Predicting Structure in Sparse Matrix Computations

John R. Gilbert*

TR 86-750
May 1986

Department of Computer Science
Cornell University
Ithaca, NY 14853

* Computer Science Department, Cornell University, Ithaca, NY 14853. This work was done while the author was a visiting professor at the University of Iceland. Publication of this report was supported by the National Science Foundation under grant DCR-8451385.

Predicting Structure in Sparse Matrix Computations

John R. Gilbert*
May 1986

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.

* Computer Science Department, Cornell University, Ithaca, NY 14853. This work was done while the author was a visiting professor at the University of Iceland. Publication of this report was supported by the National Science Foundation under grant DCR-8451385.

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, by not explicitly storing or manipulating some or all of the zero elements. Sparse matrix algorithms often need to determine the nonzero structure of the result of a computation before doing the computation, using only the nonzero structure of the input. 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.

Some of these results are not new, or at least not difficult. They are collected here because many of them are hard to dig out of papers on various topics in linear algebra and algorithms, and I wanted to put them all down in one framework. 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. Throughout, I have given the earliest reference I am aware of for each result.

2. Definitions.

Structure. Let A be an n by n matrix with nonzero diagonal. The *structure* of A is

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

The *directed graph* of A is the graph whose vertices are the integers from 1 to n and whose edges are $\{\,(i,j) : A_{ij} \neq 0\}$. Therefore the structure of a matrix is the set of edges of its directed graph. (This includes a loop at each vertex for the nonzero diagonal element.) The *structure* of a vector x is

$$\{\,i : x_i \neq 0\},$$

which can be interpreted as a set of vertices in the directed graph of A. Where ambiguity cannot arise, we will not distinguish between a matrix, its structure, and its directed graph, or between a vector and its structure.

In a few places we will use an undirected bipartite graph to represent the structure of a matrix that need not be square. If A is m by n, the *bipartite graph* of A is the graph whose vertices are $r_1, r_2, \ldots, r_m$ and $c_1, c_2, \ldots, c_n$, and whose edges are $\{\,(r_i, c_j) : A_{ij} \neq 0\}$. We will be careful to distinguish the bipartite graph of A from the directed graph of A.

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 [1] and Edenbrandt [6]. 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. We are really interested in the smallest structure that is "big enough" for the result of f with any input of the given structure, which is

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

Brayton, Gustavson, and Willoughby called an algorithm "s-minimal" if it computes this structure from structure(A).

The functions we consider in this paper all have the property that for each structure S, there is a worst-case value A with structure(A) $= S$ such that structure(B) $\subseteq S$ implies structure($f(B)$) $\subseteq$ structure($f(A)$). A function that does not have this property 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.

Graph terminology. Recall that, informally, we are not distinguishing between a graph and the structure of a matrix (which is really a set of edges). Let A be an undirected graph and let x be a subset of the vertices of A. We say x is *closed* (with respect to A) if there is no edge of A from a vertex not in x to a vertex in x; that is, if $x_j \neq 0$ and $A_{ij} \neq 0$ implies $x_i \neq 0$. The *closure* of x (with respect to A) is the smallest closed set containing x,

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

which is the set of vertices of A from which there are paths to vertices of x.

The *transitive closure* of A is the graph whose edges correspond to paths in A,

$$\text{transitive closure}(A) = \{ (i,j) : \exists \text{ a path in } A \text{ from } i \text{ to } j \}.$$

If the vertices of A are numbered $1, 2, \ldots, n$, an interesting subgraph of transitive closure(A) is the *filled graph* of A, which describes paths whose highest-numbered vertices are their endpoints,

$$\text{fill}(A) = \{ (i,j) : \exists \text{ a path in } A \text{ from } i \text{ to } j \text{ whose vertices are all less than } \min(i,j) \}.$$

A graph A is *strongly connected* if there is a path from every vertex to every other vertex. A matrix A is called *irreducible* if structure(A) is strongly connected. The *strongly connected components* (briefly, just *strong components*) of an arbitrary graph A 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 structure(A). (This partition is independent of the choice of a nonzero diagonal for A; see [2] for discussion and references to several different proofs of this fact.)

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 $A_{ij} \neq 0$. An interesting class of undirected graphs 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 $i \neq j \pm 1 \pmod{k}$.

If A is an m by n matrix with no zero columns, $A^T A$ is a symmetric matrix with nonzero diagonal. We use the notation $A^S A$ to denote an arbitrary n by n matrix with a nonzero in position (i, j) if and only if there is a row in which columns i and j of A are both nonzero. Then structure($A^T A$) $\subseteq$ structure($A^S A$); the structures are equal unless there is numerical cancellation in $A^T A$.

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 the bipartite graph of A is adjacent to at least k row vertices.) We say that A has the *strong Hall property* if, for every k with $0 < k < n$, every set of k columns of A contains nonzeros in at least $k + 1$ rows. These properties are related to matchings in the bipartite graph of A. The graph has a matching that covers all its columns if and only if A has the Hall property. If A is square, it is irreducible if and only if its bipartite graph has the strong Hall property. (See Papadimitriou and Steiglitz [15] for background on bipartite matching. Our terminology is from Coleman, Edenbrandt, and Gilbert [2].)

Other definitions. A finite set $\{x_1, \ldots, x_n\}$ of complex numbers is *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.

3. Products. The following trivial 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 structure of an m by n matrix A and an n-vector x be given.*

(i) Whatever values A and x have, structure(Ax) *is a subset of the row vertices of A adjacent to column vertices whose indices are in* structure(x).

(ii) There exist values for the nonzeros of A and x such that structure(Ax) *is equal to the set of row vertices described above.* ∎

The generalization of the theorem to products of matrices is immediate, since each column of AB is A times a column of B.

Recall that $A^S A$ denotes a square matrix whose (i, j) position is nonzero if there is a row in which columns i and j of A are both nonzero. Theorem 3.1 implies that structure($A^T A$) $\subseteq$ structure($A^S A$).

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

***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 [12]. In a later subsection we consider factorization with partial pivoting.)

Theorem 4.1 [17]. *Let a structure for A be given, with nonzero diagonal elements.*

(i) If values are chosen for which A has an LU factorization as above, then $\text{structure}(L + U - I) \subseteq \text{fill}(A)$.

(ii) There exist values for the nonzeros of A such that $\text{structure}(L + U - I) = \text{fill}(A)$. ∎

Rose and Tarjan [17] gave an algorithm for computing fill(A) from A in $O(nm)$ time, where A is n by n with m nonzeros. They also showed that transitiveclosure(A) can be computed in time asymptotically the same as that to compute fill(A), so a faster algorithm to compute fill(A) would give a faster algorithm to compute transitiveclosure(A) than the best currently known.

Remark 4.1. 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 [1] gave a counterexample. Let

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

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

Cholesky factorization. Here we consider the factorization $A = LL^T$, where A is a symmetric, positive definite matrix. 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 4.2 [18]. *Let a symmetric structure for A be given, with nonzero diagonal elements.*

(i) No matter what values A has, if A has a Cholesky factorization $A = LL^T$ then $\text{structure}(L) \subseteq \text{fill}(A)$.

(ii) There exist symmetric values for the nonzeros of A such that $\text{structure}(L + L^T) = \text{fill}(A)$. ∎

Rose, Lueker, and Tarjan [18] gave an $O(n + m + f)$ algorithm to compute fill(A) for symmetric A.

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 4.3 [16]. *Let a symmetric structure for A be given, with nonzero diagonal elements.*

(i) fill(A) *is a chordal graph.*

(ii) Conversely, if structure(A) *is a chordal graph, then its vertices can be renumbered so that* fill(A) = structure(A). ∎

Rose, Leuker, and Tarjan [18] gave an $O(n + m)$ algorithm to determine whether a graph A is chordal and, if so, to reorder its vertices so that fill(A) = structure(A).

Partial pivoting. The example in Section 2 showed that a result of the form of Theorem 4.2 is not possible for LU factorization with partial pivoting. George and Ng [10] 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 $\hat{L} = \sum_{1\le i<n}(L_i - I)$ 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.

Theorem 4.4 [10]. *Let the structure of A be given. Whatever values A has, if Gaussian elimination with partial pivoting gives the factors $\hat{L}$ and U as above, then* structure$(\hat{L} + U) \subseteq$ fill(A^SA). ∎

This theorem is not tight: There may be nonzeros in fill(A^SA) that cannot be nonzero in $\hat{L} + U$ for any pivot sequence. For example, if A is tridiagonal then fill(A^SA) is five-diagonal, predicting that U could be upper tridiagonal; but, in fact, U must be upper bidiagonal. George and Ng [11] suggest a way of predicting the structures of $\hat{L}$ and U by efficiently simulating all possible pivoting steps. When combined with permutation to block triangular form, this method may give a tight prediction, though George and Ng do not claim that it does.

***QR* factorization.** Suppose A is an m by n matrix with $m \le n$. 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^TA, the structure of R can be predicted by forming A^SA and doing structural Cholesky factorization.

Theorem 4.5 [8]. Let the structure 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 structure$(R) \subseteq$ fill(A^SA). ∎

The converse of this theorem is false; for example, if A is square with a nonzero diagonal and a full first row, then structure$(A^SA) =$ fill(A^SA) is full, but the orthogonal factor R is equal to A. Coleman, Edenbrandt, and Gilbert supplied a partial converse.

Theorem 4.6 [2]. Let the structure be given for a rectangular matrix A with at least as many rows as columns. If A has the strong Hall property, then there exist values for the nonzeros of A such that structure(R) = fill($A^S A$), where R is the orthogonal factor of A as above. ∎

If A does not have the strong Hall property, it can be permuted into a block triangular form for which the structure of R can be predicted. See [2] for details. Heath and I experimented with some typical matrices from geodetic least squares problems structural analysis. Most of them had the strong Hall property; we concluded that in practice the converse of Theorem 4.4 is often true.

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 5.1 are straightforward consequences of Tarjan's work on elimination methods for solving path problems in graphs [20]. The proof here is somewhat different.

Throughout this section A is an n by n matrix with nonzero diagonal, and the graph in question is the directed graph of 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

$$\text{structure}(A^{-1}b) \subseteq \text{closure}(b).$$

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

Proof of part (i). Let values be given for which A is nonsingular. Renumber the vertices of A so that closure(b) = $\{1, 2, \ldots, k\}$ for some $k \le n$. 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 closure(b), there is no edge (i, j) with $i \notin$ closure(b) and $j \in$ closure(b). Therefore $C = 0$. Then $Ez = 0$. Since A is nonsingular and $C = 0$, matrix E is nonsingular. Therefore $z = 0$. Thus structure(x) $\subseteq \{1, \ldots, k\}$ = closure(b).

Proof of part (ii). Choose algebraically independent values for the nonzeros of A, and let $b_i = 1$ if $i \in$ structure(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 structure(x) = $\{1, 2, \ldots, k\}$ for some $k \le n$. 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

$$\sum_{1 \le j \le k} c_{ij} y_j = e_i. \tag{5.1}$$

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 5.1 implies

$$\sum_{i \le j \le k} c_{ij} \det(B|_j^d) - e_i \det B = 0. \tag{5.2}$$

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 5.2 implies $e_i = 0$. Thus $e = 0$, so $b = \binom{d}{0}$ and structure(b) $\subseteq$ structure(x) $=$ closure(x). Therefore closure(b) $\subseteq$ structure(x). With part (i), this gives closure(b) $=$ structure(x). ∎

Remark 5.1. 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 structure($\hat{A}$) $\subseteq$ structure(A) and structure($\hat{b}$) $\subseteq$ structure(b) and $\hat{A}$ is nonsingular, then structure($\hat{A}^{-1}\hat{b}$) $\subseteq$ closure(b).

Remark 5.2. 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 structure(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 5.1. *Let the structure of A be given.*

(i) Whatever values A has, if A is nonsingular then structure(A^{-1}) $\subseteq$ transitive closure(A).

(ii) Values exist for the nonzeros of A such that structure(A^{-1}) $=$ transitive closure(A).

Proof. Note that column j of transitive closure(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. ∎

Corollary 5.1 implies that if A is irreducible, then A^{-1} is full unless numerical coincidence occurs. Duff et al. [5] gave another proof of this.

The case where A is symmetric is simpler and less interesting, but the puzzling examples like that in Remark 5.2 do not arise. If A is symmetric and its graph is not connected, then A is block diagonal, and a linear system divides 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 5.2. *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* structure($A^{-1}b$) $= \{1, 2, \ldots, n\}$; *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 5.2 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 $\{1, 2, \ldots, n\}$. ∎

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 5.2 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.

Theorem 6.1. *Let the structure of 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 $\text{structure}(u^{(i)}) \subseteq \text{closure}(e^{(i)})$. *(Recall that $e^{(i)}$ is the i^{th} unit vector and* $\text{closure}(e^{(i)})$ *is the structure of column i of the transitive closure of A.)*

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

Proof of 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 $1 < i \le n$.

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 $\text{closure}(e^{(i)}) = \text{closure}(B_j)$ (where $\text{closure}(B_j)$ 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,

$$\text{structure}(v) \subseteq \text{closure}(Ew).$$

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

Proof of part (ii). Choose algebraically independent values for A, choosing the diagonal elements so far apart that no two are closer than $2\max_j \sum_{i \neq j} |a_{ij}|$. By Gerschgorin's theorem [12], 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, and it seems that it ought to be possible, but I don't 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 structure$(u) = \{1, 2, \ldots, t\}$ for some $t \le n$. Then $Au = \lambda u$ can be partitioned as

$$\begin{pmatrix} B & D \\ C & E \end{pmatrix} \begin{pmatrix} v \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} v \\ 0 \end{pmatrix}, \tag{6.1}$$

where B is t by t and $v_k \neq 0$ for $1 \le k \le t$. We will show $C = 0$. Intuitively, it seems clear that if $C \neq 0$ then v cannot be both an eigenvector of B and a null vector of C. The details are field theoretic.

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

$$v_k = \frac{\det(B' - \lambda I)|_k^{-g}}{\det(B' - \lambda I)} \qquad \text{for } 2 \le k \le t. \tag{6.2}$$

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 $F[x]$. (Incidentally, this is our only bit of nontrivial field theory: Every element of an algebraic extension of a field is algebraic over the field. I find it amusing that this fact from field theory, which we are applying to a problem in linear algebra, is ordinarily proven by applying linear algebra to field theory.)

Since $Cv = 0$, we have

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

All the v_k are nonzero, so

$$c_{ij} = -\frac{1}{v_j} \sum_{k \ne j} c_{ik} v_k. \tag{6.3}$$

By equation 6.2, each v_k is a rational function of λ and elements of F, so $v_k \in F(\lambda)$. Each c_{ik} with $k \ne j$ is in F. Therefore the whole right hand side of equation 6.3 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 6.1, $C = 0$ 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 6.1 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 $u_i^{(i)} \ne 0$. Therefore, structure($e^{(i)}$) $\subseteq$ structure($u^{(i)}$). Since $u^{(i)}$ is closed, closure($e^{(i)}$) $\subseteq$ structure($u^{(i)}$). Part (i) gives the opposite containment, so structure($u^{(i)}$) = closure($e^{(i)}$). ∎

Corollary 6.1. *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 structure(U) $\subseteq$ transitive closure(A).

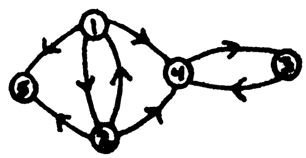
(ii) There exist values for the nonzeros of A such that the eigenvectors can be ordered so that structure(U) = transitive closure(A).

Proof. Similar to Corollary 5.1. ∎

Remark 6.1. 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 5.2 we can construct

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

The graph of A is



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.

I suspect that Theorem 6.1 holds for A with multiple eigenvalues, provided that no two diagonal blocks of the block upper triangular form of A share an eigenvalue. I do not know whether it would be enough to require that there be n linearly independent eigenvectors; that is, that each eigenvalue have 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.

Theorem 6.2. *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 5.2, 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. 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.

Matchings in bipartite graphs are important in Theorem 4.6 on orthogonal factorization. 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 [14]) and for the null space (Pothen and Coleman [4]) of a rectangular matrix with more columns than rows. It turns out that the structural range space problem can be solved in polynomial time, but the null space problem is NP-complete.

Structure prediction has several applications in the design of sparse matrix algorithms. Many such algorithms [8, 9, 10] have a phase that predicts the solution's structure without using the numbers in 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 needs to be solved, so the structural phase can be done just once.

Sparse Gaussian elimination seems to require structure prediction if the time spent manipulating data structures is not to dominate the time spent doing real arithmetic. The asymptotically fastest algorithms to compute the Cholesky factorization of a symmetric positive definite matrix are those of the Yale Sparse Matrix Package [7] and Waterloo Sparspak [9], which predict the structure of the triangular factor by a version of Theorem 4.2. Coleman, Peierls, and I [3] used prediction of the structure of the solution of a triangular system, a special case of Theorem 5.1, to develop an algorithm that performs sparse LU factorization with partial pivoting in time proportional to the number of real arithmetic operations needed.

Another application of structure prediction for triangular linear 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. Ragnar Sigurðsson [19] has used structure prediction with a simpler version of Theorem 5.1 to speed up the Sparspak triangular solver for this problem.

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 is the structure of the factor R in the QR factorization of a rectangular matrix that does not have the strong Hall property and has not been permuted to block triangular form (Section 4)? What can be said about the structure of the orthogonal factor Q in QR factorization (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)? What can be said about the structure of the singular value decomposition of a rectangular matrix [12]? The relationship between the singular values of A and the eigenvalues of A^TA, together with Theorem 6.2 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 a sparse SVD.

Acknowledgements. My thanks to Tom Coleman, Anders Edenbrandt, Mike Heath, Ragnar Sigurðsson, and Sven Sigurðsson for interesting and useful discussions of these problems. Earl Zmijewski gave this paper a careful and helpful critical reading.

References.

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

[2] Thomas F. Coleman, Anders Edenbrandt, and John R. Gilbert. Predicting fill for sparse orthogonal factorization. Cornell University Computer Science Department Report 83-578, 1983. To appear in *Journal of the ACM.*

[3] Thomas F. Coleman, John R. Gilbert, and Timothy Peierls. Sparse partial pivoting in time proportional to arithmetic operations. In preparation, 1986.

[4] Thomas F. Coleman and Alex Pothen. The null space problem I: Complexity. To appear in *SIAM Journal on Algebraic and Discrete Methods.*

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

[6] Anders Gunnar Edenbrandt. *Combinatorial Problems in Matrix Computation.* Ph. D. thesis, Cornell University, 1985.

[7] S. C. Eisenstat, M. C. Gursky, M. H. Schultz, and A. H. Sherman. Yale sparse matrix package I: The symmetric codes. *International Journal for Numerical Methods in Engineering* **18**: 1145–1151, 1982.

[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 W-H Liu. *Computer Solution of Large Sparse Positive Definite Systems.* Prentice-Hall, 1981.

[10] 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.

[11] Alan George and Esmond Ng. Symbolic factorization for sparse Gaussian elimination with partial pivoting. University of Waterloo Computer Science Department Report CS-84-43, 1984. To appear in *SIAM Journal on Scientific and Statistical Computing.*

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

[13] Frank Harary. *Graph Theory.* Addison-Wesley, 1969.

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

[15] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Prentice-Hall, 1982.

[16] 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. Academic Press, 1972.

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

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

[19] Ragnar Sigurðsson. Sparse matrix techniques in geothermal reservoir modelling. In preparation, 1986. Mathematics Department, University of Iceland, 101 Reykjavík, Iceland.

[20] Robert Endre Tarjan. A unified approach to path problems. *Journal of the ACM* **28**: 577–593, 1981.