PREDICTING STRUCTURE IN NONSYMMETRIC SPARSE MATRIX FACTORIZATIONS

JOHN R. GILBERT* AND ESMOND G. NG†

Abstract. Many computations on sparse matrices have a phase that predicts the nonzero structure of the output, followed by a phase that actually performs the numerical computation. We study structure prediction for computations that involve nonsymmetric row and column permutations and nonsymmetric or non-square matrices. Our tools are bipartite graphs, matchings, and alternating paths.

Our main new result concerns LU factorization with partial pivoting. We show that if a square matrix A has the strong Hall property (i.e., is fully indecomposable) then an upper bound due to George and Ng on the nonzero structure of $L+U$ is as tight as possible. To show this, we prove a crucial result about alternating paths in strong Hall graphs. The alternating-paths theorem seems to be of independent interest: it can also be used to prove related results about structure prediction for QR factorization that are due to Coleman, Edenbrandt, Gilbert, Hare, Johnson, Olesky, Pothen, and van den Driessche.

Key words. Gaussian elimination, partial pivoting, orthogonal factorization, matchings in bipartite graphs, strong Hall property, structure prediction, sparse matrix factorization.

AMS(MOS) subject classifications. 05C50, 05C70, 15A23, 65F05, 65F50.

1. Introduction.

Many sparse matrix algorithms predict the nonzero structure of the output of a computation before performing the computation itself. Knowledge of the output structure can be used to allocate memory, set up data structures, schedule parallel tasks, and save time by avoiding operations on zeros. Usually the output structure is predicted by doing some sort of symbolic computation on the nonzero structure of the input; the actual input values are ignored until the numerical computation begins.

This paper discusses structure prediction for orthogonal factorization and for Gaussian elimination with partial pivoting. These algorithms permute the rows and columns of an input matrix nonsymmetrically: starting with a linear system (or least-squares system) of the form $Ax = b$, they instead solve a system $(P^rAP^c)((P^c)^Tx) = (P^rb)$. Here P^r and P^c are permutation matrices; P^r reorders the rows of A (the equations), often for numerical stability or for efficiency, and P^c reorders the columns of A (the variables), often for sparsity. We are most interested in the case where P^c has already been chosen on grounds of sparsity.

Our main tools are bipartite graphs, matchings, and alternating paths. A matching corresponds to a choice of nonzero diagonal elements. Paths in graphs are important in many sparse matrix settings; the notion of alternating paths links matchings, connectivity, and irreducibility. In this paper we highlight a particular sort of irreducibility called the strong Hall property: this generalizes the notion of strong

* Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304-1314 (gilbert@parc.xerox.com). This work was supported in part by the Christian Michelsen Institute, Bergen, Norway, and by the Institute for Mathematics and Its Applications with funds provided by the National Science Foundation.

† Mathematical Sciences Section, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6367 (ngeg@ornl.gov). This author's work was supported by the Applied Mathematical Sciences Research Program of the Office of Energy Research, U.S. Department of Energy, under contract DE-AC05-84OR21400 and by the Institute for Mathematics and Its Applications with funds provided by the National Science Foundation.

connectivity (or irreducibility under symmetric permutations) to nonsymmetric permutations and nonsquare matrices. It turns out that accurate structure prediction is easier for strong Hall matrices than for general matrices. Fortunately, a non-strong-Hall linear system is often most efficiently solved by decomposing it into a sequence of strong Hall systems.

The next section gives definitions and background results, beginning with a definition of exactly what we mean by structure prediction. Section 3 discusses QR factorization. Most of this section reviews earlier work, placing it in a framework that can be used to study LU factorization as well. Section 3 also contains a new tight symbolic result on columnwise orthogonal factorization. Section 4 applies the framework from Section 3 to LU factorization. It contains the main results of the paper, which are tight upper and lower bounds on where fill can occur during LU factorization with partial pivoting. Both Sections 3 and 4 conclude with remarks and open problems; Section 5 makes some final remarks.

2. Preliminaries. We begin this section by defining various kinds of structure prediction. We then discuss several graph-theoretic models of sparse matrix structure. We define so-called "strong Hall bipartite graphs," which model a useful class of fundamental matrices. We prove a crucial result (Theorem 2.9) about matchings and alternating paths in strong Hall graphs, which is the basis for the main results in the rest of the paper. Finally, we briefly review work on structure prediction for symmetric and nonsymmetric Gaussian elimination without pivoting.

2.1. Symbolic and exact structure prediction. Suppose f is a function from matrices to matrices, and $\mathcal{F}$ is an algorithm that computes $f(A)$ by applying elementary transformations (or elementary matrices) to A. The transformations of interest to us are Gauss transforms (elimination steps), Givens rotations, Householder reflections, and row and column swaps. (See Golub and Van Loan [18] for detailed descriptions of various elementary matrix transformations.) We will discuss two kinds of structure prediction, which we call *symbolic* and *exact*.

Symbolic structure prediction models the effect of algorithm $\mathcal{F}$ by modeling the effect of each elementary transformation on the nonzero structure of a matrix. Each elementary transformation is defined to produce zeros in certain positions: a Gauss transform or a Householder reflection annihilates part of a column, a Givens rotation annihilates a single element, and a swap interchanges the zeros in two rows or columns. In symbolic structure prediction we assume that no zeros are ever produced outside those well-defined positions, whether because of numerical coincidence or structural singularity. This "no-cancellation" assumption generally guarantees that we compute an upper bound on the possible nonzero structure of $f(A)$. (At least, it does so if algorithm $\mathcal{F}$ never makes choices based on numerical comparison to zero.)

Symbolic structure prediction can sometimes produce too generous an answer for reasons that have nothing to do with numerical values. For example, consider an algorithm that solves a nonsymmetric linear system $Ax = b$ by forming the normal equations $A^TAx = A^Tb$ and factoring the matrix A^TA. If A has the structure

$$\begin{pmatrix} \times & \times & \times & \times \\ & \times & & \\ & & \times & \\ & & & \times \end{pmatrix},$$

then the symbolic approach will predict (correctly) that A^TA is full, and then (incorrectly) that the factor of this full matrix is a full triangular matrix.

Even though the no-cancellation assumption may not be strictly correct, there are situations in which symbolic structure prediction is the most useful kind. For example, an algorithm may produce *intermediate fill*, or elements that are nonzero at some point in the computation but zero in the final result. (Using the normal equations on the triangular matrix above is an example.) A symbolic prediction can be used to identify all possible intermediate fill locations, and thus to set up a static data structure in which to carry out the entire algorithm. Also, even if an element can be proved to be zero in exact arithmetic, it may not be computed as zero in floating-point arithmetic; we may wish to use symbolic structure prediction to avoid having to decide when such an element should really be considered to be zero.

Exact structure prediction, on the other hand, predicts the nonzero structure of $f(A)$ from that of A without regard to the algorithm that computes $f(A)$. For each input structure S, it yields the set of output positions that are nonzero for some choice of input A having structure S. Thus the output of an exact structure prediction is

$$\bigcup\{\text{structure}(f(A)) : \text{structure}(A) = S\}.$$

In all the interesting cases that we know, this is equal to

$$\bigcup\{\text{structure}(f(A)) : \text{structure}(A) \subseteq S\}.$$

An exact structure prediction for the normal equations algorithm on the triangular input above is that the output has the same structure as the input.

If T is the exactly predicted structure of f on input structure S, then for each nonzero position (i,j) of T there is some A (depending on i, j, and S) for which $[f(A)]_{ij}$ is nonzero. (We use $[f(A)]_{ij}$ to denote the (i,j) element of $f(A)$.) This is what we call a *one-at-a-time* result: it promises that every position in the predicted structure can be made nonzero, but not necessarily all for the same input A. A stronger result is an *all-at-once* result, saying that there is some single A depending only on S for which $f(A)$ has the structure T. Some functions f admit all-at-once exact structure predictions and some do not. For example, we will see that if $f(A)$ is the upper triangular factor in QR factorization of a strong Hall matrix, then there is an all-at-once exact prediction; but if $f(A)$ is the upper triangular factor in LU factorization with partial pivoting of a strong Hall matrix, then the tightest possible exact prediction is only one-at-a-time.

Exact structure prediction depends only on the input structure, so numerical coincidence can still produce unexpected zeros. For example, the exact structure prediction of the upper triangular factor of

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is that it is full, though in fact its $(2,3)$ element is zero (for the particular choice of numerical values).

A symbolic upper bound on structure is an exact upper bound, but not vice versa. In each of Sections 3 and 4, we prove that an exact lower bound is equal to a symbolic upper bound; it follows that the bound is tight both symbolically and exactly.

2.2. Graphs of matrices: Definitions. We assume the reader is familiar with basic graph-theoretic terminology; Harary [20] is a good general reference. We write $G_1 \subseteq G_2$ to mean that graph G_1 is a subgraph of graph G_2.

$$
\begin{pmatrix}
\times & \times & & & \\
& \times & & & \\
\times & & \times & \times & \\
& & \times & \times & \\
& \times & & & \times \\
& & & \times & \times
\end{pmatrix}
$$

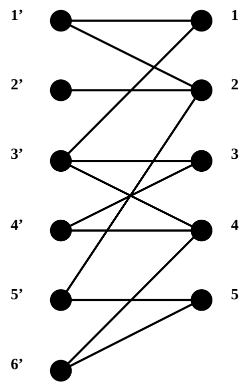


FIG. 1. *A matrix A and its bipartite graph $H(A)$.*

Suppose A is a matrix with m rows and n columns. We write $[A]_{rc}$ for the element in the (r, c) position of A.

We will use three graphs to describe the nonzero structure of A. The *bipartite graph of* A, which we write $H(A)$, has m "row vertices" and n "column vertices." The row and column vertices are drawn from two different copies of the positive integers, which we distinguish by using primes on row vertex names. Thus the row vertices are $1'$, $2'$, $\ldots$, m', and the column vertices are 1, 2, $\ldots$, n. When a variable names a vertex, we will use a prime for a row vertex; thus for example i is a column vertex, and i' is the row vertex with the same number. The graph $H(A)$ has an edge $\langle r', c\rangle$ for each nonzero element $[A]_{rc}$ of A. Figure 1 is an example.

If $m = n$ then A is square, and we also say that $H(A)$ is square. In this case the *directed graph of* A is the directed graph $G(A)$ whose n vertices are the integers $1, \ldots, n$, and whose edges are $\{\langle r, c\rangle : r \neq c \text{ and } [A]_{rc} \neq 0\}$. This graph does not include self-loops, so we cannot tell from $G(A)$ whether or not the diagonal elements of A are zero. Figure 2 is an example.

If $m = n$ and in addition A is symmetric, then the edges of $G(A)$ occur in symmetric pairs. An undirected graph with n vertices and one undirected edge for each symmetric pair of off-diagonal nonzeros is often used to represent the structure of a symmetric matrix. We will write this undirected graph as $G(A)$, and we will not distinguish between it and the directed graph of A. Figure 3 is an example.

The *column intersection graph* of an arbitrary $m \times n$ matrix A is the undirected graph $G_\cap(A)$ whose vertices are the integers $1, \ldots, n$, and whose edges are $\{\langle i, j\rangle :$

$$
\begin{pmatrix}
\times & \times & & & \\
& \times & \times & & \\
\times & & \times & & \\
& & \times & \times & \\
& \times & & \times & \times
\end{pmatrix}
$$

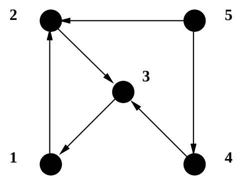


FIG. 2. *A nonsymmetric matrix A and its directed graph $G(A)$.*

$$\begin{pmatrix} \times & \times & & & \\ \times & \times & \times & & \times \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & \times & & \times & \times \end{pmatrix}$$

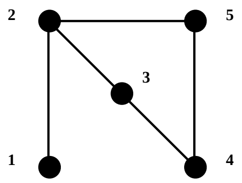


FIG. 3. *A symmetric matrix A and its undirected graph $G(A)$.*

$\exists r$ with $[A]_{ri} \neq 0$ and $[A]_{rj} \neq 0\}$. Thus the vertices of $G_\cap(A)$ are the columns of A, and an edge joins two vertices whose columns share a nonzero row in A. Unless there is numerical cancellation, $G_\cap(A)$ is equal to $G(A^TA)$; in all cases $G_\cap(A) \supseteq G(A^TA)$. Figure 4 is an example.

Table 1 summarizes this notation, as well as some that is defined in later sections. We allow both graphs and matrices as arguments to $G_\cap$ and so on; thus for example if $H = H(A)$ then $G_\cap(H)$ means the same as $G_\cap(A)$.

If x is a vertex of graph G (bipartite, directed, or undirected), we write $Adj_G(x)$ for the set of vertices y such that $\langle x, y\rangle$ is an edge of G. A *walk* is a sequence of edges $\mathcal{P} = (\langle x_0, x_1\rangle, \langle x_1, x_2\rangle, \ldots, \langle x_{p-1}, x_p\rangle)$. We can also describe this walk by listing its vertices, $(x_0, x_1, \ldots, x_p)$. The *length* of the walk is p. We count the empty sequence as a walk of length 0. A *path* is a walk in which all the vertices are distinct. We use $\mathcal{P}[x_i : x_j]$ to denote the portion of path $\mathcal{P}$ from x_i to x_j.[1] If $\mathcal{P}$ is a path from x to y, and $\mathcal{Q}$ is a path from y to z, and y is the only vertex on both $\mathcal{P}$ and $\mathcal{Q}$, then $\mathcal{PQ}$ is a path from x to z.

The *intermediate vertices* of a path $\mathcal{P}$ are all its vertices except its endpoints. If x is a vertex of G and S is a set of vertices of G, we write $Reach_G(x, S)$ to denote the set of vertices y such that G contains a path from x to y with intermediate vertices from S. In this case we also say that y is *reachable from x through S*. For a bipartite graph H, we write $ReachCol_H(x, S)$ to mean the column vertices in $Reach_H(x, S)$.

The following trivial lemma relates paths in a bipartite graph and in its column intersection graph.

[1] When the graph G is bipartite or undirected, $\mathcal{P}[x_i : x_j] = (\langle x_i, x_{i+1}\rangle, \ldots, \langle x_{j-1}, x_j\rangle)$ if $i \leq j$, and $\mathcal{P}[x_i : x_j] = (\langle x_i, x_{i-1}\rangle, \ldots, \langle x_{j+1}, x_j\rangle)$ if $i \geq j$.

$$\begin{pmatrix} \times & \times & & & \\ & \times & & & \\ \times & & \times & \times & \\ & & \times & \times & \\ & \times & & & \times \\ & & & \times & \times \end{pmatrix}$$

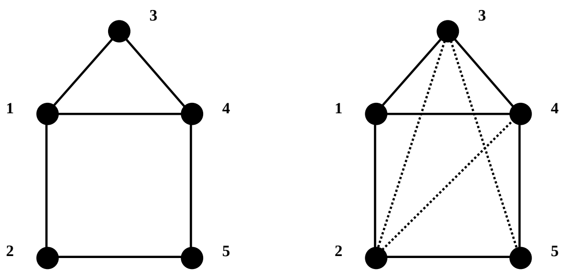


FIG. 4. *A matrix A, its column intersection graph $G_\cap(A)$, and its filled column intersection graph $G^+_\cap(A)$.*

TABLE 1
Graphs associated with the matrix A.

$H(A)$	bipartite graph of arbitrary matrix
$G(A)$	directed graph of square matrix
$G(A)$	undirected graph of square symmetric matrix
$G_\cap(A)$	column intersection graph of arbitrary matrix
$G^+(A)$	filled graph (directed or undirected) of square matrix
$G_\cap^+(A)$	filled graph of column intersection graph of arbitrary matrix
$H^\times(A)$	row merge graph (bipartite) of arbitrary matrix
$G^\times(A)$	row merge graph (directed) of square matrix

LEMMA 2.1. *Let H be a bipartite graph, and let $G_\cap(H)$ be its column intersection graph. For any subset C of the columns of H, and for any two column vertices x and y of H, there is a path in H from x to y whose intermediate column vertices all lie in C if and only if $y \in Reach_{G_\cap(H)}(x, C)$.*

Proof. Immediate. □

2.3. Bipartite matching: Definitions. We briefly summarize some terminology on matchings in bipartite graphs. Lovasz and Plummer [24] is a good general reference on matching; some of our terminology is from Coleman, Edenbrandt, and Gilbert [5]. Brualdi and Ryser [3, Chapter 4] is a good reference on decompositions of bipartite graphs.

Let H be a bipartite graph with m rows and n columns. A *matching* on H is a set $\mathcal{M}$ of edges, no two of which have a common endpoint. A vertex is *covered* or *matched* by $\mathcal{M}$ if it is an endpoint of an edge of $\mathcal{M}$. Clearly, no matching can have more than $\min(m, n)$ edges. A matching is called *column-complete* if it has n edges, and *row-complete* if it has m edges; if $m = n$ a matching with n edges is also called *perfect*. Not every bipartite graph has a column-complete or row-complete matching.

If $\mathcal{M}$ is a matching on H, an *alternating path* (with respect to $\mathcal{M}$) is a path on which every second edge is an element of $\mathcal{M}$; an *alternating walk* is a walk on which every second edge is an element of $\mathcal{M}$. Alternating paths and walks come in two flavors: an *r-alternating path* is one that follows matching edges from columns to rows and non-matching edges from rows to columns; a *c-alternating path* is one that follows matching edges from rows to columns. The reverse of an r-alternating path or walk is a c-alternating path or walk. Suppose the last vertex of one alternating walk is the first vertex of another. If the alternating walks are of the same flavor, their concatenation is an alternating walk of that flavor; if the walks are of opposite flavors, their concatenation is not an alternating walk.

Suppose that $\mathcal{P}$ is an alternating path (of either flavor) from an unmatched vertex v to a different vertex w. If the last vertex w on $\mathcal{P}$ is unmatched, or the last edge on $\mathcal{P}$ belongs to $\mathcal{M}$, then the set of edges $\overline{\mathcal{M}} = \mathcal{M} \oplus \mathcal{P} = (\mathcal{M} \cup \mathcal{P}) - (\mathcal{M} \cap \mathcal{P})$ is another matching; we say that $\overline{\mathcal{M}}$ is obtained from $\mathcal{M}$ by *alternating along path $\mathcal{P}$*. If w is matched in $\mathcal{M}$, then v is matched and w is unmatched in $\overline{\mathcal{M}}$, and $|\overline{\mathcal{M}}| = |\mathcal{M}|$. If w is unmatched in $\mathcal{M}$, then both v and w are matched in $\overline{\mathcal{M}}$, and $|\overline{\mathcal{M}}| = |\mathcal{M}| + 1$. In the latter case we also call $\mathcal{P}$ an *augmenting path* (with respect to $\mathcal{M}$). A classical result of matching theory is that a maximum-size matching can be constructed by greedily finding augmenting paths and alternating along them.

A perfect matching in the bipartite graph $H = H(A)$ of a square matrix can be thought of as a way to find a row permutation P for A so that the permuted matrix PA has nonzero diagonal. Then alternating paths in H correspond to directed paths in $G(PA)$.

LEMMA 2.2. *Suppose A has a nonzero diagonal. The directed graph $G(A)$ has a path from vertex r to vertex c if and only if the bipartite graph $H(A)$ has a path from row r' to column c that is r-alternating with respect to the matching of diagonal edges $\langle i', i\rangle$.*

Proof. Immediate. □

2.4. Hall and strong Hall bipartite graphs. A bipartite graph with m rows and n columns has the *Hall property* if every set of k column vertices is adjacent to at least k row vertices, for all $0 \leq k \leq n$. Clearly a Hall graph must have $m \geq n$. If a graph is not Hall, it cannot have a column-complete matching, because a set of columns that is adjacent only to a smaller set of rows cannot all be matched. The converse is a classical fact about bipartite matching.

THEOREM 2.3 (HALL'S THEOREM). *A bipartite graph has a column-complete matching if and only if it has the Hall property.* □

COROLLARY 2.4. *If a matrix A has full column rank, then $H(A)$ is Hall. Conversely, if H is Hall then almost all matrices A with $H = H(A)$ have full column rank.*

Proof. If $H(A)$ is not Hall, then it has a set of columns with nonzeros in a smaller number of rows; those columns must be linearly dependent. For the converse, let $\mathcal{M}$ be a column-complete matching on H and let R be the set of rows that are matched by $\mathcal{M}$. Consider any matrix A with $H(A) = H$. The submatrix of A consisting of rows R and all columns is square. Its determinant is a polynomial in the nonzero values of A. We claim that this polynomial is not identically zero: if the entries corresponding to edges of $\mathcal{M}$ have the value one and all other entries are zero, the submatrix is a permuted identity matrix and the determinant is ± 1. The set of zeros of a k-variable polynomial has measure zero in $\mathbf{R}^k$, unless the polynomial is identically zero. Thus the set of ways to fill in the values of A to make this submatrix singular has measure zero. If the submatrix is nonsingular, then all the columns of A are linearly independent and A has full column rank. □

A bipartite graph with m rows and n columns has the *strong Hall property* if every set of k column vertices is adjacent to at least $k+1$ row vertices, for all $1 \leq k < n$.[2] It is easy to see that the strong Hall property implies the Hall property.

If the Hall property is a linear independence condition, the strong Hall property is an irreducibility condition: any matrix that is not strong Hall can be permuted to a block upper triangular form called the *Dulmage-Mendelsohn decomposition* [3, 24, 29], in which each diagonal block is strong Hall.[3] Linear equation systems and least-squares problems whose matrices are not strong Hall can be solved by performing first a Dulmage-Mendelsohn decomposition, and then a block backsubstitution that solves a system with each strong Hall diagonal block. Strong Hall matrices are therefore of particular interest in sparse Gaussian elimination and least squares problems.

Brualdi and Shader [4] and Coleman, Edenbrandt, and Gilbert [5] discuss prop-

[2] This definition is from Coleman et al. [5]. Another definition that is sometimes used replaces the bounds on k by $1 \leq k < m$; the only difference is that an m by n matrix with $m > n$ and $m - n$ zero rows that is strong Hall by our definition is not strong Hall by the other definition. All the results in Section 3 and Section 4 hold no matter which definition is used.

[3] This assumes $m \geq n$. More generally, for any m and n, an $m \times n$ matrix can be permuted to a block upper triangular form in which each diagonal block is strong Hall or has a strong Hall transpose.

erties of strong Hall matrices. In the following result, an *independent set* is a set of vertices no two of which are adjacent; an independent set in a bipartite graph corresponds to the rows and columns of a zero submatrix.

THEOREM 2.5 (BRUALDI AND SHADER [4]). *A bipartite graph having m rows and $n \leq m$ columns is Hall if and only if it has no independent set of more than m vertices, and strong Hall if and only if it has no independent set of at least m vertices that includes at least one vertex from each part.* $\Box$

A square strong Hall matrix is often called *fully indecomposable*, meaning that there is no way to permute its rows and columns into a block triangular form with more than one block [3]. This gives the following (standard) result.

THEOREM 2.6. *Let $H = H(A)$ be a square strong Hall graph. Then for all row and column permutations P^r and P^c, the directed graph $G(P^r A P^c)$ is strongly connected.* $\Box$

We conclude this subsection by proving a theorem (Theorem 2.9) about strong Hall matrices that is useful in several structure prediction results. The theorem first appeared in a technical report by Gilbert [15]; other proofs have been given by Hare, Johnson, Olesky, and van den Driessche [21] and Brualdi and Shader [4]. First we need two technical lemmas.

LEMMA 2.7. *Let H be a strong Hall graph and let $\langle r', c\rangle$ be an edge of H. Then there is a column-complete matching that includes $\langle r', c\rangle$, and unless $\langle r', c\rangle$ is the only edge of H there is a column-complete matching that excludes $\langle r', c\rangle$.*

Proof. First, let $\overline{H}$ be H without vertices r' and c and their incident edges. We show that $\overline{H}$ is Hall. Every nonempty set C of columns of $\overline{H}$ is a nonempty proper subset of columns of H, and hence is adjacent to at least $|C| + 1$ rows of H. This includes at least $|C|$ rows of $\overline{H}$. Therefore $\overline{H}$ is Hall and has a column-complete matching. That matching plus edge $\langle r', c\rangle$ is a column-complete matching on H.

Now assume that H has more than one edge, and let $\widehat{H}$ be H without the single edge $\langle r', c\rangle$. We show that $\widehat{H}$ is Hall. Any nonempty proper subset C of columns is adjacent to at least $|C| + 1$ rows in H, hence to at least $|C|$ rows in $\widehat{H}$. The same argument works if C is the set of all columns and H has at least $|C| + 1$ nonzero rows.

If C is the set of all columns and H has exactly $|C|$ nonzero rows R, we argue as follows: If r' were adjacent only to c in H, then $C - c$ would be adjacent in H only to the $|C - c|$ rows $R - r'$,[4] contradicting the fact that H is strong Hall. Thus C must be adjacent in $\widehat{H}$ to all $|C|$ rows.

Whether or not H is square, then, we conclude that $\widehat{H}$ is Hall. Thus $\widehat{H}$ has a column-complete matching, which is a column-complete matching on H that excludes $\langle r', c\rangle$. $\Box$

LEMMA 2.8. *If H is strong Hall and has more nonzero rows than columns, and $\mathcal{M}$ is any column-complete matching on H, then from every row or column vertex w of H there is a c-alternating path to some unmatched row vertex r' (which depends on w and $\mathcal{M}$).*

Proof. This is a standard result on Dulmage-Mendelsohn decomposition; we include a proof here only to be self-contained. If w is an unmatched row there is nothing to prove. Otherwise, let C be the set of columns reachable by c-alternating paths from w. Then C is nonempty. Let R be the set of row vertices adjacent to vertices of C. Since H is strong Hall and has more nonzero rows than columns, $|R|$ is larger than $|C|$. Thus there is some vertex r' in R that is not matched to a vertex

[4] If C is a set of vertices and c is a vertex, we use $C - c$ to denote the set $C - \{c\}$.

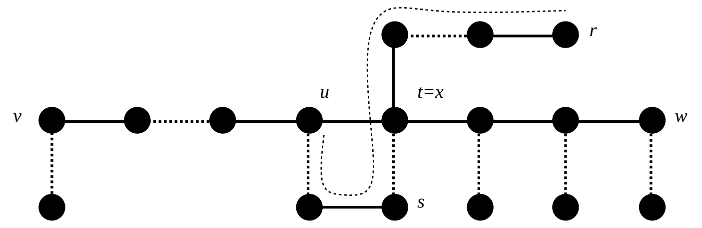


FIG. 5. *Case 1 of Theorem 2.9. The dashed edges are the matching $\mathcal{M}$. $\mathcal{P}$ is the horizontal path from v to w. The light dotted line shows path $\overline{\mathcal{R}}$ from u to r. Path $\mathcal{P}[v:u]\overline{\mathcal{R}}[u:x]$ is c-alternating with respect to $\mathcal{M}$.*

in C. Suppose r' is adjacent to $c \in C$. The c-alternating path from w to c can be extended by edge $\langle c, r' \rangle$ to r'. Now if r' were matched, it would be matched to a vertex v not in C; but then there would be a c-alternating path from w to v, contrary to the definition of C. Therefore r' is the desired unmatched row vertex. ☐

Finally we prove the main result about alternating paths in strong Hall graphs.

THEOREM 2.9 (ALTERNATING-PATHS THEOREM). *Let H be a strong Hall graph with at least two rows, let v be a column vertex of H, and let w be any row or column vertex of H such that a path exists from v to w. Then H has a column-complete matching relative to which there exists a c-alternating path from v to w (or, equivalently, an r-alternating path from w to v).*

Proof. Since some of the vertices in this proof can be either row or column vertices, we will not use primed variables; an unprimed variable may denote either a row or a column vertex.

If H is square, or if H has only as many nonzero rows as columns, then the theorem follows from Theorem 2.6 and Lemma 2.2.

Suppose that H has more nonzero rows than columns. If $v = w$ there is nothing to prove. Otherwise, by hypothesis there is at least one path from v to w. By Lemma 2.7 there is a column-complete matching that omits the first edge on that path. (Note that this edge is not the only edge of H since H has more nonzeros than columns.) If $\mathcal{P}$ is a path from v to w and $\mathcal{M}$ is a column-complete matching that omits the first edge on $\mathcal{P}$, let u (dependent on $\mathcal{P}$ and $\mathcal{M}$) be the last vertex on $\mathcal{P}$ such that $\mathcal{P}[v:u]$ is alternating. Then $\mathcal{P}[v:u]$ is c-alternating. Among all such paths and column-complete matchings, choose $\mathcal{P}$ and $\mathcal{M}$ such that the length of $\mathcal{P}[u:w]$ is minimum.

If $u = w$ the theorem holds. We shall assume $u \neq w$ and derive a contradiction. Let t be the next vertex after u on $\mathcal{P}$. Both the last edge of $\mathcal{P}[v:u]$ and the first edge of $\mathcal{P}[u:w]$ (which is $\langle u, t \rangle$) must be non-$\mathcal{M}$ edges, or else $\mathcal{P}[v:t]$ would be alternating.

Because $\mathcal{P}[v:u]$ is c-alternating and begins and ends with non-matching edges, u is a row vertex and hence t is a column vertex. Let s be the vertex matched to t in $\mathcal{M}$, which may or may not be on $\mathcal{P}$.

Lemma 2.8 implies that there is an unmatched row vertex r and a c-alternating path $\overline{\mathcal{R}}$ from u to r (possibly $u = r$). Now t is on path $\overline{\mathcal{R}}$ if and only if s is. There are two cases.

Case 1. Both t and s are on $\overline{\mathcal{R}}$. In this case $\mathcal{P}[v:u]\overline{\mathcal{R}}[u:t]$ is a c-alternating walk from v to t. Therefore there is a c-alternating path $\mathcal{R}$ from v to t. Let x be the last vertex on $\mathcal{P}$ that is also on $\mathcal{R}$ (so x is on $\mathcal{P}[t:w]$). Then $\overline{\mathcal{P}} = \mathcal{R}[v:x]\mathcal{P}[x:w]$ is a path from v to w. But this is a contradiction: $\overline{\mathcal{P}}$ is c-alternating from v at least to x, and

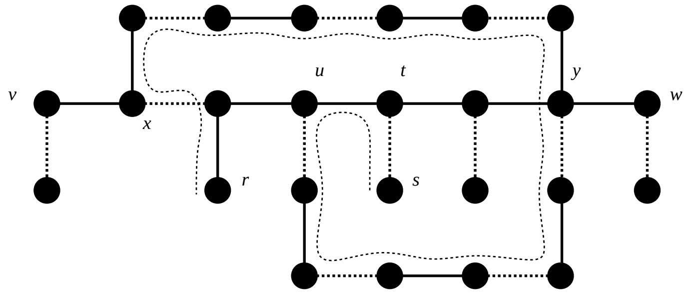


FIG. 6. *Case 2 of Theorem 2.9. The most complicated version is shown. The dashed edges are the matching $\mathcal{M}$. $\mathcal{P}$ is the horizontal path from v to w. The light dotted line shows path R from s to r. Path $\mathcal{P}[v:x]\mathcal{R}[x:y]$ is c-alternating with respect to $\mathcal{M} \oplus \mathcal{R}$. A simpler version, not shown, is if $\mathcal{R}$ does not intersect $\mathcal{P}$ after u. Then $x = u$, $y = t$, and $\mathcal{P}[v:u]\langle u, t\rangle$ is c-alternating with respect to $\mathcal{M} \oplus \mathcal{R}$.*

$\overline{\mathcal{P}}[x:w]$ is shorter than $\mathcal{P}[u:w]$. This contradicts the choice of $\mathcal{P}$. Figure 5 illustrates this case.

Case 2. Neither t nor s is on $\overline{\mathcal{R}}$. In this case $\mathcal{R} = \langle s, t\rangle\langle t, u\rangle\,\overline{\mathcal{R}}$ is a c-alternating path from s to r. Since r is an unmatched row and s is matched to t, $\overline{\mathcal{M}} = \mathcal{M} \oplus \mathcal{R}$ is a column-complete matching. Path $\mathcal{R}$ is c-alternating with respect to $\overline{\mathcal{M}}$.

Let x be the first vertex on $\mathcal{P}$ that is also on $\mathcal{R}$ (so x is on $\mathcal{P}[v:u]$), and let y be the last vertex on $\mathcal{P}$ that is also on $\mathcal{R}$ (so y is on $\mathcal{P}[t:w]$). Then $\overline{\mathcal{P}} = \mathcal{P}[v:x]\mathcal{R}[x:y]\mathcal{P}[y:w]$ is a path from v to w. The path $\mathcal{P}[v:x]$ is c-alternating with respect to both $\mathcal{M}$ and $\overline{\mathcal{M}}$, because $\mathcal{M}$ and $\overline{\mathcal{M}}$ agree on $\mathcal{P}[v:x]$. Depending on whether x precedes or follows y on $\mathcal{R}$, the path $\mathcal{R}[x:y]$ is c-alternating either with respect to $\overline{\mathcal{M}}$ or with respect to $\mathcal{M}$, because $\mathcal{M}$ and $\overline{\mathcal{M}}$ disagree on $\mathcal{R}$. Therefore $\overline{\mathcal{P}}[v:y] = \mathcal{P}[v:x]\mathcal{R}[x:y]$ is c-alternating either with respect to $\overline{\mathcal{M}}$ or to $\mathcal{M}$. Figure 6 illustrates this case.

But this is a contradiction: With respect to one of the column-complete matchings $\overline{\mathcal{M}}$ and $\mathcal{M}$, we have shown that $\overline{\mathcal{P}}$ is a path from v to w that is c-alternating from v at least as far as y, and $\overline{\mathcal{P}}[y:w]$ is shorter than $\mathcal{P}[u:w]$. This contradicts the choice of $\mathcal{P}$ and $\mathcal{M}$, and finishes the proof of Theorem 2.9. □

2.5. Gaussian elimination without pivoting. We now briefly review a graph-theoretic model of LU factorization without row or column interchanges. The undirected version of this model is due to Parter [27] and was developed extensively by Rose [30]; the directed version was developed by Rose and Tarjan [31]. George and Liu [11] is a good source for the undirected model. Gilbert [14] surveys these and related results.

If $G = G(A)$ is a directed or undirected graph, we define the *deficiency* of a vertex v of G as the set of edges

$$\{\langle r, c\rangle : v \in \mathit{Adj}_G(r), c \in \mathit{Adj}_G(v), \text{ and } c \notin \mathit{Adj}_G(r)\}.$$

The deficiency of v corresponds to the fill that occurs in A when the (v, v) element is used as a pivot in Gaussian elimination. Therefore we can define a sequence of *elimination graphs* $G_0, G_1, \ldots, G_n$, where $G_0 = G(A)$ and G_i is obtained from G_{i-1} by adding the deficiency of vertex i (in G_{i-1}) and deleting vertex i and its incident edges. Then G_i is the graph of the $(n-i)\times(n-i)$ Schur complement that remains after eliminating the first i vertices of A. This is in the symbolic sense—that is, it ignores possible numeric cancellation. We define the *filled graph* of A, which we write $G^+(A)$, as the n-vertex graph containing all the edges of all the G_i's. Thus we have the following result.

THEOREM 2.10. *Suppose the square matrix A can be factored as $A = LU$ without row or column interchanges. Then $G(L+U) \subseteq G^+(A)$ with equality unless there is cancellation in the factorization. In other words, the filled graph contains edges for all the nonzeros of L and U.* $\square$

If A is symmetric and $G(A)$ is the undirected graph, then $G^+(A)$ is undirected. (Remember that we do not distinguish between an undirected graph and a directed graph with symmetric pairs of edges.) Historically, filled graphs were studied first in the undirected case, specifically for the Cholesky factorization of symmetric positive definite matrices. The theory of undirected filled graphs, which are the same as *chordal graphs*, is quite rich [19, 30].

We can characterize the structure of $G^+(A)$ in terms of paths in the graph of A, without actually computing all the elimination graphs. In the following theorem, the paths can be interpreted as directed paths for nonsymmetric matrices and either directed or undirected paths for symmetric matrices.

LEMMA 2.11 (ROSE, TARJAN, AND LUEKER [31, 32]). *Let G be a directed or undirected graph whose vertices are the integers 1 through n, and let G^+ be its filled graph. Then $\langle x, y\rangle$ is an edge of G^+ if and only if there is a path in G from x to y whose intermediate vertices are all smaller than* $\min(x, y)$. $\square$

Paths from x to y whose intermediate vertices are all smaller than $\min(x, y)$ are sometimes referred to as *fill paths*.

A graph that is often useful in nonsymmetric structure prediction is the *filled column intersection graph* of an arbitrary $m \times n$ matrix A. This graph, which we write $G_\cap^+(A)$, is just $G^+(G_\cap(A))$; it is the n-vertex undirected filled graph of the column intersection graph of A. Figure 4 is an example. The graph $G_\cap^+(A)$ is related to the normal equations; its structure is the symbolic result of forming A^TA and then computing the Cholesky factor of that matrix. Section 3 discusses the conditions under which this symbolic structure prediction is exact.

2.6. Lemmas on exact structure in Gaussian elimination. In this final subsection we prove some easy lemmas that take into account the values of the nonzeros in the matrix. These results will be the building blocks for the exact structure predictions in the rest of the paper.

LEMMA 2.12. *Suppose A is square and nonsingular, and has a triangular factorization $A = LU$ without pivoting. Let r be a row index and c a column index of A, and let K be the submatrix of A consisting of rows 1 through $\min(r, c) - 1$ and r, and columns 1 through $\min(r, c) - 1$ and c. Then $[L + U]_{rc}$ is zero if and only if K is singular.*

Proof. Let $s = \min(r, c)$. Factor $K = L_KU_K$. Then $[U_K]_{ss} = [U]_{rc}$ if $r \le c$, and $[U_K]_{ss} = [L]_{rc}[U]_{cc}$ if $r > c$, so $[U_K]_{ss}$ is zero if and only if $[L + U]_{rc}$ is zero. The determinant of U_K is the same as that of K, and the first $s - 1$ diagonal elements of U_K are the same as those of U, so $[U_K]_{ss} = 0$ if and only if K is singular. $\square$

LEMMA 2.13. *Suppose A is square and nonsingular, and has a triangular factorization $A = LU$ without pivoting. Suppose also that all the diagonal elements of A except possibly the last one are nonzero, and that every square Hall submatrix of A is nonsingular. Then $G(L + U) = G^+(A)$; that is, every nonzero predicted by the filled graph of A is actually nonzero in the factorization.*

Proof. Suppose $\langle r, c\rangle$ is an edge of $G^+(A)$. Then there is a fill path $\mathcal{P}$ from r to c whose intermediate vertices are less than $s = \min(r, c)$.

Let K be the submatrix of A mentioned in Lemma 2.12, consisting of rows 1 through $s-1$ and r, and columns 1 through $s-1$ and c. For convenience, call the last row and column in K number r and c respectively instead of number s. Then path $\mathcal{P}$ corresponds to a path in $H(K)$ from row vertex r' to column vertex c, which is r-alternating with respect to the matching $\mathcal{M}$ of edges $\langle i', i\rangle$.

Now $\mathcal{M}$ is one edge short of being a perfect matching on K, because column c and row r' are not matched. However $\mathcal{P}$ is an augmenting path with respect to $\mathcal{M}$, and therefore $\mathcal{M} \oplus \mathcal{P}$ is a perfect matching on K. Since K has a perfect matching, it is Hall; thus its determinant is nonzero by hypothesis, and $[L+U]_{rc}$ is nonzero by Lemma 2.12. □

The hypothesis that A has nonzero diagonal in Lemma 2.13 is crucial. Brayton, Gustavson, and Willoughby [2] gave the following counterexample in the case when this hypothesis is not included. Let

$$A = \begin{pmatrix} \times & \times & \times & \\ \times & & & \\ \times & \times & \times & \\ \times & & & \times \end{pmatrix}.$$

Then the $(4,3)$ entry in $G^+(A)$ is nonzero, but $[L]_{4,3} = 0$ regardless of the nonzero values of A.

LEMMA 2.14. *Suppose bipartite graph H has a perfect matching $\mathcal{M}$. Let A be a matrix with $H(A) = H$, such that $[A]_{rc} > n$ for $\langle r', c\rangle \in \mathcal{M}$ and $0 < [A]_{rc} < 1$ for $\langle r', c\rangle \notin \mathcal{M}$. If A is factored by Gaussian elimination with partial pivoting, then the edges of $\mathcal{M}$ will be the pivots.*

Proof. When the rows of matrix A are permuted so that the edges of $\mathcal{M}$ are the diagonal elements, the values chosen make the permuted matrix strongly diagonally dominant. □

3. Orthogonal factorization. Let A be a matrix with m rows and $n \leq m$ columns, with full column rank n. In this section we consider the orthogonal factorization $A = QR$, where Q is an $m \times m$ orthogonal matrix (that is, $Q^TQ = I$), and R is an $m \times n$ upper triangular matrix with nonnegative diagonal entries. (All the nonzeros of R are in the $n \times n$ upper triangle, so we will think of R as being $n \times n$.) This factorization is unique. It arises in least squares and other optimization problems [18, 22].

To compute the QR factorization, A is transformed into R by multiplying it on the left by a sequence of orthogonal transformations that annihilate nonzeros below the main diagonal. In most applications, Q is not computed explicitly: either the orthogonal transformations are applied to a right-hand side at the same time as to A, or else a description of the sequence is saved to be applied later.

At least two structure prediction problems are of interest here. First, what is the nonzero structure of A at each step of annihilation? Second, what is the nonzero structure of R? The answer to the first question depends on the algorithm we use to compute the factorization; the answer to the second does not.

In Section 3.1 below, we review work of George, Liu, and Ng on intermediate fill during column QR factorization. We then give a new tight symbolic result on column QR factorization. In Section 3.2, we survey several authors' work on predicting the structure of R; in Section 3.3, we re-prove a result of Coleman, Edenbrandt, and

Gilbert in a framework that relates it to the new results on LU factorization in Section 4. Finally, in Section 3.4, we briefly survey some related work.

3.1. Nonzero structure of A during annihilation. In this section we develop a symbolic model of the column Givens and Householder algorithms for reducing A to upper triangular form. Our goal is a tight symbolic result, that is, an accurate description of the nonzero structure of A during the algorithm, under the assumption that no cancellation occurs.

The standard algorithms to compute R from A multiply A on the left by a sequence either of Householder reflections or of Givens rotations [18]. Multiplication by a Householder reflection reflects a vector with respect to a specified hyperplane; a Householder reflection can be chosen to annihilate all but one of the entries of the vector. Multiplication by a Givens rotation rotates a vector through a specified angle in the plane of two specified coordinate axes; a Givens rotation can be chosen to annihilate any single entry of the vector. We consider three algorithms to compute R from A in the sparse setting: *row Givens*, *column Givens*, and *column Householder*.

The sparse row Givens algorithm is due to George and Heath [8]. They first predict the nonzero structure of R, and set up a static data structure to hold R. Then they annihilate nonzeros from one row of A at a time, processing each row until either it becomes completely zero, or its structure fits into an empty row of the static data structure. This approach is attractive because only the data structure for R and the storage for the rows of A are needed in the annihilation process. Thus the only structure prediction necessary is for R, as described in Sections 3.2 and 3.3.

The column Givens and column Householder algorithms both annihilate the subdiagonal elements of one column of A at a time. We will analyze them from the symbolic point of view, that is, assuming that zeros are produced only by intentional annihilation and not by numerical cancellation or coincidence. Define the sequence $A_0, A_1, \ldots, A_n$, where $A_0 = A$ and A_i is the $(m-i) \times (n-i)$ submatrix remaining to be processed at the end of step i of the annihilation. For convenience, the columns of the $(m-i) \times (n-i)$ matrix A_i are labeled from i to n, and the rows of A_i are labeled from i to m. The matrix A_i is obtained from A_{i-1} by annihilating the nonzeros below the diagonal in column i of A_{i-1}. The Givens algorithm uses one rotation for each subdiagonal nonzero in column i of A_{i-1}; the Householder algorithm uses one reflection to annihilate the entire column. The structural effects are closely related, so we combine their descriptions.

Consider Givens rotations first. Suppose $[A_{i-1}]_{ki}$ is nonzero, $k > i$, and assume that any nonzero $[A_{i-1}]_{ji}$, $i < j < k$, has been annihilated. Then $[A_{i-1}]_{ki}$ will be annihilated by a Givens rotation, which is constructed using $[A_{i-1}]_{ii}$ and $[A_{i-1}]_{ki}$. This rotation replaces rows k and i by linear combinations of their old values; symbolically, except for the (k, i) element, it replaces both their nonzero structures with the union of their nonzero structures. Thus the structure of row k of A_i is the union of the structures of those rows j of A_{i-1} for which $i \le j \le k$ and $[A_{i-1}]_{ji} \ne 0$. Moreover, at the end of step i, the structure of row i of A_i is the union of the structures of those rows j of A_{i-1} for which $i \le j \le m$ and $[A_{i-1}]_{ji} \ne 0$.

Now consider (the row-oriented version of) Householder reflections. The Householder reflection that annihilates the subdiagonal nonzeros of column i of A_{i-1} replaces all the rows containing those nonzeros with linear combinations of their old values. Symbolically, every row with a nonzero in column i of A_{i-1} has the same structure in A_i, namely the union of their original structures in A_{i-1}.

In terms of structures, the fundamental difference between Givens rotations and

Householder reflections is the number of rows participating in one reduction operation. In one Householder reduction, *all* rows that have a nonzero in column i of A_{i-1} participate in a reduction step, whereas in a Givens reduction, only a subset of those rows are involved.

We now describe a bipartite graph model that George, Liu, and Ng [12] developed to analyze the reduction process using Givens rotations. Their model associates a bipartite graph H_i with the matrix A_i. We number the $m-i$ row vertices of H_i from $i+1$ to m, and the $n-i$ column vertices from $i+1$ to n. The changes in the structure of A_i due to the reduction process are described in terms of transformations on the graph H_i. Because of the similarity between Givens reductions and Householder reflections, this model can be extended to cover both cases. We summarize these results below; proofs can be found in the paper [12]. All these results are symbolic; they assume that zeros are introduced only by explicit annihilation, not by cancellation.

The following results contain a parameter ρ, which we introduce to cover both of the column algorithms. We define $\rho = r$ for Givens rotations, and $\rho = m$ for Householder reflections.

We begin by formalizing the symbolic effect of annihilating one column, that is, the relationship between H_{i-1} and H_i. The four statements in the lemma below are easily seen to be equivalent.

LEMMA 3.1.

- *For* $r > i$, $Adj_{H_i}(r') =$

$$\begin{cases} Adj_{H_{i-1}}(r'), & \text{if } i \notin Adj_{H_{i-1}}(r'), \\ \bigcup\{Adj_{H_{i-1}}(s') : i \le s' \le \rho, i \in Adj_{H_{i-1}}(s')\} - \{i\}, & \text{otherwise.} \end{cases}$$

- *For* $r > i$, $Adj_{H_i}(r') = ReachCol_{H_{i-1}}(r', \{i, i', (i+1)', \cdots, \rho'\})$.
- *For* $r > i$, $c \in Adj_{H_i}(r')$ *if and only if there exists a path of length 1 or 3 from* r' *to* c *through* $\{i, i', (i+1)', \cdots, \rho'\}$ *in* H_{i-1}.
- *For* $r > i$ *and* $c > i$, $c \in Adj_{H_i}(r')$ *if and only if either* $c \in Adj_{H_0}(r')$ *or for some* $k \le i$, *there is a path* (r', k, s', c) *in* H_{k-1} *with* $k \le s \le \rho$.

□

We wish to characterize fill in terms of the structure of the original matrix. George, Liu and Ng [12] provided upper and lower bounds on the structure of H_i, but neither bound is tight. Their upper bound is as follows.

THEOREM 3.2. *For* $r > i$, $Adj_{H_i}(r') \subseteq ReachCol_{H_0}(r', \{1, \cdots, i, 1', \cdots, \rho'\})$. □

Note that Theorem 3.2 provides only a *necessary* condition for a fill element to occur during the annihilation process. Figure 7 (from [12]) is an example showing that Theorem 3.2 is not tight. There is a path $(4', 2, 1', 3)$ in the graph H_0, but it is easy to verify that *no* zero element in A becomes nonzero in reducing A to upper triangular form by Givens rotations or Householder reflections.

The George, Liu, and Ng lower bound is as follows.

THEOREM 3.3. *Suppose that* H_0 *contains a path* $(r', c_1, r'_1, c_2, r'_2, \cdots, c_t, r'_t, c)$ *whose intermediate vertices are all in* $\{1, \cdots, i, 1', \cdots, \rho'\}$. *If* $c_k \le r'_k$ *for* $k \le t$ *and* $c_{k+1} \le r'_k$ *for* $k < t$, *then* $c \in Adj_{H_i}(r')$. □

Again Theorem 3.3 is a partial characterization of fill; it provides only a *sufficient* condition. Figure 8 (also from [12]) shows that the condition in Theorem 3.3 is not necessary. Consider the path $(5', 2, 1', 1, 4', 3)$ in H_0. It does not satisfy the condition in Theorem 3.3 and it is the only path from $5'$ to 3 in H_0. However it is straightforward

$$\begin{pmatrix} \times & \times & \times \\ & \times & \\ & & \times \\ & \times & \end{pmatrix}$$

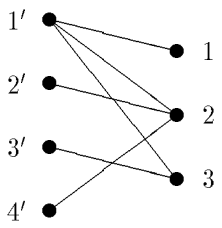


FIG. 7. *The converse of Theorem 3.2 is not true.*

to verify that $3 \in Adj_{H_2}(5')$ when either Givens rotations or Householder reflections are used.

We now provide a necessary and sufficient condition, in terms of paths in H_0, for fill to occur in the symbolic orthogonal factorizations. As in the case of sparse Gaussian elimination without pivoting, we define a class of *fill paths* in H_0 for sparse orthogonal factorizations: a path

$$\mathcal{P} = (r', c_1, r'_1, c_2, r'_2, \cdots, c_t, r'_t, c)$$

in H_0 is a fill path for sparse Givens rotation or sparse Householder transformation if either $t = 0$ or the following conditions are satisfied.

1. $c_k < \min(r', c)$ and $r'_k \leq \rho$, for all k.
2. Let c_p be the largest c_k. Then there is some q with $p \leq q \leq t$ such that $c_p \leq r'_q \leq \rho$, and the three paths $\mathcal{P}[r' : c_p]$, $\mathcal{P}[c_p : r'_q]$, and $\mathcal{P}[r'_q : c]$, are also fill paths in H_0.

By this definition, all edges in H_0 are also fill paths. The main new result of this section is the following, which generalizes the last statement of Lemma 3.1. It gives a necessary and sufficient condition for a zero element of A to become nonzero at some stage of the annihilation process, in the symbolic sense. The proof of the result is an easy induction, and is omitted.

THEOREM 3.4. *For $r', c > i$, $c \in Adj_{H_i}(r')$ if and only if there is a fill path joining r' and c in H_0.* $\square$

$$\begin{pmatrix} \times & \times & & \\ & \times & & \\ & & \times & \\ \times & & \times & \times \\ & \times & & \end{pmatrix}$$

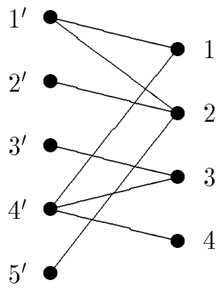


FIG. 8. *The converse of Theorem 3.3 is not true.*

Consider the path $(4', 2, 1', 3)$ in H_0 in Figure 7. Since it does not satisfy condition (2), the $(4, 3)$ element of A will remain zero throughout the computation, which is indeed the case for either Givens or Householder. Also consider the example in Figure 8. Although the path $(5', 2, 1', 1, 4', 3)$ does not satisfy the condition in Theorem 3.3, it does satisfy condition (2) above. Hence, the $(5, 3)$ element of A will become nonzero at some point during the computation, assuming exact numerical cancellation does not occur.

Unfortunately, unlike the case of sparse Gaussian elimination without pivoting, there does not appear to be a simple and non-recursive way to express the fill property.

Finally, we define a graph whose structure captures all of the H_i for the case of Householder reflections. The *(bipartite) row merge graph* of a matrix A whose diagonal is nonzero, which we write $H^{\times}(A)$, is the union of H_i (by the Householder interpretation) for $1 \leq i \leq n$. Thus $H^{\times}(A)$ has m row vertices and n column vertices, and is constructed by the following process. Begin with the bipartite graph $H(A)$, which includes all edges of the form $\langle i', i \rangle$ because A has nonzero diagonal. For each k from 1 to n, add an edge from each row $r' \geq k$ adjacent to column k to each column $c \geq k$ adjacent to any such row. (In other words, take those rows at or below row k with nonzeros in column k, and merge the parts of their nonzero structures at or to the right of column k.)

We also define a directed version of the row merge graph. The bipartite row merge graph $H^{\times}(A)$ is a bipartite graph with m rows, $n \leq m$ columns, and a column-complete matching of edges $\langle i', i \rangle$. The *(directed) row merge graph*, which we write $G^{\times}(A)$, is the n-vertex directed graph whose adjacency matrix has the structure of the first n rows of $H^{\times}(A)$.

Theorems 3.2, 3.3, and 3.4 can be translated into statements about $H^{\times}(A)$. We will need one of these later.

COROLLARY 3.5. *If A is an $m \times n$ matrix with nonzero diagonal, $m \geq n$, and $\langle r', c \rangle$ is an edge of the row merge graph $H^{\times}(A)$, then there is a path in $H(A)$ from row vertex r' to column vertex c whose intermediate column vertices are all numbered less than* $\min(r', c)$.

Proof. Immediate from Theorem 3.2 or Theorem 3.4. □

3.2. Upper bounds on nonzero structure of R. If A has full column rank and factorization $A = QR$, it follows from the column Householder algorithm (and the uniqueness of the factorization) that $G(R) \subseteq G^{\times}(A)$. In this section we state and prove a bound on the structure of R that seems weaker than this one; then we show that if A is strong Hall then the weaker bound is tight, and hence in that case the two bounds are the same.

If $A = QR$ then $A^TA = R^TQ^TQR = R^TR$. Thus (the upper triangular part of) R is equal to the Cholesky factor of the normal-equations matrix A^TA (which is symmetric and positive definite). George and Heath [8] used this fact in their implementation of sparse orthogonal factorization by Givens rotations. They predict the structure of A^TA to be the column intersection graph $G_{\cap}(A)$, which has a nonzero in position (i, j) whenever columns i and j of A have a common nonzero row; then they predict the structure of R to be $G_{\cap}^{+}(A)$, the symbolic Cholesky factor of that structure.

We will derive this prediction as a corollary of a relationship between row merge graphs and column intersection graphs. We prove this relationship for all of $G^{\times}(A)$ even though the structure of R concerns only the "upper triangle" of $G^{\times}(A)$; we will need the more general version in Section 4. A similar result for square matrices can be found in George and Ng [9].

THEOREM 3.6. *If A is an $m \times n$ matrix with $m \geq n$ and nonzero diagonal elements, then $G^{\times}(A) \subseteq G_{\cap}^{+}(A)$.*

Proof. Suppose $\langle r, c\rangle$ is an edge of $G^{\times}(A)$. Then $\langle r', c\rangle$ is an edge of $H^{\times}(A)$ with $r' \leq n$. Let $i = \min(r, c) - 1$. Then by Corollary 3.5 there is a path from r' to c in $H(A)$ whose column vertices are all numbered at most i. Since A has nonzero diagonal, $\langle r', r\rangle$ is an edge of $H(A)$. Thus $H(A)$ contains a path between column vertices r and c, whose intermediate column vertices are all smaller than $\min(r, c)$. Therefore (by Lemma 2.1), the column intersection graph $G_{\cap}(A)$ contains a path between vertices r and c, whose intermediate vertices are all smaller than $\min(r, c)$. Thus (by Lemma 2.11), $\langle r, c\rangle$ is an edge of $G_{\cap}^{+}(A)$. □

COROLLARY 3.7 (GEORGE, HEATH, LIU, AND NG [8, 10, 13]). *If $A = QR$ is the orthogonal factorization of a matrix with full column rank and nonzero diagonal, then $G(R) \subseteq G_{\cap}^{+}(A)$.* □

Corollary 3.7 says that the structure $G_{\cap}^{+}(A)$ of the Cholesky factor of A^TA is an upper bound on the structure of R. This upper bound may be an overestimate for reasons that have nothing to do with the numerical values of the nonzeros of A. An example is the upper triangular matrix in Section 2.1.

3.3. Lower bounds on nonzero structure of R. Coleman, Edenbrandt, and Gilbert [5] showed that $G_{\cap}^{+}(A)$ does not overpredict $G(R)$ if the matrix A is strong Hall. We give a proof that is related to theirs, but (unlike them) we use the alternating-paths theorem explicitly, to highlight the similarity between this result and Theorem 4.5 on LU factorization.

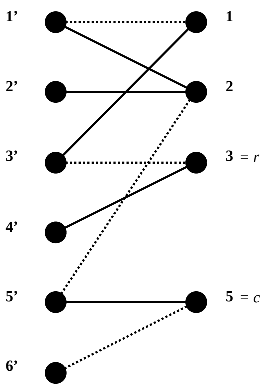
The hypotheses of Theorem 3.8 do not include a nonzero diagonal. This is because both $G(R)$ and $G_{\cap}^{+}(H)$ are independent of the row ordering of H, and since H is strong Hall its rows can be permuted to make the diagonal nonzero.

THEOREM 3.8 (COLEMAN, EDENBRANDT AND GILBERT [5]). *Let H be a bipartite graph with the strong Hall property. Then there is a matrix A with full column rank and with $H(A) = H$, such that the orthogonal factorization $A = QR$ satisfies $G(R) = G_{\cap}^{+}(H)$.*

Proof. First we show that any single edge of $G_{\cap}^{+}(H)$ can be made nonzero by an appropriate choice of A; then we show that there is one choice of A that makes all those positions nonzero at once. We shall think of the entries of A that correspond to edges of H as variables; a "choice of values for A" means an assignment to those variables. Figure 9 illustrates the proof.

Choose r and c with $r < c \leq n$. Take an arbitrary $m \times n$ matrix A with factorization QR, such that the first r columns of A are linearly independent. Now let K be the submatrix of A^TA consisting of columns 1 through $r-1$ and c, and rows 1 through r. Lemma 2.12 applies to A^TA (because A^TA is positive definite), and says that K is singular if and only if $[R]_{rc}$, the entry in the (r, c) position of R, is zero. Thus $[R]_{rc}$ is zero if and only if a certain polynomial p_{rc} in the nonzero entries of A (namely the determinant of K) is zero.

We now show that if A is a matrix with $H(A) = H$ and (r, c) is an edge of $G_{\cap}^{+}(H)$, then the polynomial p_{rc} is not identically zero. (Note that p_{rc} has a variable for each edge of H.) Let $\overline{H}$ be the subgraph of H induced by all the row vertices and the column vertices $1, 2, \ldots, r$, and c. Lemma 2.11 says that there is a path $\mathcal{P}$ from c to r in the undirected graph $G_{\cap}(H)$ whose intermediate vertices are all smaller than r. Thus $\mathcal{P}$ is also a path in $G_{\cap}(\overline{H})$. By Lemma 2.1, there is a path in $\overline{H}$ from column vertex c to column vertex r.



$$\begin{array}{c} \\ 1' \\ 2' \\ 3' \\ 4' \\ 5' \\ 6' \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 1 & & & \\ & & & & \\ 1 & & 1 & & \\ & & & & \\ & 1 & & & 1 \\ & & & & 1 \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 5 \end{array} \\ \begin{pmatrix} 2 & 1 & \\ 1 & 2 & 1 \\ 1 & & \end{pmatrix} \end{array}$$

FIG. 9. *Example for Theorem 3.8. Graph H is shown in Figure 1. Its column intersection graph and filled column intersection graph are shown in Figure 4. This figure shows the construction that makes entry $[R]_{35}$ nonzero. At left, graph $\overline{H}$ is the subgraph of H induced by column vertices 1 through $r = 3$ and $c = 5$, and all the row vertices. The dashed edges are a column-complete matching $\mathcal{M}$ with respect to which there is a c-alternating path $\mathcal{Q} = (5, 5', 2, 1', 1, 3', 3)$ from c to r. At center, A is chosen to have ones in positions $\mathcal{M}$ and $\mathcal{Q}$ and zeros elsewhere. At right, K is the submatrix of A^TA consisting of rows and columns 1 through $r - 1 = 2$, as well as row $r = 3$ and column $c = 5$. Matrix K is a permutation of a triangular matrix with nonzero diagonal and hence cannot be singular.*

Now $\overline{H}$ is strong Hall because H is. Therefore the alternating-paths theorem (Theorem 2.9) applies, and says that there is a column-complete matching $\mathcal{M}$ for $\overline{H}$ and a path $\mathcal{Q}$ from c to r that is c-alternating with respect to $\mathcal{M}$.

Choose the values of those nonzeros of A corresponding to edges of $\mathcal{M} \cup \mathcal{Q}$ to be 1, and choose the values of the other "nonzeros" to be 0. Let us examine the $r \times r$ submatrix K of A^TA defined above. (For simplicity, we will call the last column of K number c rather than number r; the last row of K is number r.) We claim that the bipartite graph $H(K)$ has exactly one perfect matching (or, equivalently, that K can be permuted to a triangular matrix with nonzero diagonal). To prove this, we match rows of K greedily to columns of K. Take a column j of K. If j is a vertex that is not on path $\mathcal{Q}$, then the only nonzero in column j of K is $[K]_{jj}$, and we match column j to row j'. If j is on $\mathcal{Q}$, i' is the vertex following j on $\mathcal{Q}$, and k is the vertex following i' on $\mathcal{Q}$, then $[K]_{kj}$ is nonzero and we match column j to row k'. (The last vertex on $\mathcal{Q}$ is column r, which is not a column of K.) This is a perfect matching on $H(K)$. Its uniqueness follows by induction on the length of $\mathcal{Q}$, the induction step being the fact that column c of K has only one nonzero (because row c' is not a row of K).

This proves the claim that $H(K)$ has exactly one perfect matching. Thus the determinant of K is just the product of the nonzero values corresponding to elements of that matching, and is itself nonzero. This shows that the polynomial p_{rc} is nonzero for at least one point, that is, for at least one choice of values for A.

Now the set of zeros of a k-variable polynomial has measure zero in $\mathbf{R}^k$, unless the polynomial is identically zero. Thus not only do values for the nonzero entries of A exist that make p_{rc} and hence $[R]_{rc}$ nonzero, but almost all choices of values (in the measure-theoretic sense) work. Therefore, almost all choices of values for A make every $[R]_{rc}$ nonzero simultaneously. Furthermore, almost all of those choices include no zero values; that is, for almost all such choices, $H(A) = H$ as desired. Finally, we observe that we can choose A to have full rank n: for some $n \times n$ submatrix of A there is a choice of values that gives nonzero determinant (namely, ones for the elements of a column-complete matching of H and zeros elsewhere), and hence almost all choices of values make that submatrix nonsingular. □

COROLLARY 3.9. *If H is strong Hall and has nonzero diagonal, then the upper triangular parts of $G^\times(H)$ and $G^+_\cap(H)$ are equal.*

Proof. By Theorem 3.6 and its corollary we have $G(R) \subseteq G^\times(H) \subseteq G^+_\cap(H)$ for any $A = QR$ with $H(A) = H$. If we choose A as in Theorem 3.8, the first and third graphs are equal, and hence the second and third are also equal. □

COROLLARY 3.10. *If H is strong Hall and has nonzero diagonal, then there is a matrix A with full column rank and with $H(A) = H$, such that the orthogonal factorization $A = QR$ satisfies $G(R) = G^\times(H)$.* □

3.4. Remarks on orthogonal factorization. Theorem 3.8 gives a tight prediction of the structure of R in QR factorization, in the exact sense, provided that A is strong Hall. Recently, Hare, Johnson, Olesky, and van den Driessche [21] extended this result significantly by giving a tight exact characterization of the structures of both Q and R, under the weaker assumption that A is Hall—that is, that A is structurally of full column rank. The Hare et al. characterization uses a notion called "Hall sets," which concerns strong Hall submatrices of A and is related to the Dulmage-Mendelsohn decomposition of $H(A)$. Hare et al. proved that their structure prediction was one-at-a-time exact; Pothen [28] then showed that in fact it is all-at-once exact. Both Hare et al. and Pothen used versions of the alternating-paths theorem in their work.

Theorem 3.4 gives a tight prediction of the structure of A at each step of column QR factorization, in the symbolic sense. This prediction is not tight in the exact sense; see Coleman et al. [5] for an example. It is an open problem to give a tight exact structure prediction for each A_i in column factorization. The techniques of Hare et al. [21] are probably relevant here.

Recently, Ng and Peyton [26] investigated the structure of the so-called matrix of Householder vectors. This is a representation of Q in which the vector that generates the i-th Householder reflection is stored in place of the i-th column of Q. Ng and Peyton gave a tight exact prediction of the structure of this matrix in the case that A is either strong Hall or has its columns permuted according to a Dulmage-Mendelsohn decomposition.

Givens rotations can be used to introduce zeros in other orders than row by row or column by column; examples are reductions of symmetric sparse matrices to tridiagonal form [33] and the Jacobi algorithm for finding eigenvalues [18]. Little work exists on structure prediction for such problems. For example, it would be interesting to prove upper and lower bounds on the work required to tridiagonalize a symmetric matrix A by Givens rotations, in terms of the structure $G(A)$.

4. *LU* factorization with partial pivoting. Let A be a nonsingular $n \times n$ matrix. The triangular factorization $A = LU$ does not always exist, and is not always numerically stable when it does exist [18, Chapter 3]. Thus some form of row or column interchanges are needed in Gaussian elimination; at each step, a nonzero must be brought into the pivotal position before elimination.

In the dense setting, the pivot is usually chosen as the element of largest magnitude in the current column (partial pivoting) or in the entire uneliminated matrix (complete pivoting). In the sparse setting, there are several strategies for choosing pivots to combine stability and sparsity. Some variations of complete pivoting choose a pivot at each step to minimize operation count from among candidates that are not too far from maximum magnitude [6]. Another approach is to preorder the matrix columns purely to preserve sparsity, and then use partial pivoting to reorder the rows for stability [13, 16].

This section parallels Section 3 in outline. In Section 4.1, we review a graph model of Gaussian elimination with row and column interchanges, and we prove some results on the structure of the matrix during elimination. These results are symbolic; that is, they assume that zeros are introduced only by explicit elimination, not by cancellation. In Section 4.2 we give upper bounds on the structure of the factors L and U obtained by Gaussian elimination with row interchanges. In Section 4.3, we give an exact lower bound on L and U. This result is tight—that is, best possible—and is the main new result of this paper. We conclude the section with remarks and open problems.

We write LU factorization with row and column interchanges as follows.

$$\begin{aligned} A_0 &= A, \\ P_1^r A_0 P_1^c &= L_1 A_1, \\ P_2^r A_1 P_2^c &= L_2 A_2, \\ &\vdots \\ P_{n-1}^r A_{n-2} P_{n-1}^c &= L_{n-1} A_{n-1} = L_{n-1} U. \end{aligned}$$

Here P_i^r is an $n \times n$ elementary permutation matrix corresponding to the row interchange at step i, P_i^c is an $n \times n$ elementary permutation matrix corresponding to the column interchange at step i, L_i is an $n \times n$ elementary lower triangular matrix whose i-th column contains the multipliers at step i, and U is an $n \times n$ upper triangular matrix. Since each elementary permutation matrix (P_i^r or P_i^c) is its own inverse, we can write the final factorization as

$$A = P_1^r L_1 P_2^r L_2 \cdots P_{n-1}^r L_{n-1} U P_{n-1}^c \cdots P_2^c P_1^c. \tag{1}$$

We define L as the $n \times n$ matrix whose i-th column is the i-th column of L_i, so that $L - I = \sum_i (L_i - I)$. Note a subtle point about L: we can also think of Gaussian elimination as computing a factorization $P^r A P^c = L^\circ U$, but this L° is not the same as L. The two matrices are both unit lower triangular, and they contain the same nonzero values, but in different positions; L° has its rows in the order described by the entire row pivoting permutation, while L has the rows of its i-th column in the order described by only the first i interchanges. The matrix L is essentially a data structure for storing L°; either can be used in solving systems of equations. The structure prediction results in Sections 4.2 and 4.3 below will be about L, not L°.

Note also that our notation is slightly different than in the previous section: now A_i is always $n \times n$, not $(n-i) \times (n-i)$.

4.1. Nonzero structure of A during elimination. In this subsection we develop a symbolic model of Gaussian elimination with row and/or column interchanges. The model is based on that of Golumbic [19] and Gilbert [17]. Theorem 4.2 is new.

Let $H_0 = H(A)$ be the bipartite graph of $A = A_0$. Assume $[A_0]_{rc}$ is nonzero and is chosen as pivot at step 1. Define the *deficiency* of the edge $\langle r', c\rangle$ of H_0 to be the set of edges

$$\{\langle i', j\rangle : c \in Adj_{H_0}(i'), j \in Adj_{H_0}(r'), \text{ and } j \notin Adj_{H_0}(i')\}.$$

We obtain the bipartite graph H_1 of the $(n-1) \times (n-1)$ submatrix that remains after eliminating $\langle r', c\rangle$ as follows: delete from H_0 vertices r' and c and all edges incident on them, then add the edges in the deficiency of $\langle r', c\rangle$. The edges in the deficiency of $\langle r', c\rangle$ correspond to the zero elements of A_0 that become nonzero when $[A_0]_{r,c}$ is eliminated. (Note that the labelling of the vertices of H_1 refers to the labelling in the original matrix A_0.) Thus, given a sequence of pivot elements $\langle r'_1, c_1\rangle, \langle r'_2, c_2\rangle, \cdots, \langle r'_{n-1}, c_{n-1}\rangle$ (some of which may be fill edges), we can follow the recipe above to construct a sequence of bipartite graphs $H_0, H_1, \cdots, H_n$, where H_i describes the structure of the $(n-i) \times (n-i)$ Schur complement remaining after step i.

It is possible to prove bipartite versions of several of the results from Section 2.5. We will use the following lemma in the exact lower bound proof later in this section.

LEMMA 4.1. *Let A be a square matrix, and let $\mathcal{M}$ be a perfect matching on $H(A)$. Let $H_0, \ldots, H_n$ be the sequence of bipartite elimination graphs described above, when elimination is carried out by pivoting on the edges of $\mathcal{M}$. If $\langle r', c\rangle$ is a non-matching edge of H_i, then there is a path from r' to c in $H(A)$ that is r-alternating with respect to $\mathcal{M}$, and whose intermediate vertices are all endpoints of edges of $\mathcal{M}$ eliminated at or before step i.*

Proof. We induce on the smallest i such that $\langle r', c\rangle$ is an edge of H_i. If $i = 0$ then $\langle r', c\rangle$ itself is the path. Otherwise, $\langle r', c\rangle$ is in the deficiency of the matching edge $\langle r'_i, c_i\rangle$ in H_{i-1}, so edges $\langle r', c_i\rangle$ and $\langle r'_i, c\rangle$ are non-matching edges of H_{i-1}. Applying the induction hypothesis to those edges, we get r-alternating paths $\mathcal{P}$ from r' to c_i and $\mathcal{Q}$ from r'_i to c in $H(A)$. Then $\mathcal{P}\langle c_i, r'_i\rangle\mathcal{Q}$ is an r-alternating walk from r' to c in $H(A)$ whose intermediate vertices are all eliminated at or before step i. Thus there exists an r-alternating path with the same property. □

One interesting fact about symbolic bipartite elimination, which is new and is stated below as a theorem, is that it preserves the Hall and strong Hall properties.

THEOREM 4.2. *Let H_0 be a bipartite graph and let $\langle r', c\rangle$ be an edge of H_0. Let H_1 be the bipartite graph resulting from the elimination of edge $\langle r', c\rangle$. If H_0 has the Hall property, then H_1 also has the Hall property. If H_0 has the strong Hall property, then H_1 also has the strong Hall property.*

Proof. Recall Theorem 2.5, which says that an $m \times n$ bipartite graph is Hall if and only if it has no independent set of more than m vertices, and strong Hall if and only if it has no independent set of exactly m vertices that includes at least one vertex from each part.

Let R_1 and C_1 be the row and column vertices in a largest independent set in H_1. It is not possible that both $r' \in Adj_{H_0}(C_1)$ and $c \in Adj_{H_0}(R_1)$, for that would imply an edge between R_1 and C_1 in H_1. Therefore either $R_1 \cup C_1 \cup \{r'\}$ or $R_1 \cup C_1 \cup \{c\}$ is an independent set in H_0. If H_0 is Hall, that set has size at most m, and hence $R_1 \cup C_1$ has size at most $m-1$, so H_1 is also Hall. The strong Hall case follows the same argument, considering only independent sets that include both rows and columns. □

4.2. Upper bounds on L and U with partial pivoting. For the remainder of this section, we restrict our attention to the case in which only row interchanges are performed during Gaussian elimination, so the column ordering is fixed initially. This subsection proves symbolic upper bounds on the structures of L and U, making no assumptions on the row pivoting strategy. For the case where A is strong Hall and rows are ordered by partial pivoting, the next subsection proves matching exact lower bounds. Therefore the symbolic upper bound is in fact a tight exact bound in this case. As we will see, the tight exact bound is a one-at-a-time result; there is no tight all-at-once bound on L and U in general.

In the rest of this section we require A to have a nonzero diagonal. The rows of any nonsingular square matrix can be permuted to put nonzeros on the diagonal (by Theorem 2.3 and Corollary 2.4). In fact, only the bounds on L below depend on a nonzero diagonal; the bounds on U hold for arbitrary nonsingular A.

Since the row interchanges P_i^r depend on the numerical values, it is in general impossible to determine where fill will occur in L and U from the structure of A. George and Ng [13] suggested a way to get an upper bound on possible fill locations. At step i of Gaussian elimination with row interchanges, call the rows that have nonzeros in column i below the diagonal *candidate pivot rows*. George and Ng observed that fill can only occur in candidate pivot rows, and only in columns that are nonzero in some candidate pivot row. Thus the structure that results from the elimination step is bounded by replacing each candidate pivot row by the union of all the candidate pivot rows (to the right of column i). We need the fact that the diagonal of A is nonzero to argue that this models the effect of row interchanges correctly: row i is itself a candidate pivot row at step i, and therefore interchanging row i with another candidate pivot row does not affect the structure of the bound.

This procedure for bounding the structures of L and U is precisely the construction of the row merge graph from Section 3. Therefore we have the following theorem. (Note that $G^\times(A) = H^\times(A)$ since A is square.)

THEOREM 4.3 (GEORGE AND NG [13]). *Let A be a nonsingular square matrix with nonzero diagonal. Suppose A is factored by Gaussian elimination with row interchanges as*

$$A = P_1 L_1 P_2 L_2 \cdots P_{n-1} L_{n-1} U,$$

and L is the union of the L_i as described above. Then

$$G(L+U) \subseteq G^\times(A),$$

that is, the structures of L and U are subsets of the lower and upper triangles of the row merge graph of A. □

COROLLARY 4.4. *Let A be a nonsingular square matrix with nonzero diagonal, factored by Gaussian elimination with row interchanges as in Theorem 4.3. Then*

$$G(L+U) \subseteq G_\cap^+(A),$$

that is, the structures of L and U are subsets of the lower and upper triangles of the (symmetric) filled column intersection graph of A. □

George, Liu, and Ng [10, 13] gave an algorithm for Gaussian elimination with partial pivoting that uses $G^\times(A)$ to build a data structure to hold the factors of A as elimination progresses. The structure may be overgenerous in the sense that it stores some zeros, but it has the advantage that it is static; the structure does not change as pivoting choices are made. George, Liu, and Ng's numerical experiments indicated that (with a judicious choice of a column reordering for sparsity) the total storage and execution time required to compute the LU decomposition using the static data structure were quite competitive with other approaches.

4.3. Lower bounds on L and U with partial pivoting. In this section we show that Theorem 4.3 is tight in the exact sense for strong Hall A. In other words, if a given input structure is strong Hall, then for every edge of the row merge graph there is a way to fill in the values so that the corresponding position of L or U is

$$
\begin{array}{c|cccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
1' & \times & \times & & & & \\
2' & & \times & & & & \times \\
3' & \times & & \times & \times & & \\
4' & & & \times & \times & & \times \\
5' & & \times & & & \times & \\
6' & & & & \times & \times & \times
\end{array}
\qquad
\begin{array}{c|cccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
1' & \times & \times & \times & \times & & \\
2' & & \times & \times & \times & \times & \times \\
3' & \times & \times & \times & \times & \times & \times \\
4' & & & \times & \times & \times & \times \\
5' & & \times & \times & \times & \times & \times \\
6' & & & & \times & \times & \times
\end{array}
$$

FIG. 10. *Example for Theorem 4.5. On the left is a matrix A. On the right is the bound $G^{\times}(A)$ on the structures of L and U. In the case $r < c$, Figure 11 shows how to make $[U]_{35}$ nonzero. In the case $r > c$, Figure 12 shows how to make $[L]_{54}$ nonzero.*

nonzero. This implies that George and Ng's static data structure [13] is the tightest possible for Gaussian elimination with partial pivoting. This is a one-at-a-time result; as we will see, no all-at-once result is possible.

The case $r < c$ of Theorem 4.5 (that is, the proof for U) first appeared in a technical report by Gilbert [15]; the case $r > c$ (for L) has not appeared before. (Gilbert actually related U to $G^{+}_{\cap}(A)$ rather than $G^{\times}(A)$, but the U parts of those graphs are the same for strong Hall A by Corollary 3.9.)

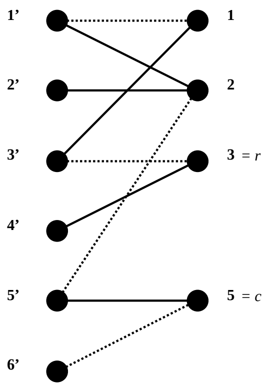
THEOREM 4.5. *Let H be the structure of a square strong Hall matrix with nonzero diagonal. Let $\langle r, c\rangle$ be an edge of the row merge graph $G^{\times}(H)$. There exists a nonsingular matrix A (depending on r and c) with $H(A) = H$, such that if A is factored by Gaussian elimination with partial pivoting into L and U as described in Theorem 4.3, then $[L+U]_{rc} \neq 0$.*

Proof. Figure 10 shows an example of the bound. The cases $r < c$ (that is, U) and $r > c$ (that is, L) are similar. Row interchanges make the L case a little more complicated; thus we prove the two cases separately.

Case $r < c$ (structure of U). Figure 11 illustrates this case. According to Corollary 3.5, there is a path $\mathcal{P}$ in H from row vertex r' to column vertex c whose intermediate column vertices are all at most r.

Let $\overline{H}$ be the subgraph of H induced by all the row vertices and the column vertices 1 through r and c. Now $\overline{H}$ is strong Hall because H is. Therefore the alternating-paths theorem (Theorem 2.9) applies, and says that there is a column-complete matching $\mathcal{M}$ for $\overline{H}$ and a path $\mathcal{Q}$ from c to r that is c-alternating with respect to $\mathcal{M}$.

Choose the values of those nonzeros of A corresponding to edges of $\mathcal{M}$ to be larger than n, and the values of the other nonzeros of A to be between 0 and 1. Further, choose the values so as to make every square submatrix of A that is Hall, including A itself, nonsingular. (Such a choice is possible by an argument like that in Theorem 3.8: the determinant of a Hall submatrix is a polynomial in its nonzero values, not identically zero because the Hall property implies a perfect matching. Therefore the set of values that make any Hall submatrix singular has measure zero, and can be avoided.)



$$\begin{array}{c} \\ 1' \\ 2' \\ 3' \\ 4' \\ 5' \\ 6' \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} 9 & x & & & & \\ & x & & & & x \\ x & & 9 & x & & \\ & & x & x & & x \\ & 9 & & & x & \\ & & & x & 9 & x \end{pmatrix} \end{array}$$

$$\begin{array}{c} \\ 1' \\ 5' \\ 3' \\ ?' \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 5 \end{array} \\ \begin{pmatrix} 9 & x & & \\ & 9 & & x \\ x & & 9 & * \\ ? & ? & ? & ? \end{pmatrix} \end{array}$$

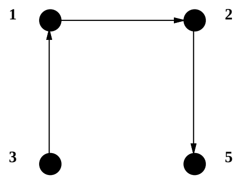
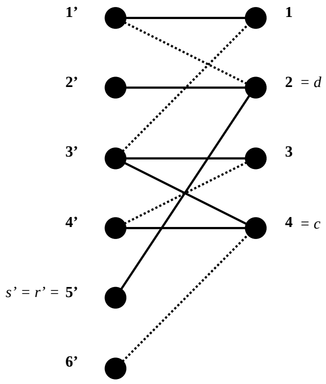


FIG. 11. *Example for Case 1 of Theorem 4.5, showing the construction that makes* $[U]_{35}$ *nonzero in the structure from Figure 10. At top left, the graph* $\overline{H}$ *is the subgraph of* H *induced by column vertices 1 through* $r = 3$ *and* $c = 5$*, and all the row vertices. The dashed edges are a column-complete matching* $\mathcal{M}$ *with respect to which there is a c-alternating path* $\mathcal{Q} = (5, 5', 2, 1', 1, 3', 3)$ *from* c *to* r*. At top right,* A *is chosen to have large values in positions* $\mathcal{M}$ *and small values elsewhere. At bottom left,* A_r *is the submatrix of* PA *with columns 1 through* r *and* c *and the rows in the corresponding positions after 3 pivot steps. The element* $[U]_{35}$ *is in position* $*$ *of the factor of* A_r*. At bottom right, the directed graph* $G(A_r)$ *has a path* $(3, 1, 2, 5)$*; therefore* $\langle 3, 5\rangle$ *fills in. In the original* A*, the first pivot step does no row swap and fills position* $(3', 2)$*; the second pivot step swaps rows* $2'$ *and* $5'$ *and fills position* $(3', 5)$*.*



$$
\begin{array}{c}
 \\ 1' \\ 2' \\ 3' \\ 4' \\ 5' \\ 6'
\end{array}
\begin{array}{c}
\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{pmatrix}
x & 9 & & & & \\
 & x & & & & x \\
9 & & x & x & & \\
 & & 9 & x & & x \\
 & x & & & x & \\
 & & & 9 & x & x
\end{pmatrix}
\end{array}
$$

$$
\begin{array}{c}
 \\ 3' \\ 1' \\ 4' \\ 6' \\ 5'
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
9 & & x & x & \\
x & 9 & & & \\
 & & 9 & x & \\
 & & & 9 & x \\
 & x & & * & x
\end{pmatrix}
\end{array}
$$

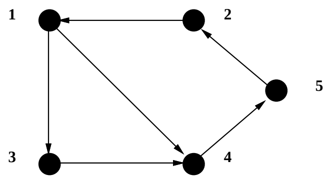


FIG. 12. *Example for Case 2 of Theorem 4.5, showing the construction that makes* $[L]_{54}$ *nonzero in the structure from Figure 10. At top left, the graph* $\overline{H}$ *is the subgraph of* H *induced by column vertices 1 through* $c = 4$*, and all the row vertices. Then* $d = 2$ *is the first column vertex on some path from* r' *to* c*. The dashed edges are a column-complete matching* $\mathcal{M}$ *with respect to which there is a* c*-alternating path* $\mathcal{Q} = (4, 4', 3, 3', 1, 1', 2)$ *from* c *to* d*. At top right,* A *is chosen to have large values in positions* $\mathcal{M}$ *and small values elsewhere. At bottom left,* A_c *is the submatrix of* PA *with columns 1 through* c *and* r *and the rows in the corresponding positions after 4 pivot steps. The element* $[L]_{54}$ *is in position* $*$ *of the factor of* A_c*. The fifth and last row of* A_c *is* $5'$*, the fifth row of* A*, because* $5'$ *was not involved in a pivoting swap during the first 4 steps; therefore* $s' = r' = 5'$ *and the argument about an alternating path from* r' *to* s' *is not needed in this example. At bottom right, the directed graph* $G(A_c)$ *has a path* $(5, 2, 1, 4)$*; therefore* $\langle 5, 4\rangle$ *fills in. In the original* A*, the first pivot step fills position* $(1', 4)$*, and the second pivot step fills position* $(5', 4)$*.*

Now we prove that this choice of values makes $[U]_{rc}$ nonzero. In the first r steps of elimination of A, the pivot elements are nonzeros corresponding to edges of $\mathcal{M}$. Let P be the permutation matrix that describes the first r row interchanges (that is, $P = P_r P_{r-1} \cdots P_1$ in Theorem 4.3). Let A_r be the $(r+1)\times(r+1)$ principal submatrix of PA that includes the first r columns and column c, and the corresponding rows. Thus the columns of A_r are those numbered 1 through r and c in H; the first r rows of A_r are those matched to columns 1 through r of H by $\mathcal{M}$; and it does not matter which row of H the last row of A_r is. We will consider the rows and columns of the bipartite graph $H(A_r)$ to have the same numbers that they did in H; thus the column vertex numbers are 1 through r and c, and the row numbers may be anything. In the directed graph $G(A_r)$, we will also number the vertices 1 through r and c, but bear in mind that the row of A_r corresponding to a vertex v was not necessarily row v' in H.

Now the first r diagonal elements of A_r are nonzero, and dominant. Let L_r and U_r be the triangular factors of A_r without pivoting, $A_r = L_r U_r$. Then the element $[U]_{rc}$ mentioned in the statement of the theorem is in fact $[U_r]_{rc}$, the element in the last column and next-to-last row of U_r. We proceed to show that $[U_r]_{rc} \neq 0$.

All square Hall submatrices of A_r are nonsingular; thus, by Lemma 2.13, $G^{+}(A_r)$ is exactly the structure of $[L_r + U_r]$. Therefore $[U]_{rc}$ is nonzero if and only if $G(A_r)$ contains a directed path from vertex r to vertex c, through vertices numbered less than r.

Recall the path $\mathcal{Q}$, which is a path in $\overline{H}$ from c to r that is c-alternating with respect to $\mathcal{M}$. The matching $\mathcal{M}$ consists of exactly the edges on the diagonal of A_r (except for the one in the last column, which cannot be an edge of $\mathcal{Q}$ because $\mathcal{Q}$ is c-alternating). Therefore $\mathcal{Q}$ corresponds to a directed path from r to c in $G(A_r)$. Every vertex of $G(A_r)$ except r and c is numbered less than r, so this is the desired directed fill path and the proof of this case is complete.

Note that the proof never explicitly identified the row of H that ended up in position (r, c) of U; it is the row matched to column r by $\mathcal{M}$, and is the second last vertex on the path $\mathcal{Q}$.

Case $r > c$ (structure of L). Figure 12 illustrates this case. The proof for this case is much like that for U, but it needs to do some extra work to identify the row of H that ends up in position (r, c) of L, because that row has not yet been matched (pivoted on) when L_{rc} is computed.

Again by Corollary 3.5, there is a path $\mathcal{P}$ in H from row vertex r' to column vertex c whose intermediate column vertices are all at most c. Let d be the first column vertex on $\mathcal{P}$ (this is the vertex after r' on $\mathcal{P}$; possibly $d = c$).

Let $\overline{H}$ be the subgraph of H induced by all the row vertices and the column vertices 1 through c. (This has one less column than in the proof for U.) Then $\mathcal{P}[d\,{:}\,c]$ is a path (possibly of length 0) in $\overline{H}$ from column vertex d to column vertex c in $\overline{H}$. Again, therefore, there is a column-complete matching $\mathcal{M}$ for $\overline{H}$ and a path $\mathcal{Q}$ from c to d that is c-alternating with respect to $\mathcal{M}$.

Again we choose A so that edges of $\mathcal{M}$ have values larger than n, other edges have values between 0 and 1, and every square Hall submatrix of A is nonsingular.

The first c steps of elimination of A pivot on nonzeros corresponding to edges of $\mathcal{M}$. Let P be the permutation matrix that describes the first c row interchanges (that is, $P = P_c P_{c-1} \cdots P_1$ in Theorem 4.3). Let A_c be the $(c+1)\times(c+1)$ principal submatrix of PA that includes the first c columns and column r, and the rows in corresponding positions of PA. Thus the columns of A_c are those numbered 1 through c and r in H; the first c rows of A_c are those matched to columns 1 through c of H

by $\mathcal{M}$. The last row of A_c is some row number s' in H that is not matched by $\mathcal{M}$. (Row s' may or may not be matched to column r in the final factorization of A.)

Again, we give the rows and columns of the bipartite graph $H(A_c)$ the same numbers they had in H; the column vertex numbers are 1 through c and r, and the row numbers may be anything (but the last row is s'). In the directed graph $G(A_c)$, we will also number the vertices 1 through c and r; again, bear in mind that the row of A_c corresponding to a vertex v was not necessarily row v' in H, and in particular the row corresponding to vertex r of $G(A_c)$ is row s' of H.

Now the first c diagonal elements of A_c are nonzero, and dominant. Let L_c and U_c be the triangular factors of A_c without pivoting, $A_c = L_c U_c$. The element $[L]_{rc}$ mentioned in the statement of the theorem is in fact $[L_c]_{rc}$, the element in the last row and next-to-last column of L_c.

As before, we show that $[L_c]_{rc} \neq 0$ by exhibiting a directed path from vertex r to vertex c of $G(A_c)$, based on a c-alternating path in H. However there is not necessarily an edge between column vertex r and row vertex s' in H; thus we must find a c-alternating path that ends at s', not r. The details of how to do that will complete the proof.

We now trace the pivoting process to discover where row s' came from. If row r' of H was not used as one of the first c pivots, then it has not moved and $s' = r'$. If row r' was used as a pivot, suppose it was in column $c_1 \leq c$, and that the row interchanged with r' at step c_1 was row r'_1. (Recall that all row and column numbers are vertex numbers of H.) Again, either $r'_1 = s'$ or else r'_1 was later used as a pivot in some column $c_2 > c_1$, when it was interchanged with some row r'_2. Continuing inductively, we eventually arrive at a row r'_k which is equal to s', which was not used as a pivot in the first c steps.

The sequence of nonzeros we followed while tracing the pivoting process was

$$r' = r'_0, c_1, r'_1, c_2, r'_2, \ldots, c_{k-1}, r'_{k-1}, c_k, r'_k = s'.$$

Each $\langle c_i, r'_i \rangle$ is an edge of one of the bipartite elimination graphs $\overline{H}_0, \overline{H}_1, \ldots, \overline{H}_c$ corresponding to the first c steps of symbolic Gaussian elimination of $\overline{H}$. Therefore, by Lemma 4.1, there is a c-alternating path in $\overline{H}$ from c_i to r'_i for each i. Furthermore each $\langle r'_{i-1}, c_i \rangle$ is an edge of $\mathcal{M}$, and is thus a one-edge c-alternating path from r'_{i-1} to c_i. Concatenating these paths yields a c-alternating walk $\mathcal{W}$ (which may repeat vertices or edges) from r' to s' in $\overline{H}$.

Now if edge $\langle d, r' \rangle$ is not an edge of $\mathcal{M}$, then $\mathcal{Q}$ followed by $\langle d, r' \rangle$ followed by $\mathcal{W}$ is a c-alternating walk from column c to row s'. Alternatively, if $\langle d, r' \rangle$ is an edge of $\mathcal{M}$, then $d = c_1$, and $\mathcal{Q}$ followed by $\mathcal{W}[d : s']$ is a c-alternating walk from column c to row s'. Either way, we have a walk in $\overline{H}$ from c to s' that is c-alternating with respect to $\mathcal{M}$. This walk corresponds to a directed walk from vertex r to vertex c of $G(A_c)$. Thus there is a directed path from vertex r to vertex c of $G(A_c)$. The intermediate vertices on this path are less than both r and c, because r and c are the last two vertices of $G(A_c)$. Therefore $\langle r, c \rangle$ is an edge of $G^{+}(A_c)$. Since all square Hall submatrices of A_c are nonsingular, therefore, $[L_c]_{rc}$ is nonzero. Thus $[L]_{rc}$ is nonzero and the proof is complete. □

4.4. Remarks on LU factorization with pivoting. Theorem 4.5 showed that $G^{\times}(A)$ is a tight exact bound on the structure of the factors L and U, assuming that the structure of A is not only strong Hall, but also has its rows permuted so that the diagonal is nonzero. We can get a tight exact bound on U without assuming a nonzero diagonal. The following result does not depend on row ordering.

Corollary 4.6. *Let H be a square bipartite graph with the strong Hall property. Let $\langle r, c\rangle$ be an edge of the filled column intersection graph $G_\cap^+(H)$. Then there is a nonsingular matrix A (depending on r and c) with $H(A) = H$, such that the upper triangular factor U of A in Gaussian elimination with partial pivoting has $[U]_{rc} \neq 0$.*

Proof. Since H is strong Hall, it has a column-complete matching. Let $\overline{H}$ be H with its row vertices permuted so that $\langle i', i\rangle$ is a matching edge for all i. The filled column intersection graph is independent of the row permutation, so $G_\cap^+(H) = G_\cap^+(\overline{H})$. Corollary 3.9 says that the upper triangles of $G_\cap^+(\overline{H})$ and $G^\times(\overline{H})$ are the same. Therefore $\langle r, c\rangle$ is an edge of $G^\times(\overline{H})$. Then, by Theorem 4.5, there is a nonsingular matrix $\overline{A}$ with $H(\overline{A}) = \overline{H}$, such that the upper triangular factor $\overline{U}$ of $\overline{A}$ in Gaussian elimination with partial pivoting has $[\overline{U}]_{rc} \neq 0$.

By a measure-theoretic argument like that in Theorem 3.8, we can choose $\overline{A}$ so that there is never a tie for the choice of pivot element, that is, so that at each elimination step all the subdiagonal nonzeros of the pivot column have different magnitudes. Under this assumption, the upper triangular factor $\overline{U}$ is independent of the row ordering of $\overline{A}$. Let A be $\overline{A}$ with its rows permuted so $H(A) = H$. The upper triangular factor U of A is equal to $\overline{U}$, and hence $[U]_{rc} \neq 0$. ◻

Theorem 4.5 on LU differs from Theorem 3.8 on QR in that the latter is all-at-once; that is, for each structure a single matrix exists that fills all the predicted nonzeros. Theorem 4.5 is not all-at-once, and no tight exact all-at-once result is possible for LU factorization with partial pivoting. To see this, consider a matrix that is tridiagonal plus a full first column,

$$A = \begin{pmatrix} \times & \times & & & \\ \times & \times & \times & & \\ \times & \times & \times & \times & \\ \times & & \times & \times & \times \\ \times & & & \times & \times \end{pmatrix}.$$

The graph $H(A)$ is strong Hall. The row merge graph $G^\times(A)$ is full. As Theorem 4.5 says, any single position in L or U can be made nonzero by an appropriate choice of pivots. But the first row of U will have the same structure as some row of A, so it is impossible for U to be full.

One application of structure prediction for partial pivoting is to predict which columns of A will update which other columns if the factorization is done with a column-by-column algorithm. For example, Gilbert [15] gave a parallel implementation of LU factorization with partial pivoting in which tasks (columns of the factorization) were scheduled dynamically to processors, based on a precedence relationship determined by precomputing the elimination tree [23] of $G_\cap(A)$. Since $[U]_{ij}$ is nonzero if and only if column i updates column j during the factorization, a corollary of Theorem 4.5 is that, for strong Hall A, this is the tightest prediction possible from the structure of A alone.

Corollary 4.7 (Gilbert [15]). *Let a strong Hall structure for the square matrix A be given. If k is the parent of j in the elimination tree of $G_\cap(A)$, then there exists a choice of nonzero values of A that will make column j update column k during factorization with partial pivoting.* ◻

This corollary is a one-at-a-time result. However, if we restrict our attention to the edges of the elimination tree of $G_\cap(A)$ instead of all of $G^\times(A)$, it may be possible to prove an all-at-once result. We conjecture that for every square strong Hall matrix H, there exists a single matrix A with $H(A) = H$ such that every edge of the elimination tree of $G_\cap(A)$ corresponds to a nonzero in the upper triangular factor U of A with partial pivoting.

Little if anything is known about the case when $H(A)$ is not strong Hall. Hare et al. [21] gave a complete exact result for QR factorization assuming only the Hall property; is a similar analysis possible for partial pivoting? In particular, since the upper triangles of $G^\times(A)$ and $G_\cap^+(A)$ can differ in the non-strong Hall case, how tight is the former for partial pivoting? There are non-strong Hall structures for which $G^\times(A)$ is tight but $G_\cap^+(A)$ is not; an example is a matrix whose only nonzeros are the diagonal and the first row.

5. Remarks. The theme of this paper is that, when solving a nonsymmetric linear system, structure prediction is easier if the matrix is strong Hall. On the other hand, a system whose matrix is not strong Hall can be partitioned (by Dulmage-Mendelsohn decomposition) into smaller strong Hall systems. This useful coincidence makes some intuitive sense. Symbolic independence of vectors (the Hall property) is a weaker condition than numeric linear independence. In a sense, Dulmage-Mendelsohn decomposition tries to wring as much as possible out of symbolic relationships before Gaussian elimination takes over to handle numeric relationships; the tight exact (i.e. numeric) lower bounds in this paper say that Dulmage-Mendelsohn decomposition is doing its job.

Predicting structure in algorithms that combine numerical and structural information is an interesting challenge. Murota et al. [25] have studied block triangular decompositions that take some but not all of the numerical values into account.

We point out once more that Hare, Johnson, Olesky, van den Driessche, and Pothen [21, 28] have recently obtained tight exact bounds on both Q and R in the general Hall case, thus extending the work of Coleman, Edenbrandt, and Gilbert that we reviewed in Section 3. It would be interesting to see whether our bounds on L and U for partial pivoting, in Section 4, could be similarly extended.

We conclude by mentioning three open problem areas for nonsymmetric structure prediction.

First, it would be interesting to understand the relationship between the structure of L and the structure of L°, both of which are different ways of storing the lower triangular factor in Gaussian elimination with partial pivoting. Can the techniques discussed in this paper be used to obtain bounds on the structure of L°?

Second, it would be useful to achieve a complete structural understanding of the Bunch-Kaufmann symmetric indefinite factorization [18, Chapter 4.4]. Here a symmetric indefinite matrix is factored symmetrically by choosing pivots from the diagonal, but each pivot may be either an element or a 2×2 submatrix. Thus the factorization is $PAP^T = LDL^T$, where P is a permutation, L is lower triangular, and D is block diagonal with 1×1 and 2×2 blocks. This factorization is particularly useful for solving "augmented systems" of the form

$$\begin{pmatrix} K & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix},$$

where A is rectangular and K is symmetric and (perhaps) positive definite [1]. Even the common case $K = I$ is not well understood.

Third, it would be interesting to understand the structural issues in the incomplete LU factorizations sometimes used to precondition iterative methods for solving linear systems [7].

REFERENCES

[1] Åke Björck. A note on scaling in the augmented system methods, 1991. Unpublished manuscript.

[2] Robert K. Brayton, Fred G. Gustavson, and Ralph A. Willoughby. Some results on sparse matrices. *Mathematics of Computation*, 24:937–954, 1970.

[3] Richard A. Brualdi and Herbert J. Ryser. *Combinatorial Matrix Theory*. Cambridge University Press, 1991.

[4] Richard A. Brualdi and Bryan L. Shader. Strong Hall matrices. IMA Preprint Series #909, Institute for Mathematics and Its Applications, University of Minnesota, December 1991.

[5] Thomas F. Coleman, Anders Edenbrandt, and John R. Gilbert. Predicting fill for sparse orthogonal factorization. *Journal of the Association for Computing Machinery*, 33:517–532, 1986.

[6] I. S. Duff and J. K. Reid. Some design features of a sparse matrix code. *ACM Transactions on Mathematical Software*, 5:18–35, 1979.

[7] Howard Elman. A stability analysis of incomplete LU factorization. *Mathematics of Computation*, 47:191–218, 1986.

[8] Alan George and Michael T. Heath. Solution of sparse linear least squares problems using Givens rotations. *Linear Algebra and its Applications*, 34:69–83, 1980.

[9] Alan George and Joseph Liu. Householder reflections versus Givens rotations in sparse orthogonal decomposition. *Linear Algebra and its Applications*, 88:223–238, 1987.

[10] Alan George, Joseph Liu, and Esmond Ng. A data structure for sparse QR and LU factorizations. *SIAM Journal on Scientific and Statistical Computing*, 9:100–121, 1988.

[11] Alan George and Joseph W. H. Liu. *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall, 1981.

[12] Alan George, Joseph W. H. Liu, and Esmond Ng. Row ordering schemes for sparse Givens transformations I. Bipartite graph model. *Linear Algebra and its Applications*, 61:55–81, 1984.

[13] Alan George and Esmond Ng. Symbolic factorization for sparse Gaussian elimination with partial pivoting. *SIAM Journal on Scientific and Statistical Computing*, 8:877–898, 1987.

[14] John R. Gilbert. Predicting structure in sparse matrix computations. Technical Report 86–750, Cornell University, 1986. To appear in *SIAM Journal on Matrix Analysis and Applications*.

[15] John R. Gilbert. An efficient parallel sparse partial pivoting algorithm. Technical Report 88/45052-1, Christian Michelsen Institute, 1988.

[16] John R. Gilbert and Tim Peierls. Sparse partial pivoting in time proportional to arithmetic operations. *SIAM Journal on Scientific and Statistical Computing*, 9:862–874, 1988.

[17] John Russell Gilbert. *Graph Separator Theorems and Sparse Gaussian Elimination*. PhD thesis, Stanford University, 1980.

[18] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, second edition, 1989.

[19] Martin Charles Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, 1980.

[20] Frank Harary. *Graph Theory*. Addison-Wesley Publishing Company, 1969.

[21] Donovan R. Hare, Charles R. Johnson, D. D. Olesky, and P. van den Driessche. Sparsity analysis of the QR factorization, 1991. To appear in *SIAM Journal on Matrix Analysis and Applications*.

[22] Michael T. Heath. Numerical methods for large sparse linear least squares problems. *SIAM Journal on Scientific and Statistical Computing*, 5:497–513, 1984.

[23] Joseph W. H. Liu. The role of elimination trees in sparse factorization. *SIAM Journal on Matrix Analysis and Applications*, 11:134–172, 1990.

[24] L. Lovász and M. D. Plummer. *Matching Theory*. North Holland, 1986.

[25] Kazuo Murota, Masao Iri, and Masataka Nakamura. Combinatorial canonical form of layered mixed matrices and its application to block-triangularization of systems of linear/nonlinear equations. *SIAM Journal on Algebraic and Discrete Methods*, 8:123–149, 1987.

[26] Esmond G. Ng and Barry W. Peyton. A tight and explicit representation of Q in sparse QR factorization. Technical Report ORNL/TM–12059, Oak Ridge National Laboratory, 1992.

[27] S. Parter. The use of linear graphs in Gauss elimination. *SIAM Review*, 3:119–130, 1961.

[28] Alex Pothen. Predicting the structure of sparse orthogonal factors. Manuscript, 1991.

[29] Alex Pothen and Chin-Ju Fan. Computing the block triangular form of a sparse matrix. *ACM Transactions on Mathematical Software*, 16:303–324, 1990.

[30] Donald J. Rose. Triangulated graphs and the elimination process. *Journal of Mathematical Analysis and Applications*, 32:597–609, 1970.

[31] Donald J. Rose and Robert Endre Tarjan. Algorithmic aspects of vertex elimination on directed graphs. *SIAM Journal on Applied Mathematics*, 34:176–197, 1978.

[32] Donald J. Rose, Robert Endre Tarjan, and George S. Lueker. Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing*, 5:266–283, 1976.

[33] H.R. Schwartz. Tridiagonalization of a symmetric band matrix. *Numerische Mathematik*, 12:231–241, 1968.