Separators and Structure Prediction in Sparse Orthogonal Factorization

John R. Gilbert* Esmond G. Ng† Barry W. Peyton†

July 8, 1993

Abstract

In the factorization $A = QR$ of a matrix A, the orthogonal matrix Q can be represented either explicitly (as a matrix) or implicitly (as a matrix H of Householder vectors). We derive both upper and lower bounds on the number of nonzeros in H and the number of nonzeros in Q, in the case where the graph of A^TA has "good" separators and A need not be square. We also derive an upper bound on the number of nonzeros in the null-basis part of Q in the case where A is the edge-vertex incidence matrix of a planar graph. The significance of these results is that they both illuminate and amplify a folk theorem of sparse QR factorization, which holds that the matrix H of Householder vectors represents the orthogonal factor of A much more compactly than Q itself. To facilitate discussion of this and related issues, we review several related results which have appeared previously.

Keywords: Sparse matrix algorithms, QR factorization, separators, column intersection graph, strong Hall property.

AMS(MOS) subject classifications: 65F05,65F50,68R10.

*Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304-1314 (gilbert@parc.xerox.com).

†Mathematical Sciences Section, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6367 (esmond@msr.epm.ornl.gov, peyton@sky.epm.ornl.gov). These authors' work was supported by the Applied Mathematical Sciences Research Program of the Office of Energy Research, U.S. Department of Energy, under contract DE-AC05-84OR21400.

1 Introduction

The orthogonal factorization $A = QR$ is central to direct methods for least squares and eigenvalue problems. Let A be an m by n matrix of full column rank, so that $m \geq n$. The "narrow" QR factorization is $A = QR$, where Q is an m by n matrix with orthonormal columns that span the range of A, and R is an n by n upper triangular matrix with positive diagonal elements. This factorization is unique. In the "wide" QR factorization, Q is square, and its last $m - n$ columns span the null space of A^T. The narrow factorization is sufficient for many applications. Most of this paper discusses the narrow factorization; Section 5 considers the wide one in a special case.

The orthogonal factor Q is often stored implicitly, as follows. Let $A = H_1 H_2 \cdots H_n R$, where H_i is a Householder reflection that zeros out column i of A below the main diagonal. The reflection $H_i = I - h_i h_i^T$ is a rank-1 modification to the identity, and can be represented compactly by the column vector h_i. The *Householder matrix* $H = [h_1, \ldots, h_n]$ is the concatenation of these vectors. Since h_i is zero in positions 1 through $i - 1$, the Householder matrix is lower trapezoidal. A matrix-vector product Qa can be computed efficiently from H and a [6].

Figure 1: A matrix A with 452 rows and columns, and its factors Q, R, and H.

In the context of sparse factorization, Q is often observed to have many more nonzeros than H. For example, in the factorization of the square

Figure 2: A matrix A with 219 rows and 85 columns, and its factors Q, R, and H.

matrix in Figure 1, Q is more than ten times as dense as H, which in turn is about twice as dense as A. (The triangular factor R here is about four times as dense as A.) Figure 2 shows a rectangular matrix in which the difference between H and Q is much less dramatic: here Q has only 50% more nonzeros than H, and both are an order of magnitude denser than A.

Hereafter we write $|B|$ for the number of nonzero entries in a matrix B. In this paper, we analyze $|Q|$, $|H|$, and $|R|$ in terms of the sizes of *separators* in the graph of A^TA (i.e., the column intersection graph of A). We use the bounds on $|H|$ and $|Q|$ to measure the cost of storing H against that of storing Q under various realistic scenarios.

The rest of the paper is organized as follows. Section 2 briefly reviews the graph-theoretic model of QR factorization that we use. George and Ng [11] obtained upper bounds on $|H|$ and $|R|$ in the case where A is a certain square finite element matrix whose column intersection graph has good separators. Section 3 gives upper bounds on $|Q|$, $|H|$, and $|R|$ in the case where the only restriction on A is that its column intersection graph has good separators. These upper bounds follow in a straightforward manner from results in George, Heath, Lipton, Liu, Ng, Rose, and Tarjan [8, 9, 11, 22]. In Section 4 we show that the upper bounds given in Section 3 are the best possible, at least for a large class of matrices. The argument combines some new analysis of separators in the intersection graph of A with a sequence of results of Coleman, Edenbrandt, Gilbert, Hare, Johnson, Ng, Olesky, Peyton, Pothen, and van den Driessche [4, 19, 26, 28]. In Section 5 we show that, in one interesting special case (i.e., when A is an edge-vertex incidence matrix of a planar graph), the null space part of Q has about the same density as H and as the narrow part of Q. Finally, Section 6 concludes with a few remarks.

2 Graph models of QR factorization

Three graphs model sparsity structure in QR factorization: the column intersection graph, the filled column intersection graph, and the column elimination tree. Gilbert and Ng [16] discuss these graphs in detail; here we review them briefly. All three graphs have the same set of vertices, namely the integers $1, \ldots, n$, representing the columns of A.

Matrix $A = (a_{ij})$ is m by n, with $m \geq n$ and rank n. For a matrix B and a graph G, we write $B \subseteq G$ if for each entry $b_{ij} \neq 0$ the edge $\langle i, j \rangle$ is included in G. Throughout the paper we assume that the diagonal elements of A are nonzero. Every matrix with full column rank admits a row permutation that makes the diagonal elements nonzero [18]. The three graphs we define below are independent of the row ordering of A.

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

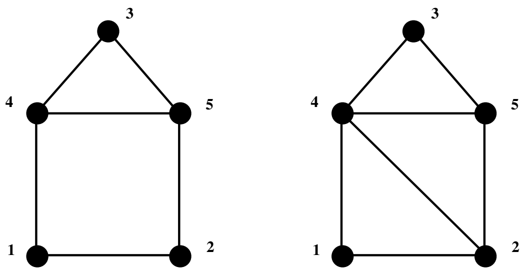


Figure 3: A matrix A, its column intersection graph $G_\cap(A)$, and its filled column intersection graph $G_\cap^+(A)$.

The *column intersection graph* of A is the undirected graph $G_\cap(A)$ in which an edge joins two vertices whose columns share a nonzero row in A:

$$G_\cap(A) = (\{1:n\}, \{\langle i, j \rangle : \exists k \text{ with } a_{ki} \neq 0 \text{ and } a_{kj} \neq 0\}).$$

Figure 3 shows an example. This graph corresponds to the matrix of the normal equations: each symmetric pair of off-diagonal nonzeros in A^TA is represented by an edge of $G_\cap(A)$, hence we write

$$A^TA \subseteq G_\cap(A). \tag{1}$$

Clearly, containment in (1) is proper if and only if no inner product of a pair of "intersecting" columns of A is zero. Since a zero inner product of this sort occurs with probability zero over all choices of values for the nonzeros of A, we call equality in (1) the *generic case*.

The *filled column intersection graph* of A is the undirected graph $G_\cap^+(A)$ that results from performing symbolic Cholesky factorization [10] on $G_\cap(A)$. In one characterization of $G_\cap^+(A)$, due to Rose, Tarjan, and Lueker [29], an edge joins vertices i and j in $G_\cap^+(A)$ whenever $G_\cap(A)$ contains a path between i and j that passes only through vertices with numbers less than $\min(i,j)$. Figure 3 shows an example of a filled column intersection graph.

The filled column intersection graph represents the nonzero structure of the Cholesky factor of the normal equations matrix A^TA. Since that Cholesky factor is equal to R in the QR factorization,

$$R + R^T \subseteq G_\cap^+(A). \tag{2}$$

George and Heath [8] use this containment to design data structures for sparse QR factorization. The conditions under which we have equality in (2) are subtle; several of the references study this question, and we return to it in Section 4. In cases where equality in (2) occurs with probability one, we again call equality the *generic case*; we apply the term to such statements for other factors of A in the same fashion.

In the *column elimination tree* of A, denoted $T_\cap(A)$, the parent of vertex i is the smallest $j > i$ adjacent to i in $G_\cap^+(A)$; if i has no larger neighbor then i is a root. This "tree" is actually a forest with one connected component for each connected component of $G_\cap(A)$. Figure 4 shows an example. Elimination trees have many uses in sparse factorization; Liu [24] gives a survey. In a sense, this tree represents dependences among the columns in QR (and other) factorizations. The column elimination tree is a depth-first spanning tree of $G_\cap^+(A)$, which means that every edge of $G_\cap^+(A)$ joins an ancestor and a descendant in the tree.

George, Liu, Ng, and Peyton [9, 26] use paths in this tree to bound the nonzero structure of the rows of Q and H. Write $Q = [Q_1; \cdots; Q_m]$, $H = [H_1; \cdots; H_m]$, and $A = [A_1; \cdots; A_m]$ by rows (as in Matlab notation). Then the structure of Q_i (as a set of column indices) is a subset of the

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

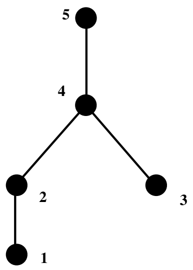


Figure 4: The orthogonal factor Q and the column elimination tree $T_\cap(A)$ for the matrix in Figure 3.

path in $T_\cap(A)$ from the vertex representing first nonzero index in row A_i to the root of $T_\cap(A)$ (more precisely, the root of that connected component of $T_\cap(A)$ containing the first nonzero index in row A_i). Note that all the nonzero column indices in row A_i occur along this path.

The structure of H_i is also a subset of this path in $T_\cap(A)$, but since H is lower trapezoidal, a stronger statement is true: the structure of H_i (as a set of column indices) is a subset of the path in $T_\cap(A)$ from the first nonzero index in row A_i to the appropriate root or to vertex i, whichever is smaller.

Ng and Peyton [26] also give conditions under which the containments are tight. We discuss some of these conditions in Section 4.

3 Upper bounds

We now bound $|Q|$, $|R|$, and $|H|$ from above under the hypothesis that the column intersection graph $G_\cap(A)$ and its subgraphs have good *separators*. A separator is a small set of vertices whose removal divides a graph approximately in half. Formally, we say that a class of graphs has *n^λ-separators*, for some $\lambda < 1$, if every n-vertex subgraph of a graph in the class has a set of $O(n^\lambda)$ vertices whose removal leaves no connected component with more than $\frac{2}{3}n$ vertices. (Any constant between 0 and 1 could be substituted for $\frac{2}{3}$, changing only the constant in $O(n^\lambda)$ [23].) For example, planar graphs have $n^{1/2}$-separators [23], as do graphs of bounded genus [14] and graphs that omit any fixed minor [2]; finite element meshes of bounded aspect ratio in d dimensions have $n^{1-1/d}$-separators [25].

Theorem 1 *Let A be an m by n matrix of full column rank, such that $G_\cap(A)$ has $\sqrt{n}$-separators for all its subgraphs. Then there exists a column permutation P such that the matrices Q, R, and H in the narrow QR factorization of AP satisfy the following bounds:*

- $|R| = O(n \log n)$.
- $|Q| = O(m\sqrt{n}\,)$.
- $|H| = O(n \log n + (m-n)\sqrt{n}\,)$.

Proof: A column permutation P of A corresponds to a labeling of the vertices of the column intersection graph $G_\cap(A)$ with the integers $[1 : n]$. We take P to be a nested dissection ordering: find a separator in the graph, number its vertices last, then recursively number the fragments of the graph.

To prove the theorem for all $\sqrt{n}$-separable graphs, we must use Lipton, Rose, and Tarjan's "generalized nested dissection" ordering [22], which includes the separator vertices themselves in each recursive call (though it does not reorder them). George's original, simpler form of nested dissection [7], which does not include the separators in the recursive call, is actually sufficient for all of the specific classes of $\sqrt{n}$-separable graphs named above: planar graphs, graphs of bounded genus or bounded excluded minor, and two-dimensional finite element meshes of bounded aspect ratio [17].

The labeling of the vertices of $G_\cap(A)$ yields a column permutation P for A. Now we count the nonzeros in the factors R, Q, and H of AP. (We shall assume that the rows of AP have been permuted to make the diagonal nonzero; but to simplify notation we won't mention this row permutation explicitly.)

First, consider R. Since $R + R^T \subseteq G^+_\cap(AP)$, Lipton, Rose, and Tarjan's analysis of Cholesky factorization with generalized nested dissection [22] shows that $|R| = O(n \log n)$.

Second, consider Q. First we bound the height of the column elimination tree $T_\cap(AP)$. A path from a leaf to a root of $T_\cap(AP)$ includes vertices from at most one separator at each level of the nested dissection recursion. Since the relative decrease in graph size is bounded above by a constant $c \le \frac{2}{3} < 1$ at each level, and the separator size is proportional to the square root of the graph size, the sum of the separator sizes on the path (and the height of the tree) is $O(\sqrt{n}\,)$. Each row Q_i is a subset of a path in the tree, so $|Q_i| = O(\sqrt{n}\,)$ for each i, and $|Q| = O(m\sqrt{n}\,)$.

Third, consider the first n rows of H, which are zero above the main diagonal. George and Ng [11] prove that the structure of these rows of H

is a subset of the structure of the symbolic Cholesky factor of $P^T A^T A P$, which is $G_\cap^+(AP)$. (George and Ng actually consider only the case of square A, but their result extends easily to the non-square case [16].) Therefore these n rows of H have $O(n \log n)$ nonzeros in all.

Finally, consider the last $m - n$ rows of H. The structure of each such row H_i is a subset of the path in $T_\cap(AP)$ from the first nonzero in row A_i to a root. Therefore $|H_i| = O(\sqrt{n}\,)$ for each row i, $n+1 \le i \le m$, and thus $|H| = O(n \log n + (m-n)\sqrt{n}\,)$. □

A similar analysis leads to bounds for graphs with larger separators.

Proposition 1 *Let A be an m by n matrix of full column rank, such that $G_\cap(A)$ has n^λ-separators for all its subgraphs, for some $\frac{1}{2} < \lambda < 1$. Then there exists a column permutation P such that the factors Q, R, and H of AP satisfy the following bounds:*

- $|R| = O(n^{2\lambda})$.
- $|Q| = O(mn^\lambda)$.
- $|H| = O(n^{2\lambda} + (m-n)n^\lambda)$.

□

Theorem 1 says roughly that there are two regimes in the comparison between the storage required for Q and that for H. If A is square or nearly so (that is, if $m - n = o(n)$), then Q has asymptotically fewer nonzeros than a dense matrix ($O(n^{3/2})$ instead of n^2), but many more than H (which has only $O(n \log n)$). On the other hand, if $m - n$ is on the same order as n, then H is smaller than Q only by a constant factor, and if m is much larger than n then the difference between the sizes of Q and H is likely to be relatively small.

4 Lower bounds

We can ask two questions about the tightness of the results in Section 3. First, are the upper bounds on nonzero count in Theorem 1 tight for the column nested dissection ordering? This amounts to asking whether the containments $R + R^T \subseteq G_\cap^+(A)$ and row $Q_i \subseteq$ path in $T_\cap(A)$ are tight. Second, is column nested dissection in fact the best ordering for the matrices in the hypothesis of Theorem 1? This section answers both questions affirmatively, in a sense: the upper bounds are tight if the matrix satisfies an irreducibility condition, and the column ordering is within a constant of best possible for a large class of matrices.

An m by n matrix A is *strong Hall* [4] if, for k satisfying $0 < k < n$, every set of k columns of A collectively contains nonzeros in at least $k+1$ rows. (This condition implies $m \geq n$.) The strong Hall property can be viewed as an irreducibility condition under row and column permutations: for a square matrix A, "strong Hall" is the same as "fully indecomposable"; further, any matrix has a so-called "Dulmage-Mendelson decomposition" that permutes its rows and columns to a block triangular form with diagonal blocks that are strong Hall. The following proposition, which collects several known results, says that our various containments are tight if the matrix is strong Hall.

Proposition 2 *Suppose A is an m by n strong Hall matrix, with factors Q, R, and H. Then in the generic case the following are all true:*

1. *A has full column rank n [18].*
2. *The structure of $R + R^T$ is equal to that of $G^+_\cap(A)$ [4].*
3. *The nonzero column indices of row Q_i are exactly those on the tree path in $T_\cap(A)$ from the first nonzero column index of row A_i to a root of $T_\cap(A)$ [19, 28, 26].*
4. *The rows of A can be permuted to make the diagonal nonzero, and in this case the structure of H is equal to the lower trapezoidal part of the structure of Q [26].*

□

Hare et al. [19] and Pothen [28] also characterize the structures of Q and R when A is of full rank but is not strong Hall; and Ng and Peyton [26] also discuss the general case. Gilbert and Ng [16] perform a related analysis of the structures of the factors of A in Gaussian elimination with partial pivoting.

If A is not strong Hall, the containments may not be tight. For example, consider a square matrix A whose first row and main diagonal are full, and whose other entries are all zero. In this case, Q is diagonal (as is H) and R has the same structure as A; but $T_\cap(A)$ is a single path and $G^+_\cap(A)$ is a complete graph, so the containments would predict that Q and R are full upper triangular matrices.

We now turn to the question of optimality of the nested dissection ordering. It is known in several contexts that the best attainable matrix orderings are related to the best separators in a graph [1, 3, 12], and that turns out to be the case here as well.

Consider a square grid graph, with $n = k^2$ vertices in a square array and each vertex adjacent to its four nearest neighbors. Hoffman, Martin, and Rose [21] show that in the generic case the Cholesky factor of a matrix with this structure, must have $\Omega(n \log n)$ nonzeros, for any elimination ordering. (We use the Ω notation for asymptotic lower bounds: $f(n) = \Omega(g(n))$ means the same as $g(n) = O(f(n))$.) Thus the nested dissection bound on R in Theorem 1 is tight to within a constant factor.

Proposition 3 *Let A be a strong Hall matrix such that $G_\cap(A)$ is an n-vertex square grid graph. Then, assuming the generic case, the upper triangular factor R in the orthogonal factorization $AP = QR$ has $\Omega(n \log n)$ nonzeros for every column permutation P.* $\square$

For Q, we can show directly that the best separator in $G_\cap(A)$ bounds the best column order from below.

Theorem 2 *Let A be a strong Hall matrix with n columns, and suppose that at least s vertices must be deleted from the graph $G_\cap(A)$ to leave connected components of size at most $n/2$. Then, assuming the generic case, the orthogonal factor Q in the factorization $AP = QR$ has $\Omega(ns)$ nonzeros for every column permutation P.*

Proof: Consider the column elimination tree $T_\cap(AP)$. We consider a vertex v as one of its own descendants; the *proper descendants* of v are all its descendants except v itself.

Let S be the set of tree vertices that have more than $n/2$ descendants. If S is empty, then the graph $G_\cap(AP)$ is disconnected and has no component with more than $n/2$ vertices, whence $s = 0$ and there is nothing to prove. If S is nonempty, it includes exactly one root, and each vertex in S has at most one child in S, since two disjoint subtrees cannot both have more than $n/2$ vertices. Therefore S consists of a path from a root to some vertex y. If we delete the vertices of S from $G_\cap(AP)$, each remaining connected component is a subtree of $T_\cap(AP)$. Each such subtree's root has at most $n/2$ descendants (since it is not in S). Thus each connected component has at most $n/2$ vertices. The hypotheses of the theorem then say that S has at least s vertices.

Let Y be the set of vertices (i.e. integer column indices of AP) that are proper descendants of y. Since $y \in S$ has more than $n/2$ descendants, Y has at least $n/2$ elements. Let X be the set of indices of rows of AP that have nonzeros in columns from Y. Since A is strong Hall, X is larger than Y and has more than $n/2$ elements. Now consider a row index $i \in X$. The tree path in $T_\cap(AP)$ that describes the structure of row Q_i includes some vertex in Y, and therefore includes y and the entire path S from y to the root. Therefore row Q_i has at least s nonzeros. Since X has more than $n/2$ elements, Q has more than $ns/2 = \Omega(ns)$ nonzeros in all. □

As corollaries to this theorem, we can show that several of the bounds in Section 3 are tight. The square grid graph satisfies the hypothesis of Theorem 2 with $s = \Omega(\sqrt{n}\,)$, as does a planar graph or 2-dimensional finite element graph that is not "too skinny." Let G be any of these graphs—for example, the square grid graph with n vertices and $m \approx 2n$ edges—and let A be a matrix that has a column for each vertex of G, a row for each edge of G, and a nonzero in each position corresponding to an endpoint of an edge. (This is the nonzero structure of the *edge-vertex incidence matrix* of G, which we discuss further in the next section.) Then the column intersection graph of A is G. Theorem 2 applies with $s = \Omega(\sqrt{n}\,)$, and shows that the bounds on $|Q|$ and $|H|$ in Theorem 1 are tight.

That example has $m = O(n)$. To construct examples where m is arbitrarily large relative to n, simply take multiple copies of the rows of A. This does not change $G_\cap(A)$, but the analysis in the proof of Theorem 2 now shows that $|Q|$ and $|H|$ are $\Omega(m\sqrt{n}\,)$.

5 The null basis

The QR factorization of A is given by $A = QR$, where Q is m by m and orthogonal, and R is m by n and upper trapezoidal. In this case the last $m - n$ columns of Q are an orthonormal basis for the null space of A^T, and are not in general uniquely determined by A. Sparse null bases for sparse matrices have been studied extensively [5, 13, 20, 30]. In some cases, the sparsest orthonormal null basis is much denser than the sparsest unrestricted null basis [27].

We have not made a complete analysis of orthonormal null bases in the context of sparse QR factorization, but we present in this section a result showing that in one interesting special case the null-basis part of Q is about as sparse as the rest of Q. The special case is the edge-vertex incidence matrix of a planar graph, which we now define.

Let G be any graph with m edges and n vertices. Its *edge-vertex incidence matrix* A has m rows, one for each edge, and n columns, one for each vertex. Each row has two nonzeros, one equal to $+1$ and one equal to -1. If the edge corresponding to row k has endpoint vertices i and j, then $a_{ki} = +1$ and $a_{kj} = -1$ or vice versa. (The choice of a positive and negative endpoint for each edge is arbitrary; we think of it as establishing an arbitrary direction for each edge, from its $+1$ to its -1 endpoint.) This matrix arises in many settings, including finite element analysis and electrical circuits. Strang [31, pp. 70–74] gives a good introduction to incidence matrices and their properties.

It is easy to see that the original graph G is equal to $G_\cap(A)$, the column intersection graph of its edge-vertex incidence matrix. Since each row of A contains one $+1$ and one -1, the sum of the columns of A is zero, and the rank of A is at most $n - 1$. In fact, if the graph G has c connected components, then the rank of A is exactly $n - c$.

Theorem 3 *Let A be the m by n edge-vertex incidence matrix of a connected planar graph G with n vertices and m edges. Then after a suitable column permutation, A has a QR factorization in which R has $O(n \log n)$ nonzeros, and Q (including the null-basis part) has $O(n^{3/2})$ nonzeros.*

Proof: Since $G = G_\cap(A)$ is planar, it has $\sqrt{n}$-separators, and Theorem 1 says that ordering the columns of A by nested dissection gives a QR factorization in which R has $O(n \log n)$ nonzeros and the first n columns of Q have $O(n^{3/2})$ nonzeros. It remains to show that A^T has an orthonormal null basis with $O(n^{3/2})$ nonzeros. (Note that permuting the columns of A does not change the null space of A^T.)

Since G is connected, A^T has rank $n - 1$. Thus any null basis of A^T has $m - n + 1$ columns. Every cycle in G determines a null vector of A^T as follows: traverse the edges around the cycle, adding the columns of A^T whose edges are traversed "forward" from their $+1$ to -1 ends, and subtracting the columns whose edges are traversed "backward" from their -1 to $+1$ ends. The sum is zero; thus the vector with $+1$ for each forward edge in the cycle and -1 for each backward edge is a null vector of A^T.

Now embed G in the plane. Euler's theorem says that the embedded graph divides the plane into $m - n + 1$ finite faces (plus one infinite face). Each face is bounded by a cycle in the graph G. The $m - n + 1$ null vectors corresponding to cycles bounding finite faces are linearly independent, and constitute the *facial null basis* of A^T. Let B be the matrix whose columns are those null vectors, so that B has m rows and $m-n+1$ columns. The null basis B has no more than $2m$ nonzeros (because each edge occurs at most twice in a facial cycle), but the columns of B are not mutually orthogonal.

Consider the column intersection graph $G_\cap(B)$. The vertices of $G_\cap(B)$ are the finite faces of the embedding of G; two faces are adjacent in $G_\cap(B)$ if their boundary cycles share an edge in G. Therefore $G_\cap(B)$ is the planar dual of the graph G, and is itself planar. Theorem 1 thus implies that there is a QR factorization of a column permutation of B, say $BP = Q_B R_B$, in which Q_B has only $O(m\sqrt{m-n+1}\,)$ nonzeros. The columns of Q_B are an orthonormal basis of the column space of B, which is the null space of A^T. Thus Q_B is an orthonormal null basis for A^T. Since G is planar and connected, we have $n-1 \le m \le 3n-3$, and hence Q_B has $O(n^{3/2})$ nonzeros, as desired. $\Box$

We conjecture that the bound $O(n^{3/2})$ is the tightest possible in general, and that the edge-vertex incidence matrix of a planar graph satisfying the separator condition in the hypothesis of Theorem 2 has $\Omega(ns)$ nonzeros in every null basis.

Comparing the density of Q and H for these matrices, we see that H is sparser (with $O(n\log n + (m-n)\sqrt{n}\,)$ nonzeros as against $O(m\sqrt{n}\,) = O(n^{3/2})$ for the "wide" Q); but if the graph has enough edges that $m-n$ is comparable to n, the densities differ only by a constant factor.

Stephen Vavasis (personal communication) has extended the result in this section to all matrices A for which the column intersection graph $G_\cap(A)$ has $\sqrt{n}$-separators.

6 Remarks

A folk theorem of sparse QR factorization holds that the matrix H of Householder vectors represents the orthogonal factor much more compactly than the matrix Q itself. This paper is essentially an attempt to make this folk theorem precise. We have shown that for a common class of square matrices, namely matrices whose column intersection graphs have $\sqrt{n}$-separators, Q is much sparser than a full matrix (having only $O(n^{3/2})$ nonzeros), but much denser than H or R (both with $O(n\log n)$ nonzeros). On the other hand, if the matrix is m by n with m/n bounded away from one, then both H and Q can have $O(m\sqrt{n}\,)$ nonzeros; and if m is much larger than n then the difference between H and Q is a lower-order term. Thus, for matrices that are far from square, there may not be much savings in storing the Householder vectors instead of the explicit matrix Q.

This comparison is in terms of the "narrow" QR factorization, in which only the range-space part of Q is stored. The Householder matrix has the advantage that it implicitly represents the null-space part of Q as well; and if the null space is important, that advantage grows as the matrix becomes farther from square. For one limited class of matrices—edge-vertex incidence matrices of planar graphs—we showed that the null-space part of Q, like H, can have only $O(n^{3/2})$ nonzeros; but even then H is more compact by a significant constant factor.

This work was partly motivated by a project to add sparse matrix data structures and algorithms to Matlab [15]. In this setting, there is an esthetic advantage to returning Q instead of H as the result of a QR factorization: the philosophy of Matlab is that everything is a matrix, and an orthogonal matrix should be represented no differently than any other. At the time of this writing, Matlab is still without a built-in sparse QR factorization, but at least we now have a more quantitative understanding of the tradeoff between esthetics and storage.

References

[1] Ajit Agrawal and Philip Klein. Cutting down on fill using nested dissection: Provably good elimination orderings. In Alan George, John R. Gilbert, and Joseph W. H. Liu, editors, *Graph Theory and Sparse Matrix Computation*. Springer-Verlag, 1993.

[2] Noga Alon, Paul Seymour, and Robin Thomas. A separator theorem for graphs with an excluded minor and applications. In *Proceedings of the 22th Annual Symposium on Theory of Computing*. ACM, 1990.

[3] Hans Bodlaender, Hjálmtýr Hafsteinsson, John R. Gilbert, and Ton Kloks. Approximating treewidth, pathwidth, and minimum elimination tree height. In *Proceedings of the 17th International Workshop on Graph-Theoretic Concepts in Computer Science*, volume 570, pages 1–12. Springer-Verlag, 1992. Submitted to *Journal of Algorithms*.

[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 II: algorithms. *SIAM Journal on Algebraic and Discrete Methods*, 8:544–563, 1987.

[6] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart. *LINPACK Users' Guide*. SIAM Press, 1979.

[7] Alan George. Nested dissection of a regular finite element mesh. *SIAM Journal on Numerical Analysis*, 10:345–363, 1973.

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

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

[11] 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.

[12] John R. Gilbert. Some nested dissection order is nearly optimal. *Information Processing Letters*, 26:325–328, 1988.

[13] 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.

[14] John R. Gilbert, Joan P. Hutchinson, and Robert Endre Tarjan. A separator theorem for graphs of bounded genus. *Journal of Algorithms*, 5:391–407, 1984.

[15] John R. Gilbert, Cleve Moler, and Robert Schreiber. Sparse matrices in Matlab: Design and implementation. *SIAM Journal on Matrix Analysis and Applications*, 13:333–356, 1992.

[16] John R. Gilbert and Esmond Ng. Predicting structure in nonsymmetric sparse matrix factorizations. In Alan George, John R. Gilbert, and Joseph W. H. Liu, editors, *Graph Theory and Sparse Matrix Computation*. Springer-Verlag, 1993.

[17] John R. Gilbert and Robert Endre Tarjan. The analysis of a nested dissection algorithm. *Numerische Mathematik*, 50:377–404, 1987.

[18] P. Hall. On representatives of subsets. *Journal of the London Mathematical Society*, 10:26–30, 1935.

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

[20] M. T. Heath, R. J. Plemmons, and R. C. Ward. Sparse orthogonal schemes for structural optimization using the force method. *SIAM Journal on Scientific and Statistical Computing*, 5:514–532, 1984.

[21] Alan J. Hoffman, Michael S. Martin, and Donald J. Rose. Complexity bounds for regular finite difference and finite element grids. *SIAM Journal on Numerical Analysis*, 10:364–369, 1973.

[22] Richard J. Lipton, Donald J. Rose, and Robert Endre Tarjan. Generalized nested dissection. *SIAM Journal on Numerical Analysis*, 16:346–358, 1979.

[23] Richard J. Lipton and Robert Endre Tarjan. A separator theorem for planar graphs. *SIAM Journal on Applied Mathematics*, 36:177–189, 1979.

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

[25] Gary L. Miller, Shang-Hua Teng, William Thurston, and Stephen A. Vavasis. Automatic mesh partitioning. In Alan George, John R. Gilbert, and Joseph W. H. Liu, editors, *Graph Theory and Sparse Matrix Computation*. Springer-Verlag, 1993.

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

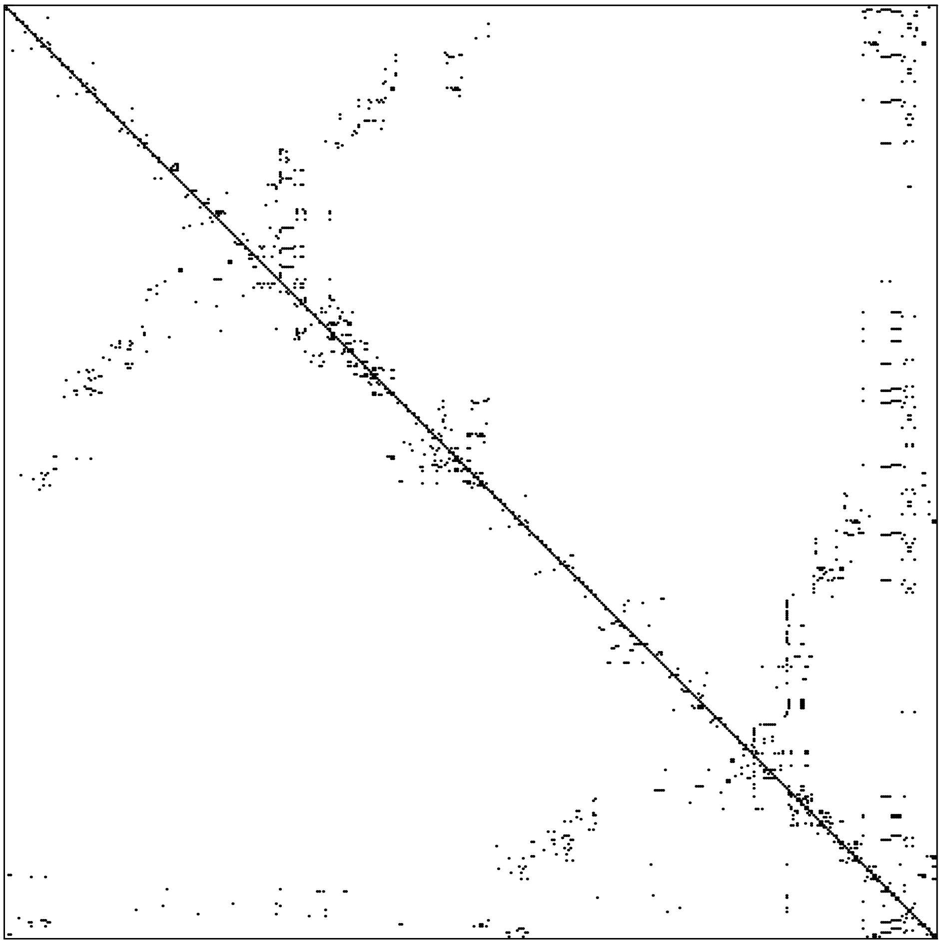
[27] Alex Pothen. *Sparse Null Bases and Marriage Theorems*. PhD thesis, Cornell University, 1984.

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

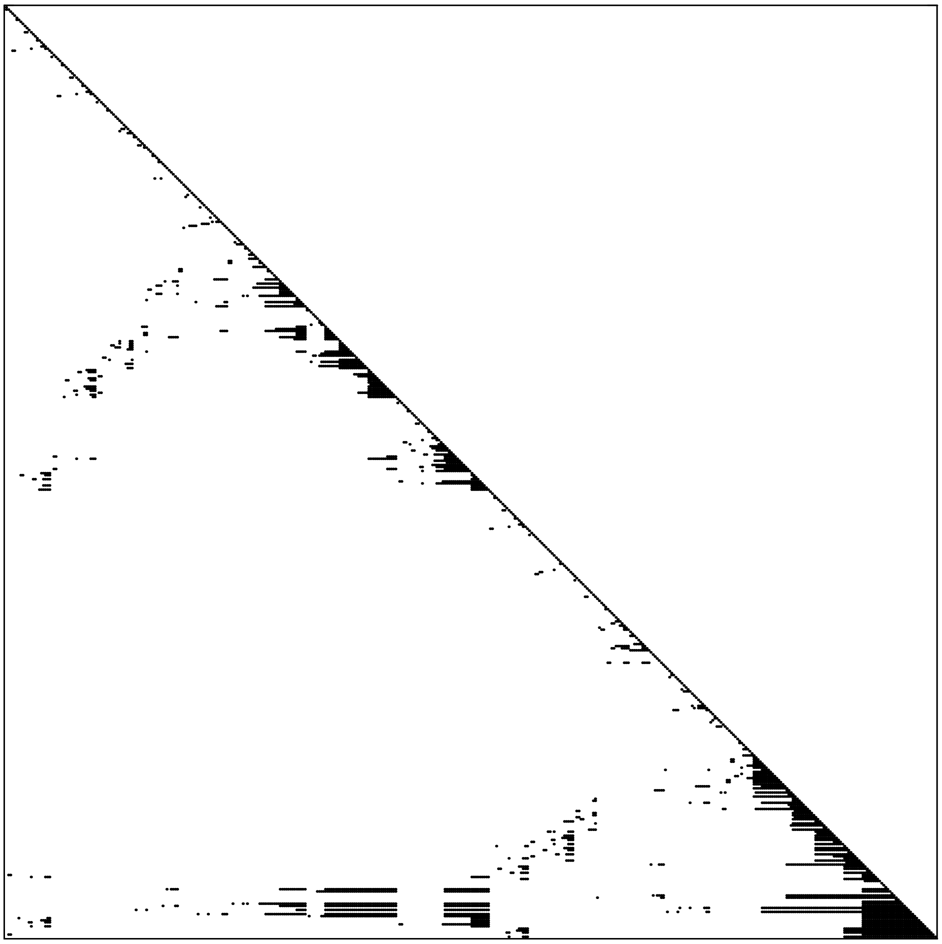
[29] Donald J. Rose, Robert Endre Tarjan, and George S. Lueker. Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing*, 5:266–283, 1976.

[30] Julio M. Stern and Stephen A. Vavasis. Nested dissection for sparse nullspace bases. Technical Report TR–90–1173, Cornell University, 1990.

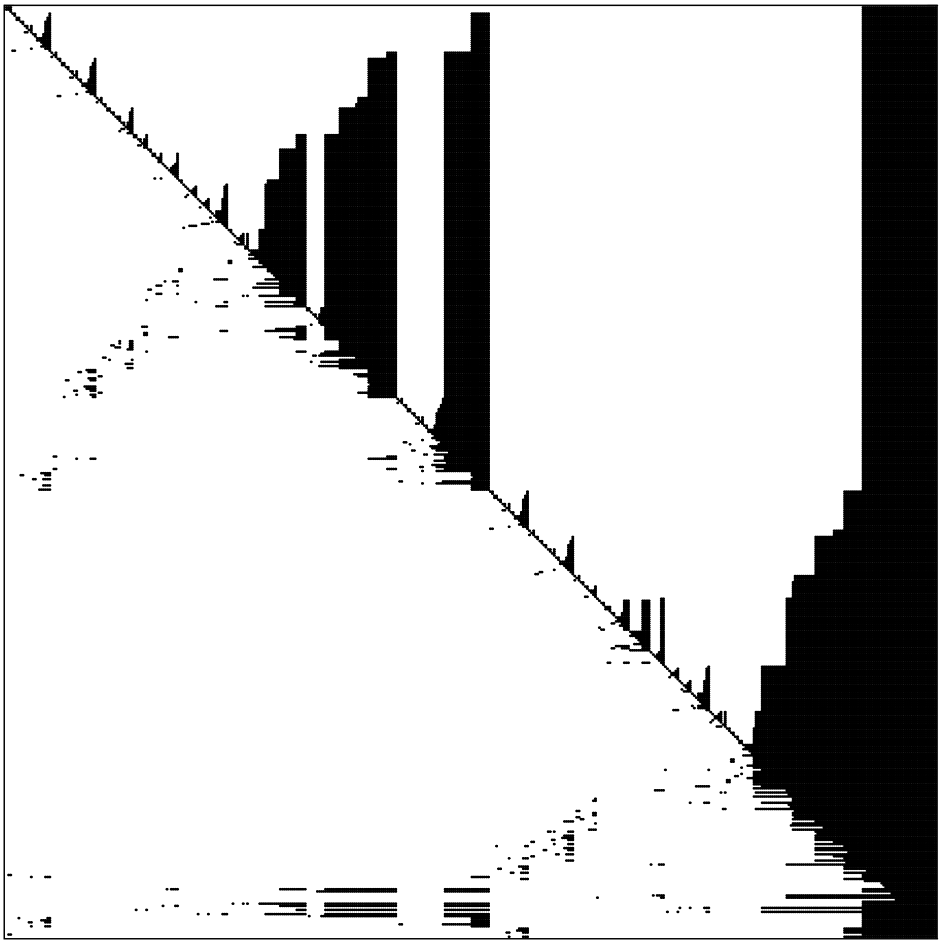
[31] Gilbert Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Box 157, Wellesley, MA 02181, 1986.



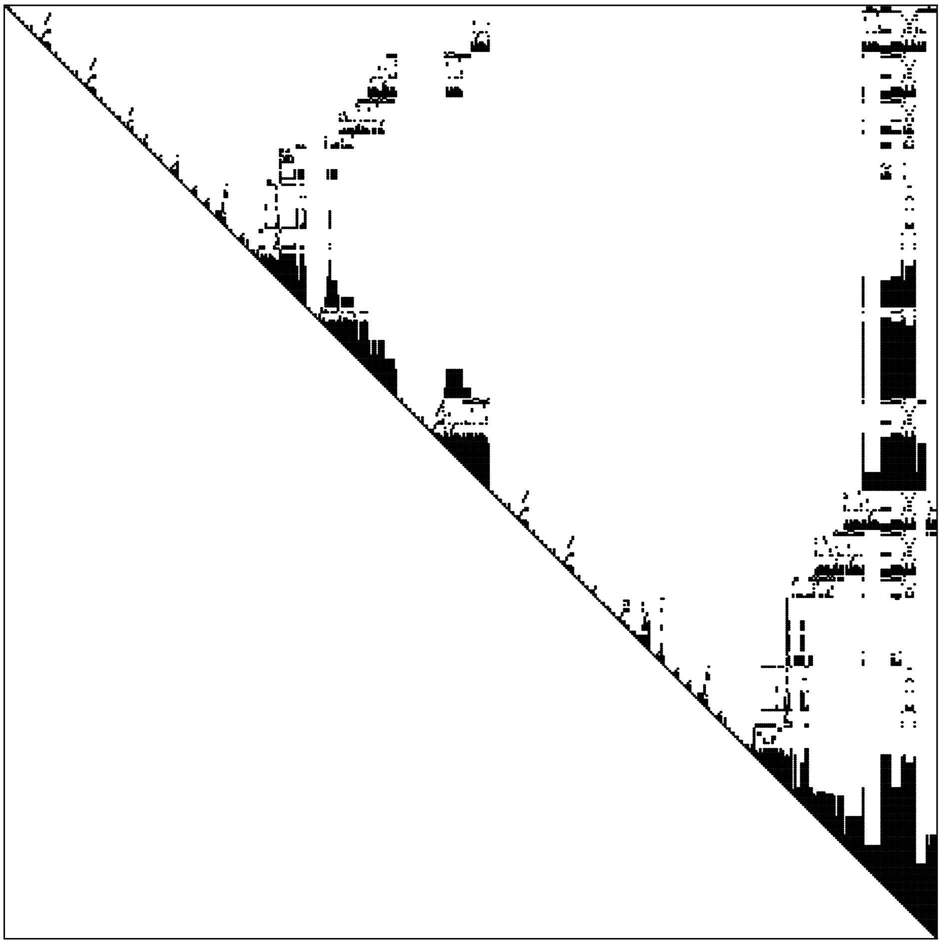
nz = 1966



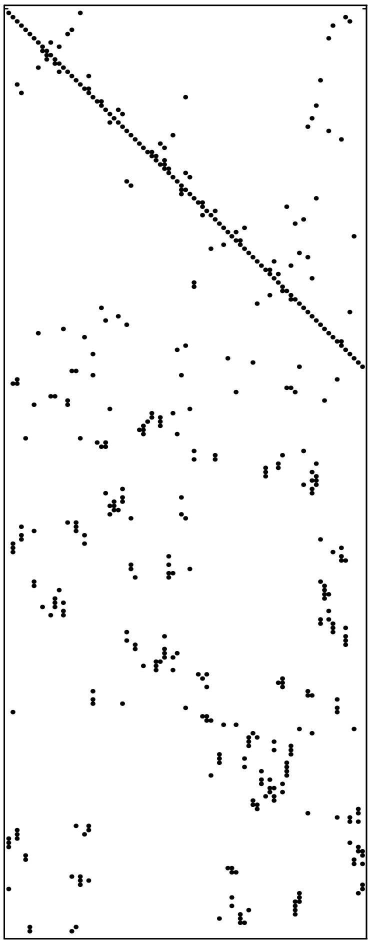
nz = 3579



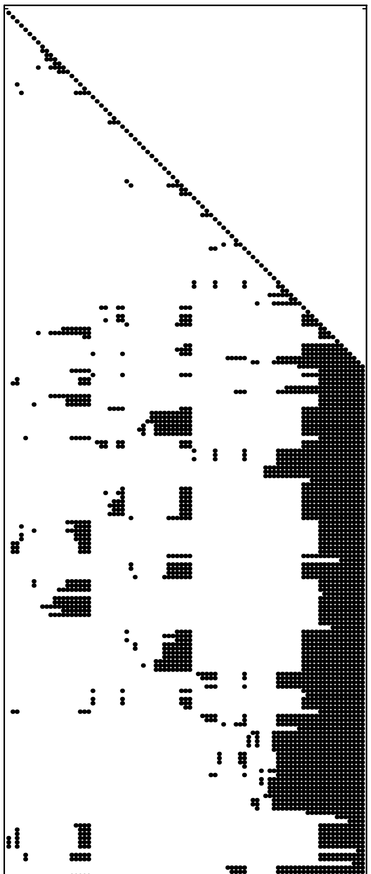
nz = 36381



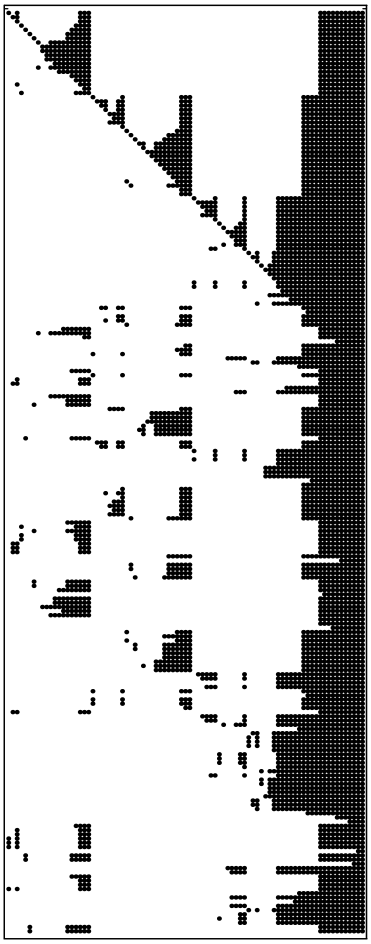
nz = 8500



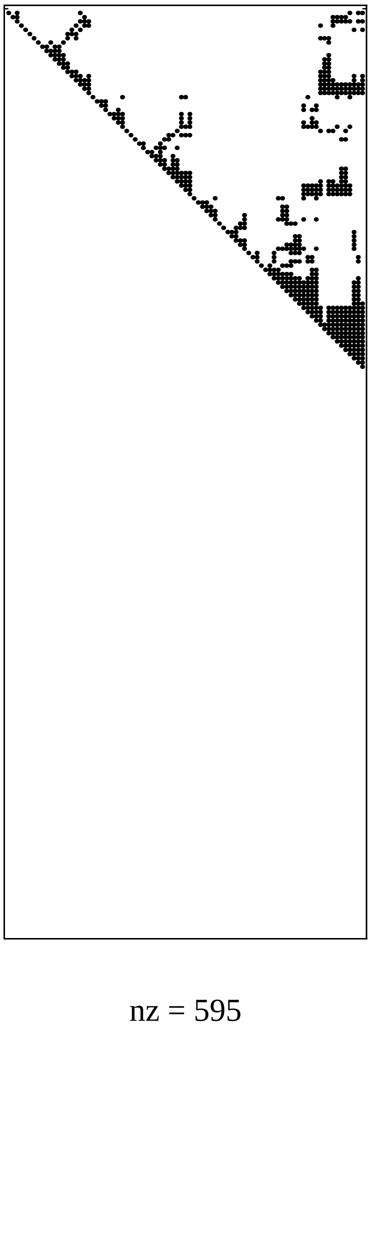
nz = 438



nz = 2954



nz = 4445



nz = 595