Classifications (2000). Primary 05C22, Secondary 05B35, 05C20, 90C10.

Key words. Bidirected graph, incidence matrix, half-integrality, signed-graph matroid.

A bidirected graph B is a graph in which every edge has an independent direction at each endpoint. The node-edge incidence matrix $\mathrm{H}(B)$ generalizes the incidence matrix of an ordinary directed graph G. Every nonsingular square minor of a graphical incidence matrix has determinant equal to $+1$ or -1; this property is the basis of the theory of network matrices. A *network matrix* is obtained from a graphic incidence matrix by deleting dependent rows (call the result $\bar{\mathrm{H}}(G)$), choosing a maximal forest T (that is, a basis of the column space of $\mathrm{H}(G)$), premultiplying by the inverse of the square submatrix $\bar{\mathrm{H}}(G,T)$ indexed by the columns corresponding to T, and deleting those columns. Since $\det \bar{\mathrm{H}}(G,T) = \pm 1$, $\bar{\mathrm{H}}(G,T)^{-1}$ is integral and therefore so is any network matrix. Appa and Kotnyek [1] generalized this idea to bidirected graphs. An essential lemma for their work is that the submatrix $\bar{\mathrm{H}}(B,T)$ indexed by a basis of the column space of $\bar{\mathrm{H}}(B)$ has an inverse that is half-integral; thus they improve on [8, Lemma 8A.2], which showed the weaker fact that $\det \bar{\mathrm{H}}(B,T)$ is a signed power of 2. Appa and Kotnyek provide an algorithm [1, Algorithm 1] that proves the half-integrality in a constructive way. Here we give a similar but simpler algorithm, which is implicit in an insight of Bolker [3, 4], and was first published (in a more complicated form) by Bouchet [5, proof of Corollary 2.3]. The algorithm permits simplification of other parts of [1], but we do not discuss that in depth.

First we state precise definitions. A graph G with n nodes $v_1, \ldots, v_n$ and m edges $e_1, \ldots, e_m$ may have four kinds of edges. A *link* is an edge with two distinct endpoints; a *loop* has two coinciding endpoints; a *halfedge* has one endpoint, and a *loose edge* has no endpoints. A *circle* is a connected, 2-regular edge set; a loop is a circle of length 1. A *signed graph* is a graph together with a *signature*,

σ, that assigns to each link or loop e a sign $\sigma(e) \in \{+1, -1\}$. The sign of a circle is the product of the signs of its edges.

We indicate the bidirection in a bidirected graph B by a function $\eta : I \to \{+1, -1\}$, where I is the set of endpoint-edge pairs (v_i, e_j); one can think of the value $+1$ as indicating that the edge is directed into the node, -1 as indicating direction away from the node. The *incidence matrix* is the $n \times m$ matrix $\mathrm{H}(B)$ whose (i, j) entry is $\eta(v_i, e_j)$, except that it is 0 if e_j is a positive loop or not incident with v_i and it is $2\eta(v_i, e_j) = \pm 2$ if e_j is a negative loop incident with v_i. (These rules are explained more fully in [8]. Our notation, to avoid overcomplication, is a bit sloppy since it fails to distinguish the two ends of a loop; we trust the reader will understand the meaning.) The column indexed by edge e is denoted by c_e.

A bidirected graph implies the signature $\sigma(e) := -\eta(v, e)\eta(w, e)$ where v and w are the nodes incident with e. Thus, a positive link or loop has ends which are oriented in the same direction along the edge, one end leaving its node and the other end entering its node; as this is just like an ordinary directed edge, a directed graph is the same as a bidirected all-positive signed graph. The signature is unchanged by *reorienting* an edge, which means negating the values of η (i.e., reversing the arrows) on that edge; the incidence matrix is changed in that the column of e is negated. *Switching* a node means negating all the values of η at that node; it corresponds to negating a row of the incidence matrix. *Reducing* a node v means deleting v but not any of the incident edges; instead, an incident edge loses its endpoint(s) at v, becoming a halfedge or loose edge. Reducing a node corresponds to deleting a row of the incidence matrix, just as deleting an edge corresponds to deleting a column. A *reduction* of B is a result of applying any combination of edge deletions and node reductions.

Thus, the matrix that results from negating and deleting rows and columns in $\mathrm{H}(B)$ is the incidence matrix of a bidirected graph obtained from B. Note that $\mathrm{H}(B)$ has full row rank if and only if each component of B contains a halfedge or a negative circle [8, Theorems 5.1 and 8B.1]. If $\mathrm{H}(B)$ does not have full row rank, B can be converted, by reducing one or more nodes, to B' such that the rows of $\mathrm{H}(B')$ are rows of $\mathrm{H}(B)$ and are a basis of its row space.

In a walk $n_0, f_1, n_1, \dots, f_l, n_l$, a node n_i is *consistent* if $\eta(n_i, e_i) = -\eta(n_i, e_{i+1})$. (This definition applies to n_0, with subscripts modulo l, if $n_0 = n_l$ and $l > 0$.) A *consistent orientation* of the edges of the walk is an orientation in which every node is consistent.

Lemma. *Let T be the edge set corresponding to a basis of the column space of* $\mathrm{H}(B)$. *Let e be another edge in B. Then c_e is a half-integral combination of the columns c_f for $f \in T$, the possible nonzero coefficients being* $\pm\frac{1}{2}, \pm 1, \pm 2$.

Proof. Our constructive proof begins as does Appa and Kotnyek's. We may assume (by reduction) that $\mathrm{H}(B)$ has full row rank. Since a basis matrix has full row rank, each component of T consists of a tree and one more edge that either is a halfedge or forms a negative circle with the tree. Thus, $S = T \cup \{e\}$ contains a unique circuit (minimal dependent subset) of the set of columns of $\mathrm{H}(B)$. Let $c_1, \dots, c_n$ be the columns of $\mathrm{H}(B)$ that correspond to the edges of T.

According to [8, Theorems 5.1(e) and 8B.1], a circuit has one of three forms. It may be a positive circle (or a loose edge, but we may safely ignore this trivial case), or a pair of negative circles with exactly one common node, or a pair of disjoint negative circles along with a minimal connecting path. The latter two types are called *handcuffs*. In a handcuff, either negative circle, or both, may be replaced by a halfedge.

A *minimal covering walk* of a circuit C is a walk of minimum length that covers each edge and has no endpoints. Thus, it is a closed walk if C contains no halfedge, but otherwise the walk begins and ends with a halfedge. A minimal covering walk covers each edge of a connecting path twice and each other edge exactly once, except that if one circle is replaced by a halfedge, the halfedge is also covered twice, and if both circles are replaced by halfedges, then every edge of C is covered once (indeed, C is its own minimal covering walk). A *consistent orientation* of a circuit is an orientation such that in a minimal covering walk every node is consistent. It is easy to verify that every circuit has a consistent orientation. (One may consult [9] for a detailed discussion of how to orient a signed graph. We note that in papers by Zaslavsky the word used for "consistent" is "coherent", while in [1] the corresponding term is "incoherent".)

Here is the procedure for producing c_e as a linear combination of the c_i. Let C be the unique circuit contained in $T \cup \{e\}$. First, reorient edges of C so that a minimal covering walk becomes consistently oriented. (This is independent of which minimal covering walk one chooses.) Then assign weights -1 to each singly covered edge, -2 to each doubly covered edge, and 0 to the other edges in T. Then, negate the values assigned to edges that were reoriented. Divide by 2 if necessary, and negate all signs if necessary, to ensure that e has weight -1. The edge weights on T are now the coefficients in the linear combination of the c_i that equals c_e. □

The procedure can be made more precise. We follow a suggestion of the referee. Find a minimal covering walk W. Initialize all edge weights in $T \cup \{e\}$ at 0. Trace the edges of W, reorienting each newly encountered edge so as to make W consistent. Start with e (and do not reorient it) if W is a closed walk, but otherwise start from a halfedge of W. Each time an edge is encountered, assign it weight $+1$ or -1, respectively, if its weight was zero and it was, or was not, reoriented; but double its weight if the weight was nonzero (as happens on the second encounter with an edge). When done tracing W, divide all weights by the negative of the weight of e (whose weight is -1 or -2 if W is closed, ± 1 or ± 2 otherwise). It is clear that all edges of C have been visited and assigned the right nonzero weights.

We derived our procedure from Bolker's realization that one can simply write out the linear-dependence coefficients of a signed-graph circuit (see [3, p. 160, second paragraph] and [4, proof of Theorem 7]). Zaslavsky remembered that result when he encountered this problem and turned it into our simple method. Although Bouchet stated a similar procedure, he did not apply it to the question addressed in the lemma; rather, he was interested in the dual question of nowhere-zero integral flows.

Since the rows of a submatrix M of $\mathrm{H}(B)$ are indexed by nodes and the columns by edges, the rows of M^{-1} (if it exists) are indexed by edges and the columns by nodes.

Proposition. *The inverse of any nonsingular square submatrix M of $\mathrm{H}(B)$ is half integral. Let B' be the reduction of B that corresponds to M. Then the half integers in M^{-1} are $\pm\frac{1}{2}$'s in positions (e, v) such that e lies in a circle in the component of B' that contains v. The other entries are integers 0 and ± 1.*

Proof. By reducing B we may assume that M has all the rows of $\mathrm{H}(B)$; thus it is the $n \times n$ matrix $\mathrm{H}(B, T)$ indexed by a basis $\{c_1, \dots, c_n\}$ of the column space of $\mathrm{H}(B)$, each c_i corresponding to an edge in B'. Now we replace the edge set of B by T together with a new halfedge h_i at each node v_i of B. The matrix of the new halfedges is the identity matrix, the ith column being indexed by h_i so it is the ith unit basis vector u_i. The ith column w of M^{-1} satisfies $Mw = u_i$. By the lemma, u_i is a half-integral linear combination of the columns of M. Therefore, M^{-1} is half integral. The half integers appear as stated because of the exact form of our algorithm. □

This proposition is contained in [1, Example 4 and proof of Theorem 9]. The idea of using appended halfedges is the same. The proofs are similar; ours seems simpler due to the simpler algorithm. [1, Theorem 9] points out that each row of M^{-1} is either integral (all entries belong to $\{0, \pm 1\}$) or strictly half-integral (all entries in $\{0, \pm\frac{1}{2}\}$); their proof shows, just as does our statement of the proposition, that the $\frac{1}{2}$'s appear in the rows indexed by the edges in circles of B'.

We suggest that the proofs of some other interesting results in [1], though much the same in essentials, also become more transparent by using our procedure instead of their Algorithm 1. We discuss, in particular, a main result of [1]. A *binet matrix* [1] is a matrix obtained from a bidirected-graph incidence matrix $\mathrm{H}(B)$ of full row rank by choosing a set T of edges whose columns form a basis of the column space, premultiplying $\mathrm{H}(B)$ by $\mathrm{H}(B,T)^{-1}$, and deleting the columns that correspond to T. It is the bidirected generalization of a network matrix.

Corollary ([1, Theorem 20])**.** *A consistent integral linear system $Ax = b$, in which A is a binet matrix and b is integral, has a half-integral solution. An integral linear program $Ax = b$, $x \geq 0$ with finite optimum whose coefficient matrix A is a binet matrix has a half-integral optimum.*

Proof. First, we may assume that A is the incidence matrix of a bidirected graph. If not, $A = \mathrm{H}(B,T)^{-1}\mathrm{H}(B,T^c)$, where T is a basis of the column space of $\mathrm{H}(B)$ and T^c is its complement. Then $Ax = b$ can be rewritten as $\mathrm{H}(B,T^c)x = \mathrm{H}(B,T)b$. Since the product on the right is integral, we can replace $Ax = b$ by this equation; that is, we assume A is the incidence matrix of a bidirected graph.

We may also assume A is invertible. For the proof of this, let A_i denote a row of A and b_i the corresponding entry of b. We have constraint equations $A_i x = b_i$ and, in the LP case, the nonnegativity bounds $x \geq 0$. In the case of a linear system, we discard redundant equations. In the LP case, we focus on a particular x that is an optimal vertex of the feasible region, determined by some constraint equations and some equations $x_j = 0$; it may not satisfy all constraint equations, so we discard the ones it does not satisfy as well as any redundant equations. In both cases, A remains a bidirected incidence matrix because, graph theoretically, we are only reducing vertices.

Our situation is now that x is a solution of $Ax = b$ where A has full row rank; however, x may be underdetermined by $Ax = b$. In a linear system, we specify x by setting the free variables equal to zero. In the LP case, x is already determined by having some coordinates equal to zero. If we discard the columns of A that correspond to zero coordinates of x, we have a square binet matrix A' of full row rank and remaining equations $A'_i x = b'_i$ whose constant terms are some of the entries of $\mathrm{H}(B,T)b$, hence integers.

We conclude that the original constraints can be rewritten as $A'x = b'$ where A' is an invertible incidence matrix of a bidirected graph. The corollary then follows from the proposition. □

It is easy to see (by adding a halfedge at every node) that the incidence matrix of a bidirected graph is itself a binet matrix [1]. The corollary in this special case is known (in different terminology), e.g., by [6]; see the introduction to [1, Section 6] for a discussion. The first part is also derived independently in [2, Lemma 4.7] from difficult work of Lee [7].

Acknowledgement

We thank G. Appa for his valuable suggestions.

References

[1] Gautam Appa and Balázs Kotnyek, A bidirected generalization of network matrices. *Networks* 47 (2006), 185–198.

[2] Matthias Beck and Thomas Zaslavsky, The number of nowhere-zero flows in graphs and signed graphs. *J. Combin. Theory Ser. B* (to appear).

[3] Ethan D. Bolker, Bracing grids of cubes. *Environment and Planning B* 4 (1977), 157–172.

[4] Ethan D. Bolker, Bracing rectangular frameworks. II. *SIAM J. Appl. Math.* 36 (1979), 491–503.

[5] André Bouchet, Nowhere-zero integral flows on a bidirected graph. *J. Combin. Theory Ser. B* 34 (1983), 279–292.

[6] Dorit S. Hochbaum, Nimrod Megiddo, Joseph (Seffi) Naor, and Arie Tamir, Tight bounds and 2-approximation algorithms for integer programs with two variables per inequality. *Math. Programming Ser. B* 62 (1993), 69–83.

[7] Jon Lee, Subspaces with well-scaled frames. *Linear Algebra Appl.* 114/115 (1989), 21–56.

[8] Thomas Zaslavsky, Signed graphs. *Discrete Appl. Math.* 4 (1982), 47–74. Erratum. *Discrete Appl. Math.* 5 (1983), 248.

[9] Thomas Zaslavsky, Orientation of signed graphs. *European J. Combin.* 12 (1991), 361–375.