PREDICTING STRUCTURE IN SPARSE MATRIX COMPUTATIONS

JOHN R. GILBERT*

Abstract. Many sparse matrix algorithms—for example, solving a sparse system of linear equations—begin by predicting the nonzero structure of the output of a matrix computation from the nonzero structure of its input. This paper is a catalog of ways to predict nonzero structure. It contains known results for problems including various matrix factorizations, and new results for problems including some eigenvector computations.

Key words. sparse matrix algorithms, graph theory, matrix factorization, systems of linear equations, eigenvectors

AMS(MOS) subject classifications. 15A18, 15A23, 65F50, 68R10

1. Introduction. A sparse matrix algorithm is an algorithm that performs a matrix computation in such a way as to take advantage of the zero/nonzero structure of the matrices involved. Usually this means not explicitly storing or manipulating some or all of the zero elements; sometimes sparsity can also be exploited to work on different parts of a matrix problem in parallel. Large sparse matrix computations arise in structural design, geodetics, fluid dynamics, heat transport, semiconductor modeling, circuit analysis, molecular dynamics, geophysical reservoir analysis, and many other areas. It is common for problems to be so large that they could not be solved at all without sparse techniques.

Many sparse matrix algorithms [6, 15, 16, 17, 21] have a phase that predicts the nonzero structure of the solution from the nonzero structure of the problem, followed by a phase that does the numerical computation in a static data structure. This saves space, because the space used by the pointers in a dynamic data structure during the first phase can be reused by the numeric values in the second phase. Also, in many applications a sequence of problems with the same nonzero structure must be solved, and the structural phase can be done just once. The structural phase may also be used to schedule the numerical phase efficiently on a parallel machine [20, 29].

Structure prediction can be used to save time as well as space in sparse Gaussian elimination. The asymptotically fastest algorithms to compute the Cholesky factorization of a symmetric positive definite matrix are those of the Yale Sparse Matrix Package [12] and Sparspak [15], which predict the structure of the triangular factor by a version of Theorem 4.3 below. Gilbert and Peierls [24] have used prediction of the structure of the solution of a triangular system of equations, a special case of Theorem 5.1 below, to develop the first algorithm that performs sparse LU factorization with partial pivoting in worst-case time proportional to the number of real arithmetic operations.

Graph theory is a useful language in which to state and prove structure prediction results. One reason for this is that the structural effect of a matrix computation often depends on path structure, which is easier to describe in terms of graphs than in terms of matrices. Seymour Parter [33] was among the first to use graph theory as a tool to

* Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304. This work was done while the author was visiting the University of Iceland and the Institute for Mathematics and Its Applications at the University of Minnesota. Research partially supported by the National Science Foundation under grant DCR-8451385.

investigate sparse matrix computation; Miroslav Fiedler [13] also pioneered many of these ideas.

This paper is a catalog of the effects on nonzero structure of several common matrix computations. It includes arithmetic, linear systems, various factorizations, and some eigenvector problems. Where appropriate, it cites algorithms to compute nonzero structure as well as theorems that describe it. Some of these results are not new. The known results are scattered among papers on various topics in linear algebra and algorithms, published in journals on numerical analysis, theoretical computer science, operations research, and engineering. Here they are presented in a common framework, together with a few new results.

Three or four different graph models are used for structure prediction. Undirected graphs model symmetric matrices; directed graphs model unsymmetric matrices under symmetric permutations; bipartite graphs model arbitrary rectangular matrices; and column intersection graphs can sometimes be used to apply undirected graph results to rectangular matrices. This paper describes results using all the models, but concentrates most heavily on those that use directed graphs to model unsymmetric matrices with nonzero diagonal elements.

Sections 3 through 6 contain the results of the paper: Roughly speaking, the results in Section 3 are immediate; those in Section 4 are known; most of those in Section 5 are consequences of known results; and those in Section 6 are new. This paper is based on an earlier technical report [19].

2. Definitions. We assume that the reader is familiar with such basic graph theoretic terms as directed graph, undirected graph, and path. Harary [26] is a good general reference.

2.1. Directed graphs and matrix structures. Let A be an n by n matrix. The *structure* of A is its directed graph

$$\mathrm{struct}(A) = G(A),$$

whose vertices are the integers $1, \ldots, n$ and whose edges are

$$\{\, (i, j) : i \neq j \text{ and } A_{ij} \neq 0 \}.$$

When no amgibuity can arise, we shall sometimes not distinguish between a matrix, its graph, and the set of edges of its graph.

The graph $G(A)$ does not specify whether the diagonal elements of A are zero or not. In this paper we will use $G(A)$ only for matrices with nonzero diagonal elements; in the context of structure prediction, matrices with zero diagonal elements are more usefully studied by means of their bipartite graphs, as described below.

Applying the same permutation to the columns and rows of A corresponds to renumbering the vertices of the directed graph of A. In other words, if P is a permutation matrix then the directed graph of PAP^T is isomorphic to the directed graph of A. In general, if P and Q are different permutation matrices then the directed graph of PAQ^T is not isomorphic to that of A.

The *structure* of a vector x is

$$\mathrm{struct}(x) = \{\, i : x_i \neq 0 \},$$

which can be interpreted as a set of vertices of the directed graph of any n by n matrix.

We will use the notation $G_1 \subseteq G_2$ to mean that graph G_1 is a subgraph of graph G_2, that is, that both the edge and vertex sets of G_1 are subsets of those of G_2.

2.2. Predicting structure in a computation. To say more precisely what we mean by the structural effect of a computation, we make some remarks based on those of Brayton, Gustavson, and Willoughby [3] and Edenbrandt [10]. Let f be a function from one or more matrices or vectors to a matrix or vector. The structure of A may not determine the structure of $f(A)$; for example, in general the sum of two full vectors is full, but $(1,1)^T + (1,-1)^T$ is not full. We wish to ignore zeros created by coincidence in the numerical values of A, and determine the smallest structure that is "big enough" for the result of f with any input of the given structure. That is, given f and $\text{struct}(A)$, we want to determine

$$\bigcup_B \{\text{struct}(f(B)) : \text{struct}(B) \subseteq \text{struct}(A)\}.$$

Brayton, Gustavson, and Willoughby called an algorithm "s-minimal" if it computes this structure from $\text{struct}(A)$. We sometimes call this a "one-at-a-time" structure prediction, since each position in the predicted structure can be made nonzero but there is no guarantee that all can be made nonzero at the same time.

Most of the functions we consider in this paper have the property that for each input structure S, there is a worst-case value A with $\text{struct}(A) = S$ such that $\text{struct}(B) \subseteq S$ implies $\text{struct}(f(B)) \subseteq \text{struct}(f(A))$. In other words, each input structure corresponds to a unique maximal output structure. We sometimes call such an analysis an "all-at-once" structure prediction.

A function for which there is no "all-at-once" structure prediction is $f(A) = U$, the upper triangular factor of A in Gaussian elimination with partial pivoting. Suppose the structure of A is

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

Depending on the relative magnitudes of the elements in the first column, the structure of $f(A)$ may be

$$\begin{pmatrix} \times & & \\ & \times & \\ & & \times \end{pmatrix}, \quad \begin{pmatrix} \times & \times & \\ & \times & \times \\ & & \times \end{pmatrix}, \quad \begin{pmatrix} \times & \times & \\ & \times & \\ & & \times \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} \times & & \times \\ & \times & \times \\ & & \times \end{pmatrix}.$$

The smallest structure big enough for $f(A)$ is a full upper triangular matrix, even though $f(A)$ cannot be full.

2.3. Graph terminology. Representing a matrix structure as a graph has the advantage that it is easy to describe properties that depend on paths in the graph. Here we define several graph notions that depend on path structure.

The notation $i \overset{A}{\to} j$ means that there is an edge from i to j in $G(A)$, that is, that $A_{ij} \neq 0$. The notation $i \overset{A}{\Longrightarrow} j$ means that there is a directed path from i to j in $G(A)$. Such a path may have length zero; that is, $i \overset{A}{\Longrightarrow} i$ always holds. The matrix A may be omitted if it is clear from context.

Now let $G(A)$ be a directed graph and let x be a subset of the vertices of G. We say x is *closed* (with respect to A) if there is no edge of G from a vertex not in x to a vertex in x; that is, if $x_j \neq 0$ and $A_{ij} \neq 0$ imply $x_i \neq 0$. The *closure* of x (with respect to A) is the smallest closed set containing x,

$$\text{closure}(x) = \bigcap \{\, y : x \subseteq y \text{ and } y \text{ is closed}\},$$

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

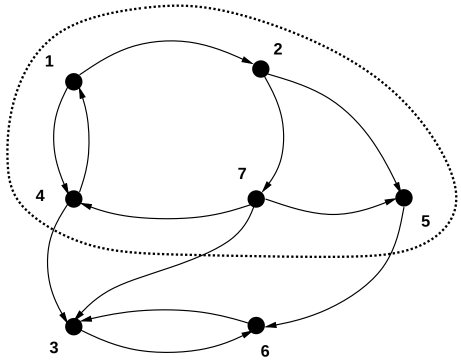

FIG. 1. *A matrix and its directed graph.* closure($\{5\}$) *is outlined.*

which is the set of vertices of A from which there are paths to vertices of x. Figure 1 shows a matrix and its graph, with closure($\{5\}$) outlined.

The *transitive closure* of A is the graph $G^*(A)$ whose edges correspond to paths in A. Thus

$$i \xrightarrow{G^*(A)} j \quad \text{if and only if} \quad i \neq j \text{ and } i \stackrel{A}{\Longrightarrow} j.$$

The structural effect of Gaussian elimination can be described in terms of a subgraph of $G^*(A)$ called the *filled graph* of A, which we write as $G^+(A)$. This graph has edges corresponding to paths whose highest-numbered vertices are their endpoints:

$$i \xrightarrow{G^+(A)} j \quad \text{if and only if} \quad i \neq j \text{ and } i \stackrel{A}{\Longrightarrow} j \text{ through vertices less than } \min(i, j).$$

Figure 2 is an example. Notice that the filled graph depends on the numbering of the vertices of A, whereas the transitive closure and the closure of a vertex are preserved under renumbering (that is, under graph isomorphism).

REMARK 2.1. Rose [35] introduced the notation $G^*(A)$ for the filled graph of A, but that notation is also widely used for transitive closure. Since we want to refer to both transitive closures and filled graphs, we use $G^+(A)$ for the "smaller" of the two.

A graph G is *strongly connected* if there is a path from every vertex to every other vertex, or equivalently if G^* is a complete directed graph. A square matrix A is called *irreducible* if struct(A) is strongly connected. Clearly, for any permutation matrix P, PAP^T is irreducible if and only if A is. The *strongly connected components* (or just *strong components*) of a graph G are its maximal strongly connected subgraphs. Every vertex of a graph is in exactly one strong component, and every edge is in at most one strong component. If a square matrix A is permuted into block triangular form with as many diagonal blocks as possible, the diagonal blocks partition the rows and columns of A into sets corresponding to the strong components of struct(A). Figure 3 shows the strong components of the graph in Figure 1, with the vertices renumbered so that the matrix has a diagonal block for each strong component.

A square matrix A is called *fully indecomposable* if struct(PAQ^T) is strongly connected for all permutation matrices P and Q. Clearly this is equivalent to struct(PA)

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

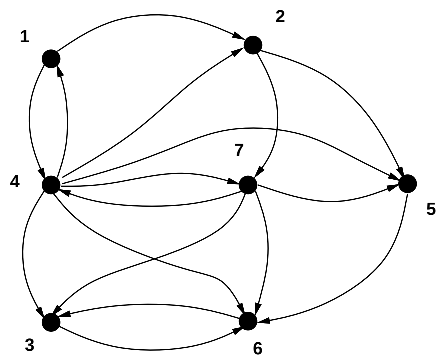

FIG. 2. *Filled graph of example from Figure 1.*

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

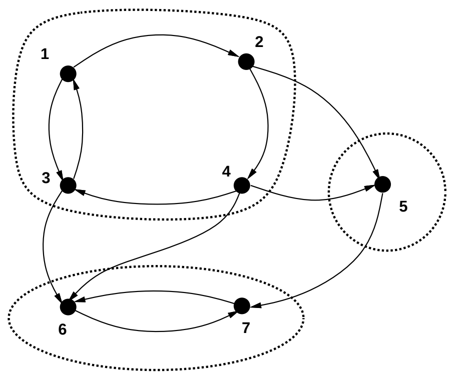

FIG. 3. *Matrix from Figure 1 permuted to block triangular form. Strong components are outlined.*

being strongly connected for all permutations P. As we will remark in the next section, fully indecomposable matrices are the same as square strong Hall matrices.

If matrix A is symmetric, its directed graph contains edge (i, j) if and only if it contains edge (j, i). Informally, we shall not distinguish between this graph and the *undirected graph* of A, which has an undirected edge $\{i, j\}$ if $A_{ij} \neq 0$. A class of undirected graphs that are useful for describing symmetric Gaussian elimination is the chordal graphs: An undirected graph is *chordal* if every cycle of length at least four has a *chord*, that is, if for every cycle $v_1, v_2, \ldots, v_k, v_1$ with $k \geq 4$ there is some edge $\{v_i, v_j\}$ for which $i \neq j \pm 1 \pmod{k}$.

2.4. Non-square matrices. This paper concentrates on structure prediction results that use directed graphs and that (for the most part) apply to square matrices with nonzero diagonals. For completeness, however, we include some results on matrices that need not be square, or that may have zero diagonal elements. Two other graph models are appropriate in that case.

If A has m rows and n columns, then $A^T A$ is an n by n symmetric matrix, whose diagonal is nonzero if A has no zero columns. Its structure is related to the *column intersection graph* of A, which is the undirected graph $G_\cap(A)$ whose vertices are the integers $1, \ldots, n$ (corresponding to the columns of A), and whose edges are

$$\{\, \{i, j\} : i \neq j \text{ and } \exists k \text{ with } A_{ki} \neq 0 \text{ and } A_{kj} \neq 0\}.$$

Thus $G_\cap(A)$ has an edge between any pair of columns that share a nonzero row. This implies that

$$G(A^T A) \subseteq G_\cap(A),$$

with equality unless there is numerical cancellation in $A^T A$. Permuting the rows of A does not change the column intersection graph: if P is a row permutation matrix, then $G_\cap(PA)$ is the same as $G_\cap(A)$.

If A has m rows and n columns, the *bipartite graph* of A is the undirected graph $H(A)$ whose vertices are $1', 2', \ldots, m'$ and $1, 2, \ldots, n$, and whose edges are $\{\, \{i', j\} : A_{ij} \neq 0\}$. The superscript prime notation is intended to indicate that the row and column vertices of $H(A)$ are chosen from two different copies of the positive integers. Permuting the rows and columns of A only relabels the vertices of the bipartite graph: if P and Q are row and column permutation matrices, then $H(PAQ^T)$ is isomorphic to $H(A)$.

Several structure prediction problems use matchings and alternating paths in the bipartite graph of a matrix [4, 6, 7, 20, 23, 21, 32]. This paper does not consider such problems in detail, but we include enough definitions here to state some of these results in later sections.

Let A be an m by n matrix with $m \geq n$. We say that A has the *Hall property* if, for every k with $0 \leq k \leq n$, every set of k columns of A contains nonzeros in at least k rows. (That is, every set of k column vertices of $H(A)$ is adjacent to at least k row vertices.) We say that A has the *strong Hall property* if, for every k with $0 < k < m$, every set of k columns of A contains nonzeros in at least $k+1$ rows. The graph $H(A)$ has a matching that covers all its columns if and only if A has the Hall property. A square matrix A with nonzero diagonal is irreducible if and only if $H(A)$ has the strong Hall property; an arbitrary square matrix A is fully indecomposable if and only if $H(A)$ has the strong Hall property. See Lovasz and Plummer [30] for background on bipartite matching. Our terminology is from Coleman, Edenbrandt, and Gilbert [4].

Incidentally, although permuting the rows and the columns of A independently can change the directed graph of A, it does not change the row partition and the column partition induced by the strong components of the graph of A. Another way to say this is that, given a matrix A, if we first permute rows and columns (asymmetrically) to make the diagonal elements nonzero, and then permute rows and columns symmetrically to block triangular form with as many diagonal blocks as possible, then regardless of the initial choice of nonzero diagonal we always get the same block triangular form, up to possible permutation of the diagonal blocks and reordering of the rows and columns within each block.

2.5. Other definitions. We will call a finite set $\{x_1, \ldots, x_n\}$ of complex numbers *algebraically independent* if the point $(x_1, \ldots, x_n)$ is not a zero of any nonzero n-variable polynomial with integer coefficients. Then x_i is transcendental over the field $\mathbf{Q}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ of the rationals extended by all the x's except x_i. There exist arbitrarily large algebraically independent sets, even of real numbers, by a simple countability argument.

3. Products. The following simple result is used in the proof of Theorem 6.1. The structure is the bipartite graph of the matrix in question.

THEOREM 3.1. *Let the structures of an m by n matrix A and an n-vector x be given.*

- *(i) Whatever values A and x have,* struct(Ax) *is a subset of the row vertices of $H(A)$ adjacent to column vertices whose indices are in* struct(x).
- *(ii) There exist values for the nonzeros of A and x such that* struct(Ax) *is equal to the set of row vertices described above.* □

The generalization to products of matrices is immediate, since each column of AB is A times a column of B. Theorem 3.1 also implies that $G(A^T A) \subseteq G_\cap(A)$.

4. Factorizations. In this section we describe the structural effect of several matrix factorizations. The necessary definitions are in Section 2.

4.1. ***LU*** **factorization.** For the factorization $A = LU$ where L is lower triangular with unit diagonal and U is upper triangular, we consider square matrices A with nonzero diagonal, and the graph in question is the directed graph of A. The square matrix $L + U - I$ represents the entire factorization. (Not all nonsingular matrices have LU factorizations without pivoting. In a later subsection we consider factorization with partial pivoting.)

THEOREM 4.1 ([36]). *Let a structure $G(A)$ be given, with nonzero diagonal elements.*

- *(i) If values are chosen for which A has an LU factorization as above, then $G(L + U - I) \subseteq G^+(A)$.*
- *(ii) There exist values for the nonzeros of A such that $G(L+U-I) = G^+(A)$.* □

Rose and Tarjan [36] gave an algorithm for computing $G^+(A)$ from A in $O(nm)$ time, where A is n by n with m nonzeros. They also showed that $G^*(A)$ can be computed in time asymptotically the same as that to compute $G^+(A)$, so a faster algorithm to compute $G^+(A)$ would give a faster algorithm to compute transitive closures than the best currently known. Eisenstat, Gilbert, and Liu [11, 22] give algorithms to compute $G^+(A)$ that are more efficient in practice than transitive closure, by using various transitively reduced graphs.

REMARK 4.2. A nonsingular square matrix may have an LU factorization even though it has zeros on the diagonal. In this case, Theorem 4.1(i) still holds; but the converse, part (ii), is false. Brayton, Gustavson, and Willoughby [3] gave a counterexample. Let

$$\text{struct}(A) = \begin{pmatrix} \times & \times & \times & \\ \times & & & \\ \times & \times & \times & \\ \times & & & \times \end{pmatrix}.$$

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

4.2. Cholesky factorization. Here we consider the factorization $A = LL^T$, where A is a symmetric, positive definite matrix, and L is lower triangular with positive diagonal. Then A has a nonzero diagonal because it is positive definite, and the directed graph of A corresponds to an undirected graph because A is symmetric.

THEOREM 4.3 ([37]). *Let a symmetric structure $G(A)$ be given, with nonzero diagonal elements.*

(i) No matter what values A has, if A has a Cholesky factorization $A = LL^T$ then $G(L) \subseteq G^+(A)$.
(ii) There exist symmetric values for the nonzeros of A such that $G(L + L^T) = G^+(A)$. $\square$

Rose, Tarjan, and Lueker [37] gave an $O(n+f)$ algorithm to compute $G^+(A)$ for a symmetric n by n matrix A with fill of size f. Another such algorithm is implemented in several standard sparse matrix computation packages [12, 15].

Rose showed that the graphs of Cholesky factors of symmetric matrices are exactly the chordal graphs; or, equivalently, that a structure can be reordered to have no fill if and only if it is chordal.

THEOREM 4.4 ([37]). *Let a symmetric structure $G(A)$ be given, with nonzero diagonal elements.*

(i) $G^+(A)$ is a chordal graph.
(ii) Conversely, if $G(A)$ is a chordal graph, then there is a permutation matrix P such that $G^+(PAP^T) = G(PAP^T)$. $\square$

Rose, Tarjan, and Lueker [37] and Tarjan and Yannakakis [39] gave linear-time algorithms to determine whether a $G(A)$ is chordal and, if so, to reorder its vertices so that $G^+(PAP^T) = G(PAP^T)$. Such a reordering is called a "perfect elimination order."

4.3. Partial pivoting. The example in Section 2 showed that a result of the form of Theorem 4.3 is not possible for LU factorization with partial pivoting, because no single choice of nonzero values for A is guaranteed to produce nonzeros in all possible positions of U. George and Ng [16] gave an upper bound. A few remarks are necessary to understand the bound.

There are two ways to write the LU factorization, with partial pivoting, of a square matrix A. One is as $A = PLU$, where L is unit lower triangular, U is upper triangular, and P is a permutation matrix. The other is as $A = P_1L_1P_2L_2 \cdots P_{n-1}L_{n-1}U$, where P_i is a permutation that just transposes row i and a higher-numbered row, and L_i is a Gauss transform (a unit lower triangular matrix with nonzeros only in column i). To get the first factorization, use the standard outer product form of Gaussian elimination to replace A by its triangular factors, pivoting by interchanging two rows of the matrix at the beginning of each major step; at major step k, each row thus interchanged contains entries of L in the first $k-1$ positions and entries of the partially factored A in the remaining positions. To get the second factorization, pivot by interchanging, at the beginning of major step k, only columns k through n of the two rows in question. In this case an entry of the lower triangle is never moved once it is computed, and only the rows of the partially factored matrix are interchanged.

The factorizations are equivalent in the sense that the same arithmetic is performed in each case, the two U's are the same, and the values of the nonzeros in $L-I$ and $\hat{L} = \sum_{1 \leq i < n}(L_i - I)$ are the same; only the positions of the nonzeros in the lower triangular factors are different. The George-Ng theorem describes the structures of $\hat{L}$ and U, saying that both of them are subsets of the structure of the symbolic Cholesky factor of $A^T A$, that is of the filled graph of the column intersection graph of A.

Theorem 4.5 ([16]). *Let a structure $G(A)$ be given. Whatever values A has, if Gaussian elimination with partial pivoting gives the factors $\hat{L}$ and U as above, then $G(\hat{L}+U) \subseteq G_{\cap}^{+}(A)$.* $\square$

The example in Section 2 showed that the converse of this theorem is not true. Various partial converses hold, however. If we restrict ourselves to structures that are strong Hall (fully indecomposable), which includes irreducible structures with nonzero diagonals, then there is a "one-at-a-time" converse to Theorem 4.5 for the upper triangular factor U:

Theorem 4.6 ([20]). *Let a square strong Hall structure $H(A)$ be given. For any choice of $i < j$ with $i \xrightarrow{G_{\cap}^{+}(A)} j$, there is a choice of values for A for which the factorization with partial pivoting $PA = LU$ makes $U_{ij} \neq 0$.* $\square$

The "all-at-once" version of this statement is not true even for irreducible matrices. For example, take $G(A)$ to be a tridiagonal matrix plus a full first column. Then $G(A)$ is irreducible (and strong Hall), and it is easy to see that $G_{\cap}^{+}(A)$ is full. As Theorem 4.6 says, it is possible to choose values for A to make any single position in U nonzero; however, the first row of U will always be some row of A, so the entire first row of U cannot be nonzero at the same time.

For the lower triangular factor L, Theorem 4.5 is not as tight as possible even in the "one-at-a-time" sense for strong Hall structures. For example let $G(A)$ be tridiagonal (and hence strong Hall). Then $G_{\cap}^{+}(A)$ is five-diagonal, predicting that $\hat{L}$ could be lower tridiagonal; but, in fact, $\hat{L}$ must be lower bidiagonal. George and Ng [17] suggest a way of predicting the structures of $\hat{L}$ and U by efficiently simulating all possible pivoting steps. Gilbert and Ng [20] have recently shown that this method (which we do not describe here in detail) gives a tight "one-at-a-time" prediction of the structures of both L and U in the strong Hall case.

4.4. QR factorization. Suppose A is an m by n matrix with $m \leq n$. Here we consider the factorization $A = QR$, where Q is an orthogonal matrix and R is upper triangular with nonnegative diagonal. George and Heath observed that, since this R is the same as the Cholesky factor of $A^T A$, the structure of R can be predicted by forming $G_{\cap}(A)$ and doing structural Cholesky factorization.

Theorem 4.7 ([14]). *Let the structure $H(A)$ be given for a rectangular matrix A with at least as many rows as columns. Whatever values A has, if A has full column rank then its orthogonal factorization $A = QR$ satisfies $G(R) \subseteq G_{\cap}^{+}(A)$.* $\square$

The converse of this theorem is false; for example, if A is upper triangular with a nonzero diagonal and a full first row, then $G_\cap(A) = G_\cap^+(A)$ is full, but the orthogonal factor R is equal to A. Coleman, Edenbrandt, and Gilbert supplied a converse in the strong Hall case.

THEOREM 4.8 ([4]). *Let a structure $H(A)$ be given with at least as many rows as columns. If $H(A)$ has the strong Hall property, then there exist values for the nonzeros of A such that $G(R) = G_\cap^+(A)$, where R is the orthogonal factor of A as above.* □

Johnson, Olesky, and van den Driessche [27] gave a more complicated prediction of the structure of both R and the orthogonal factor Q. They showed that their prediction was tight in the "one-at-a-time" sense for all Hall structures, that is, for all structures with full symbolic column rank. Pothen [34] then proved that Johnson et al.'s prediction was in fact tight in the "all-at-once" sense, thus finishing off the problem for both Q and R in the Hall case.

George and Ng [18] studied another representation of the structure of the orthogonal factor Q in the case that A is square and has nonzero diagonal elements. They showed that a suitable representation of Q has a structure that also satisfies Theorem 4.7. Suppose A is reduced to upper triangular form by a sequence of Householder transformations that zero the subdiagonal elements of the first column, then the second column, and so on. The Householder transformation used to zero column j (for $1 \leq j < n$) has the form $Q_j = I - w_j w_j^T$, where w_j is a column vector whose first $j-1$ entries are zero. The orthogonal factor $Q = Q_1^T Q_2^T \cdots Q_{n-1}^T$ is conveniently represented by the lower triangular matrix W whose columns are the w_j. The George-Ng result describes the structure of this triangular matrix.

THEOREM 4.9 ([18]). *Let a structure $G(A)$ be given for a square matrix A with nonzero diagonal elements. Whatever values A has, if its orthogonal factor is represented by the matrix W described above, then $G(W) \subseteq G_\cap^+(A)$.* □

This result is useful because in practice it often suffices to represent the orthogonal factor by the sequence of Householder vectors, and that representation is often sparser than an explicit representation of Q. For example, consider a symmetric structure whose graph $G(A)$ is a square grid with n vertices, corresponding to the standard five-point finite difference stencil on a k by k mesh with $n = k^2$. The number of nonzeros in A is $\Theta(n)$. Take the numeric values of A to be algebraically independent. It is straightforward to use the results cited above [18, 27] to show that a nested dissection ordering of $G_\cap(A)$ asymptotically minimizes the number of nonzeros in all of Q, R, and W, and that for such an ordering the number of nonzeros in both R and W is $\Theta(n \log n)$ while the number of nonzeros in Q is $\Theta(n^{3/2})$. Thus in this model problem W is a much more efficient way to store the orthogonal factor than Q.

5. Solutions of linear systems. In this section we determine the structure of the solution x to the square system of linear equations $Ax = b$. We solve the related problem of determining the structure of A^{-1}. These results have not appeared before in this form, but the upper bounds in Theorem 5.1 and Corollary 5.4 are straightforward consequences of Tarjan's work on elimination methods for solving path problems in graphs [40], and are closely related to work of Fiedler [13]. The proofs here are somewhat different.

The extremes of path structure in directed graphs are a strongly connected graph (which corresponds to an irreducible matrix) and an acyclic graph (which corresponds to a permutation of a triangular matrix). Some of these results become almost trivial at the strongly connected extreme—for example, the inverse of an irreducible matrix is full in the absence of coincidental cancellation. In solving general nonsingular linear systems it is often advantageous to begin by partitioning the matrix into strong components, and then to factor only the irreducible blocks of the partition. This approach is taken, for example, in Duff and Reid's MA28 code [9].

Curiously enough, the most important applications of the results in this section are at the opposite extreme, for triangular systems. Structure prediction for sparse triangular systems is used in efficient algorithms for LU factorization with partial pivoting [24] and in parallel triangular solution [1].

Throughout this section A is an n by n matrix with nonzero diagonal, and the graph in question is the directed graph $G(A)$.

THEOREM 5.1. *Let the structures of A and b be given.*

(i) Whatever the values of the nonzeros in A and b, if A is nonsingular then

$$\mathrm{struct}(A^{-1}b) \subseteq \mathrm{closure}(b).$$

(ii) There exist nonzero values for which $\mathrm{struct}(A^{-1}b) = \mathrm{closure}(b)$. *(In fact, all the nonzeros in b can have the value 1.)*

Proof. Part (i). Let values be given for which A is nonsingular. Renumber the vertices of A so that $\mathrm{closure}(b) = \{1, 2, \ldots, k\}$ for some $k \leq n$. Then $Ax = b$ can be partitioned as

$$\begin{pmatrix} B & D \\ C & E \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix},$$

where B is k by k. By the definition of $\mathrm{closure}(b)$, there is no edge (i, j) with $i \notin \mathrm{closure}(b)$ and $j \in \mathrm{closure}(b)$. Therefore $C = 0$. Then $Ez = 0$. Since A is nonsingular and $C = 0$, matrix E is nonsingular. Therefore $z = 0$. Thus $\mathrm{struct}(x) \subseteq \{1, \ldots, k\} = \mathrm{closure}(b)$.

Part (ii). Choose algebraically independent values for the nonzeros of A, and let $b_i = 1$ if $i \in \mathrm{struct}(b)$. Then A is nonsingular because $\det A$ is a nonzero polynomial in the nonzeros of A. Let $x = A^{-1}b$. Renumber the vertices of A so that $\mathrm{struct}(x) = \{1, 2, \ldots, k\}$ for some $k \leq n$. Then $Ax = b$ can be partitioned as

$$\begin{pmatrix} B & D \\ C & E \end{pmatrix} \begin{pmatrix} y \\ 0 \end{pmatrix} = \begin{pmatrix} d \\ e \end{pmatrix},$$

where B is k by k and all entries of y are nonzero. Consider row i of C. We have

$$\sum_{1 \leq j \leq k} c_{ij} y_j = e_i. \tag{1}$$

Now B is nonsingular, since $\det B$ is a nonzero polynomial. By Cramer's rule, $By = d$ implies $y_j = \det(B|_j^d)/\det B$, where $B|_j^d$ is B with column j replaced by d. Then Equation 1 implies

$$\sum_{i \leq j \leq k} c_{ij} \det(B|_j^d) - e_i \det B = 0. \tag{2}$$

This is a polynomial with rational coefficients in the entries of A, so it is the zero polynomial. Now $y_j \neq 0$ implies that $\det(B|_j^d)$ is not the zero polynomial, so c_{ij} must be zero. Thus $C = 0$. This implies that $x = \binom{y}{0}$ is closed. Furthermore, $\det B \neq 0$, so Equation 2 implies $e_i = 0$. Thus $e = 0$, so $b = \binom{d}{0}$ and $\mathrm{struct}(b) \subseteq$

struct(x) = closure(x). Therefore closure$(b) \subseteq$ struct(x). With part (i), this gives closure(b) = struct(x). $\square$

REMARK 5.2. The proof of part (i) never assumes that the "nonzero values" of A are in fact different from 0. Thus we have the slightly stronger result that if $G(\hat{A}) \subseteq G(A)$ and struct$(\hat{b}) \subseteq$ struct(b) and $\hat{A}$ is nonsingular, then struct$(\hat{A}^{-1}\hat{b}) \subseteq$ closure(b).

REMARK 5.3. It seems natural to conjecture in part (i) that if A is singular and $Ax = b$ has a solution, then it has some solution with struct$(x) \subseteq$ closure(b). Oddly enough, this is false. Consider

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 2 & 2 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

All solutions to $Ax = b$ are of the form $(\alpha, -\alpha, 1, 1)^T$, none of which is a subset of closure$(b) = (\times, \times, 0, 0)^T$.

COROLLARY 5.4. *Let the structure $G(A)$ be given.*

(i) Whatever values A has, if A is nonsingular then $G(A^{-1}) \subseteq G^*(A)$.
(ii) Values exist for the nonzeros of A such that $G(A^{-1}) = G^*(A)$.

Proof. Note that column j of $G^*(A)$ is closure$(e^{(j)})$, where $e^{(j)}$ is the j^{th} unit vector. The corollary is immediate from Theorem 5.1, noting that part (ii) of the theorem holds even if the right-hand side entries are all zeros and ones. $\square$

Corollary 5.4 implies that if A is irreducible with nonzero diagonal, then A^{-1} is full unless numerical coincidence occurs. Duff et al. [8] gave another proof of this special case.

The case where A is allowed to have zeros on the diagonal is a straightforward extension. First, for A^{-1} to exist, $H(A)$ must be Hall. That implies that a permutation P exists for which PA has nonzero diagonal. Then the structure of the solution to $Ax = b$ is the structure of $A^{-1}b = (PA)^{-1}(Pb)$, which is the closure of Pb with respect to the graph $G(PA)$ by Theorem 5.1. The structure of A^{-1} can be predicted similarly by permuting to nonzero diagonal, forming the transitive closure, and permuting back. This also implies a slightly stronger version of Duff et al.'s result: if nonsingular A is fully indecomposable, or equivalently strong Hall, (with no restriction on the diagonal) then A^{-1} is full unless numerical coincidence occurs.

The case where A is symmetric is simpler and less interesting, but the puzzling examples like that in Remark 5.3 do not arise. If A is symmetric and its graph is not connected, then A is block diagonal, and a linear system splits into a separate problem for each block. If A is connected, then it is strongly connected and the closure of every nonempty set is the whole graph. Then the upper bound in part (i) of Theorem 5.1 is trivial, and values exist to achieve it.

THEOREM 5.5. *Let a symmetric structure for A be given along with a nonzero structure for b. If the structure for A is connected (i.e. irreducible, or not block diagonal) then there exist symmetric values for A such that* struct$(A^{-1}b) = \{1, 2, \ldots, n\}$*; that is, x is full. Also, in this case, A^{-1} is full.*

Proof. The proof is almost identical to that of Theorem 5.1 (ii), so this is just a sketch: Choose algebraically independent values for the lower triangle of A and make the upper triangle symmetric. Then A is nonsingular. The polynomial in Equation 2 does not contain c_{ji}, so we can still conclude $c_{ij} = 0$ from the fact that it occurs multiplied by a nonzero polynomial. Therefore $A^{-1}b$ is closed. But if symmetric

A is connected then it is strongly connected, so the only nonempty closed set is $\{1, 2, \ldots, n\}$. $\square$

6. Eigenvectors. In this section we determine the structure of the eigenvectors of a square matrix A. The results in this section are new. We deal only with the case of distinct eigenvalues. As described at the end of the section, the reason we cannot handle multiple eigenvalues is related to Remark 5.3 above.

Throughout this section A is an n by n matrix with nonzero diagonal, and the graph in question is the directed graph of A. Recall that $e^{(i)}$ is the i^{th} unit vector and closure$(e^{(i)})$ is the structure of column i of the transitive closure of A.

THEOREM 6.1. *Let the structure $G(A)$ be given.*

(i) Whatever the values of the nonzeros in A, if A has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ then the eigenvectors of A can be numbered $u^{(1)}, \ldots, u^{(n)}$ such that $\mathrm{struct}(u^{(i)}) \subseteq \mathrm{closure}(e^{(i)})$.

(ii) There exist nonzero values for which A has n distinct eigenvalues, and the eigenvectors satisfy $\mathrm{struct}(u^{(i)}) = \mathrm{closure}(e^{(i)})$.

Proof. Part (i). Let values be given for A. Renumber the vertices of A to put A in block upper triangular form—that is, to put the strongly connected components of A in topological order. Then A is partitioned as

$$A = \begin{pmatrix} B_1 & C_{1,2} & \ldots & C_{1,s} \\ & B_2 & \ldots & C_{2,s} \\ & & \ddots & \vdots \\ 0 & & & B_s \end{pmatrix},$$

where each B_j is square and strongly connected. Renumber the eigenvalues and eigenvectors in nondecreasing order of the highest-numbered nonzero in the eigenvector. That is, if $u_k^{(i)} = u_{k+1}^{(i)} = \cdots = u_n^{(i)} = 0$, then $u_k^{(i-1)} = u_{k+1}^{(i-1)} = \cdots = u_n^{(i-1)} = 0$, for $1 < i \leq n$.

Consider some eigenvector $u^{(i)}$. Suppose its highest-numbered nonzero is in a row that runs through block B_j. Then $Au^{(i)} = \lambda_i u^{(i)}$ is partitioned as

$$\begin{pmatrix} D & E & F \\ 0 & B_j & G \\ 0 & 0 & H \end{pmatrix} \begin{pmatrix} v \\ w \\ 0 \end{pmatrix} = \lambda_i \begin{pmatrix} v \\ w \\ 0 \end{pmatrix}, \qquad \text{where } D = \begin{pmatrix} B_1 & \ldots & C_{1,j-1} \\ & \ddots & \vdots \\ 0 & & B_{j-1} \end{pmatrix}, \quad \text{etc.}$$

Then $B_j w = \lambda_i w$ with $w \neq 0$, so λ_i is an eigenvalue of B_j. In fact, each λ_i is an eigenvalue of one B_j, with j increasing as i increases. Since no B_j has more eigenvalues than its dimension, we conclude by counting rows that row i and column i of A run through B_j. Now B_j is strongly connected, so $\mathrm{closure}(e^{(i)}) = \mathrm{closure}(B_j)$ (where $\mathrm{closure}(B_j)$ denotes the closure of the set of vertices of B_j with respect to A).

We have $Dv + Ew = \lambda_i v$, so

$$(D - \lambda_i I)v = Ew.$$

Since the eigenvalues of A are simple, λ_i is not an eigenvalue of D and $D - \lambda_i I$ is nonsingular. Thus, by Theorem 5.1,

$$\mathrm{struct}(v) \subseteq \mathrm{closure}(Ew).$$

Now if D is m by m and B_j is t by t, $\mathrm{struct}(w) \subseteq \{m+1, m+2, \ldots, m+t\}$ and $\mathrm{struct}(Ew) \subseteq \{1 \le k \le m : a_{kl} \ne 0 \text{ for some } l \in \mathrm{struct}(w)\}$ by Theorem 3.1, so $\mathrm{struct}(Ew) \subseteq \mathrm{closure}(w)$ (closure still with respect to A) and $\mathrm{struct}(v) \subseteq \mathrm{closure}(w)$. Therefore, $\mathrm{struct}(u^{(i)}) = \mathrm{struct}(v) \cup \mathrm{struct}(w) \subseteq \mathrm{closure}(w)$. Since $w \subseteq B_j$, this implies that $\mathrm{struct}(u^{(i)}) \subseteq \mathrm{closure}(B_j) = \mathrm{closure}(e^{(i)})$.

Part (ii). Choose algebraically independent values for A, choosing the diagonal elements so far apart that no two are closer than $2\max_j \sum_{i \ne j} |a_{ij}|$. By Gerschgorin's theorem [25], this guarantees that there are n distinct, simple eigenvalues. (It would be more elegant to conclude that the eigenvalues are simple from the algebraic independence of the elements, but I do not know how to prove it.)

First we will show that each eigenvector is closed. Let u be an eigenvector with $Au = \lambda u$. Renumber the vertices of A so that $\mathrm{struct}(u) = \{1, 2, \ldots, t\}$ for some $t \le n$. Then $Au = \lambda u$ can be partitioned as

$$\begin{pmatrix} B & D \\ C & E \end{pmatrix} \begin{pmatrix} v \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} v \\ 0 \end{pmatrix}, \tag{3}$$

where B is t by t and $v_k \ne 0$ for $1 \le k \le t$. We will show $C = 0$. Intuitively, it seems clear that if $C \ne 0$ then v cannot be both an eigenvector of B and a null vector of C. A field-theoretic argument makes this intuition precise.

Since $Bv = \lambda v$ and the diagonal elements of B are far enough apart that their Gerschgorin discs do not overlap, λ is in the Gerschgorin disc of exactly one b_{kk}. Renumber vertices 1 through t so that b_{kk} is b_{11}. Choose v such that $v_1 = 1$. Then $Bv = \lambda v$ partitions into

$$\begin{pmatrix} b_{11} & f^T \\ g & B' \end{pmatrix} \begin{pmatrix} 1 \\ v_2 \\ \vdots \\ v_t \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ v_2 \\ \vdots \\ v_t \end{pmatrix},$$

where f and g are $t-1$-vectors. Now we have

$$(B' - \lambda I) \begin{pmatrix} v_2 \\ \vdots \\ v_t \end{pmatrix} = -g.$$

By Gerschgorin's theorem, λ is not an eigenvalue of B', so $B' - \lambda I$ is nonsingular and

$$v_k = \frac{\det(B' - \lambda I)|_k^{-g}}{\det(B' - \lambda I)} \qquad \text{for } 2 \le k \le t. \tag{4}$$

Now we fix i and j and show that $c_{ij} = 0$ (for $1 \le i \le n-t$ and $1 \le j \le t$).

Let F be the field obtained by adjoining to $\mathbf{Q}$ (the rationals) all the nonzeros of B and also all the nonzeros of row i of C except c_{ij}. Now $F[x]$ is the ring of one-variable polynomials with coefficients in F, and $F(\lambda)$ is the field obtained by adjoining λ to F. We know λ is a zero of a nonzero polynomial in $F[x]$, namely $\det(B - xI)$. Therefore λ is algebraic over F, so every element of $F(\lambda)$ is a zero of some nonzero polynomial in $F[x]$.

Since $Cv = 0$, we have

$$\sum_{1 \le k \le t} c_{ik} v_k = 0.$$

All the v_k are nonzero, so

$$c_{ij} = -\frac{1}{v_j}\sum_{k\neq j} c_{ik} v_k. \tag{5}$$

By Equation 4, each v_k is a rational function of λ and elements of F, so $v_k \in F(\lambda)$. Each c_{ik} with $k \neq j$ is in F. Therefore the whole right hand side of Equation 5 is in $F(\lambda)$, so $c_{ij} \in F(\lambda)$. This means that c_{ij} is a zero of a nonzero polynomial in $F[x]$. But if c_{ij} is nonzero, then c_{ij} was chosen to be transcendental over F. Thus $c_{ij} = 0$; and, since i and j were arbitrary, $C = 0$.

Recalling the partition of A in Equation 3, $C = 0$ implies that the eigenvector $u = \binom{v}{0}$ is closed.

Now all the eigenvectors of A are closed. Renumber the eigenvectors so that λ_i is in the Gerschgorin disc of a_{ii}. The argument following Equation 3 shows that λ_i is in a Gerschgorin disc whose index j corresponds to a nonzero $u_j^{(i)}$ of $u^{(i)}$; since λ_i is in only one disc, this means $u_i^{(i)} \neq 0$. Therefore, $\mathrm{struct}(e^{(i)}) \subseteq \mathrm{struct}(u^{(i)})$. Since $u^{(i)}$ is closed, $\mathrm{closure}(e^{(i)}) \subseteq \mathrm{struct}(u^{(i)})$. Part (i) gives the opposite containment, so $\mathrm{struct}(u^{(i)}) = \mathrm{closure}(e^{(i)})$. □

Corollary 6.2. *Let the structure of A be given.*

- (i) *No matter what nonzero values A has, if A has only simple eigenvalues then its eigenvectors can be ordered so that the matrix U whose columns are the eigenvectors has $G(U) \subseteq G^*(A)$.*
- (ii) *There exist values for the nonzeros of A such that the eigenvectors can be ordered so that $G(U) = G^*(A)$.*

Proof. Similar to Corollary 5.4. □

Remark 6.3. It is natural to conjecture that if A has multiple eigenvalues, then there is some choice of a maximal set of eigenvectors whose structure is a subset of the transitive closure of A. Again, oddly, this is false. From the example in Remark 5.3 we can construct

$$A = \begin{pmatrix} 3 & 1 & 0 & 1 & -1 \\ 2 & 4 & 0 & 1 & -1 \\ 0 & 0 & 3 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}, \quad \text{so } G^*(A) = \begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & & \times & \times & \\ & & \times & \times & \\ & & & & \times \end{pmatrix}.$$

The graph of A is shown in Figure 4. The characteristic equation of A is $\det(xI - A) = (x-2)^4(x-5)$, so the eigenvalues are 2 and 5. The eigenspace of 5 is one-dimensional and consists of multiples of $(1, 2, 0, 0, 0)^T$, which is a column of the transitive closure. However, the eigenspace of 2 is also one-dimensional and consists of multiples of $(0, 0, 1, 1, 1)^T$, which is not a subset of any column of the transitive closure.

Mascarenhas [31] has recently extended Theorem 6.1 to the case where A has multiple eigenvalues, provided that no two diagonal blocks of the block upper triangular form of A share an eigenvalue. He has also proved a similar result for the more general case where there are n linearly independent eigenvectors, that is, where each eigenvalue has equal geometric and algebraic multiplicity.

For symmetric A, the situation for eigenvectors is the same as for symmetric linear systems: If A is block diagonal then each block is a separate problem; if A is not block diagonal (i.e., A is irreducible or connected) then the upper bounds are both trivial and tight.

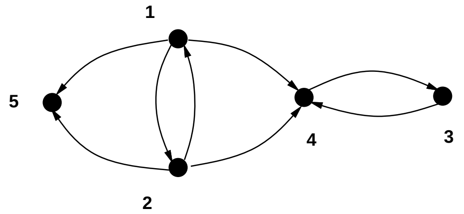

FIG. 4. *Graph of counterexample in Remark 6.3.*

THEOREM 6.4. *Let a symmetric structure for A be given. If the structure is connected, then there exist symmetric values for A such that A has n distinct eigenvalues, and all its eigenvectors are full.*

Proof. Just as in Theorem 5.5, the proof of Theorem 6.1 part (ii) goes through even if A is required to be symmetric. □

7. Remarks, applications, and open problems. We have described several matrix computations in which the nonzero structure of the result of the computation can be inferred, partly or completely, from the nonzero structure of the input to the computation. The language of graph theory seems most appropriate to state these results, primarily because path structure is most easily described in graph-theoretic terms.

Matchings in bipartite graphs are important in several of the results of Section 4. Bipartite matching theory plays a central role in two other structural problems that we have not described here: finding the sparsest basis for the range space (McCormick [32]) and for the null space (Coleman and Pothen [5, 6], Gilbert and Heath [21]) of a rectangular matrix. It turns out that the structural range space problem can be solved in polynomial time, but the null space problem is NP-complete.

Several applications of structure prediction to solving systems of linear equations were cited in Section 1. Some of the present work was motivated by Gilbert and Peierls's use of structure prediction for triangular linear systems as part of an efficient algorithm for sparse LU factorization with partial pivoting [24]. Another application of structure prediction for triangular systems is in a practical problem in reservoir analysis. Here a finite-element model of an underground reservoir of hot water (to be tapped for power and heating for the city of Reykjavík) requires the solution of hundreds of positive definite linear systems with the same coefficient matrix. All the systems have very sparse right-hand sides, and in addition only a few of the unknown values are required for each system. Sigur∂sson [38] has used structure prediction with a simpler version of Theorem 5.1 to speed up the Sparspak triangular solver for this problem.

We have been concerned exclusively with predicting nonzero structure in this paper. A related question is, given a matrix and a matrix function, which entries of the matrix are unchanged in value by application of the function. Barrett, Johnson, Olesky, and van den Driessche [2, 28] have given such characterizations for functions including LU factorization and Schur complement.

A few open problems in structure prediction, some of which have already been mentioned, are as follows. Is it possible to give a tight bound on the nonzero structures of the factors in Gaussian elimination with partial pivoting (Section 4)? What can be said about solutions to singular linear systems in light of the counterexample in

Section 5? What can be said about eigenvector structures for matrices with multiple eigenvalues (Section 6, [31])? What can be said about the structure of the singular value decomposition of a rectangular matrix [25]? The relationship between the singular values of A and the eigenvalues of $A^T A$, together with Theorem 6.4 on eigenvectors of symmetric matrices, suggests that the SVD of a connected matrix is always full (ignoring numerical cancellation). This would certainly confirm the conventional wisdom that there is no such thing as an SVD with sparse singular vectors.

Acknowledgements. My thanks to Tom Coleman, Anders Edenbrandt, Mike Heath, Joseph Liu, Esmond Ng, Ragnar Sigur∂sson, and Sven Sigur∂sson for interesting and useful discussions of these problems. Earl Zmijewski gave this paper a careful and helpful critical reading. Gene Golub nudged me to finish revising it for publication.

REFERENCES

[1] Fernando Alvarado and Robert Schreiber. Optimal parallel solution of triangular sparse systems. To appear in the *SIAM Journal on Scientific and Statistical Computing*, 1992.

[2] Wayne W. Barrett, Charles R. Johnson, D. D. Olesky, and P. van den Driessche. Inherited matrix entries: Principal submatrices of the inverse. *SIAM Journal on Algebraic and Discrete Methods*, 8:313–322, 1987.

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

[4] 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.

[5] Thomas F. Coleman and Alex Pothen. The null space problem I: complexity. *SIAM Journal on Algebraic and Discrete Methods*, 7:527–537, 1986.

[6] Thomas F. Coleman and Alex Pothen. The null space problem II: algorithms. *SIAM Journal on Algebraic and Discrete Methods*, 8:544–563, 1987.

[7] I. S. Duff. On algorithms for obtaining a maximum transversal. *ACM Transactions on Mathematical Software*, 7:315–330, 1981.

[8] I. S. Duff, A. M. Erisman, C. W. Gear, and J. K. Reid. Some remarks on the inverses of sparse matrices. Technical Report Mathematics and Computer Science Division Report 51, Argonne National Laboratory, 1985.

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

[10] Anders Gunnar Edenbrandt. *Combinatorial Problems in Matrix Computation*. PhD thesis, Cornell University, 1985.

[11] Stanley C. Eisenstat and Joseph W. H. Liu. Exploiting structural symmetry in unsymmetric sparse symbolic factorization. Technical Report CS–90–12, York University, 1990. To appear in SIMAX.

[12] Stanley C. Eisenstat, Martin H. Schultz, and Andrew H. Sherman. Algorithms and data structures for sparse symmetric Gaussian elimination. *SIAM Journal on Scientific and Statistical Computing*, 2:225–237, 1981.

[13] Miroslav Fiedler. Inversion of bigraphs and connection with the Gauss elimination. In *Graphs, Hypergraphs, and Block Systems*, pages 57–68. Zielona Gora, 1976.

[14] 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.

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

[16] Alan George and Esmond Ng. An implementation of Gaussian elimination with partial pivoting for sparse systems. *SIAM Journal on Scientific and Statistical Computing*, 6:390–409, 1985.

[17] 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.

[18] Alan George and Esmond Ng. On the complexity of sparse QR and LU factorization of finite-element matrices. *SIAM Journal on Scientific and Statistical Computing*, 9:849–861, 1988.

[19] John R. Gilbert. Predicting structure in sparse matrix computations. Technical Report 86–750, Cornell University, 1986.

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

[21] John R. Gilbert and Michael T. Heath. Computing a sparse basis for the null space. *SIAM Journal on Algebraic and Discrete Methods*, 8:446–459, 1987.

[22] John R. Gilbert and Joseph W. H. Liu. Elimination structures for unsymmetric sparse LU factors. Technical Report CSL 90–11, Xerox Palo Alto Research Center, 1991. To appear in SIMAX.

[23] John R. Gilbert and Esmond Ng. Notes on the prediction of fill in some sparse matrix factorizations. In preparation, 1992.

[24] 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.

[25] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 1983.

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

[27] Charles R. Johnson, D. D. Olesky, and P. van den Driessche. Sparsity analysis of the QR factorization, 1991. IMA Sparse Matrix Workshop; to appear in SIMAX.

[28] Charles R. Johnson, D. D. Olesky, and P. van den Driessche. Inherited matrix entries: LU factorization. *SIAM Journal on Matrix Analysis and Applications*, pages 94–104, 89.

[29] Joseph W. H. Liu. Computational models and task scheduling for parallel sparse Cholesky factorization. *Parallel Computing*, 3:327–342, 1986.

[30] L. Lovasz and M. D. Plummer. *Matching Theory.* North Holland, 1986.

[31] Walter Mascarenhas. Predicting structure in eigenvector computation. Technical report, Institute for Mathematics and Its Applications, University of Minnesota, to appear 1992.

[32] S. Thomas McCormick. *A Combinatorial Approach to Some Sparse Matrix Problems.* PhD thesis, Stanford University, 1983.

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

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

[35] Donald J. Rose. A graph-theoretic study of the numerical solution of sparse positive definite systems of linear equations. In Ronald C. Read, editor, *Graph Theory and Computing*, pages 183–217. 1972.

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

[37] 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.

[38] Sven Sigur∂sson. Sparse matrix techniques in geothermal reservoir modelling. In *ICIAM*, 1991.

[39] Robert E. Tarjan and Mihalis Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal on Computing*, 13:566–579, 1984.

[40] Robert Endre Tarjan. A unified approach to path problems. *Journal of the Association for Computing Machinery*, 28:577–593, 1981.