126-4114-02

4 February 1987

Dear Ms. DeAngelis,

Thank you for the help recently extended to me in the form of Research Report CS-81-09. Four of the ten dollars enclosed is intended as compensation for that report.

I often get the disquieting impression that a sole reliance upon the references found in loosely associated publications is not the optimal way to pursue a course of research. Given the fact that the University of Waterloo is widely known as a center of Numerical Analysis research, I am hoping that you can provide me with a list of the Department of Computer Science Research Reports which are currently available for distribution. Six of the ten dollars enclosed is, hopefully, adequate compensation for that listing of reports.

Thank you.

WITH RESPECT,

VBW

Vern B. Winterton
3821 South Terrace Heights Road
Salt Lake City, Utah
84109-3619
U.S.A.

received payment

S.D.

Sent 1986 listings

S.D.

# University of Waterloo

INVOICE

Waterloo, Ontario, Canada
N2L 3G1

Faculty of Mathematics
Department of Computer Science
519/885-1211

Telex Number
069-55259

January 22, 1987.

Mr. Vern B. Winterton,
3821 South Terrace Heights Road,
Salt Lake City, Utah,
84109-3619.
U.S.A.

Dear Mr. Winterton:

Thank you for your letter of January 3, 1987.

Enclosed please find our Research Report CS-81-09 by A. George and E. Ng. Please be advised that the cost of sending this report is $4.00 Canadian.

Would you please make your cheque or money order payable to the University of Waterloo, Computer Science Department and forward to my attention.

Thank you for your interest in our department.

Yours truly,

Susan DeAngelis

Susan DeAngelis (Mrs.),
Technical Report Secretary.

/sd
Encl.

3 January 1987

I am involved in a generalized course of research oriented around Matrix Analysis and Applications. I am very interested in studying the following report:

ON ROW AND COLUMN ORDERINGS FOR SPARSE LEAST SQUARES PROBLEMS
authored by: A. GEORGE & E. NG
Report CS-81-09

Would you please inform me of the cost of acquiring this report? In turn, I will mail to you a bank cheque in Canadian dollars.

Thank You. $4.00

WITH RESPECT,

Vern B. Winterton

Vern B. Winterton
3821 South Terrace Heights Road
Salt Lake City, Utah, 84109-3619
U.S.A.

# On Row and Column Orderings for Sparse Least Squares Problems*

*Alan George*

*Esmond Ng*

Department of Computer Science
University of Waterloo
Waterloo, Ontario, CANADA

Research Report CS-81-09
March, 1981

# On Row and Column Orderings for Sparse Least Squares Problems*

*Alan George*

*Esmond Ng*

Department of Computer Science
University of Waterloo
Waterloo, Ontario, CANADA

Research Report CS-81-09
March, 1981

## *ABSTRACT*

Let $P_1$ and $P_2$ be respectively $m$ by $m$ and $n$ by $n$ permutation matrices, with $m \geqslant n$. Suppose the $m$ by $n$ sparse matrix $\bar{A} = P_1AP_2$ is reduced to upper trapezoidal form $\begin{bmatrix} R \\ 0 \end{bmatrix}$ through the application of Givens rotations sequentially to the rows of $\bar{A}$. It is well-known that the sparsity of $R$ depends only on the choice of $P_2$, but the choice of $P_1$ can drastically affect the arithmetic required to compute $R$. In this paper we provide a mechanism for studying the connection between good row and good column orderings, along with a modified nested dissection algorithm for finding a good $P_2$ which automatically induces a good $P_1$. An analysis for a model problem is given, along with some experimental results.

* Research of the first author was supported in part by the Canadian Natural Sciences and Engineering Research Council under Grant A8111 and also supported in part by the U.S. Dept. of Energy under contract W-7405-eng-26 with the Union Carbide Corporation. Research of the second author was supported in part by the Canadian Natural Sciences and Engineering Research Council through a postgraduate scholarship.

# On Row and Column Orderings for Sparse Least Squares Problems*

*Alan George*
*Esmond Ng*
Department of Computer Science
University of Waterloo
Waterloo, Ontario, CANADA

Research Report CS-81-09
March, 1981

## 1. Introduction

Let $A$ be a large sparse $m$ by $n$ matrix which is of full rank. Consider the least squares problem

$$\min_{x} \| Ax - b \|_2 ,$$

where $b$ and $x$ are vectors of length $m$ and $n$ respectively. One way to solve this problem is to reduce $A$ to upper trapezoidal form using orthogonal transformations [6]. That is, we find an $m$ by $m$ orthogonal matrix $Q$ such that

$$QA = \begin{bmatrix} R \\ O \end{bmatrix} ,$$

where $R$ is $n$ by $n$ and upper triangular. If $c$ denotes the first $n$ components of $Qb$, the least squares solution $x$ is then obtained by solving the triangular system

$$Rx = c \; .$$

An efficient way of computing $Q$ and $R$ is as follows [4]. Let $R^0$ be the $n$ by $n$ zero matrix. Then for $1 \leqslant k \leqslant m$, we obtain $R^k$ by rotating the $k-th$ row of $A$ into $R^{k-1}$ using Givens transformations. Let $Q_k$ be the product of these Givens rotations. Then we have

$$Q = Q_m Q_{m-1} \cdots Q_2 Q_1 ,$$

and

* Research of the first author was supported in part by the Canadian Natural Sciences and Engineering Research Council under Grant A8111 and also supported in part by the U.S. Dept. of Energy under contract W-7405-eng-26 with the Union Carbide Corporation. Research of the second author was supported in part by the Canadian Natural Sciences and Engineering Research Council through a postgraduate scholarship.

$$R = R^m .$$

Note that

$$R^T R = \begin{bmatrix} R^T & O \end{bmatrix} \begin{bmatrix} R \\ O \end{bmatrix} = A^T Q^T Q A = A^T A ,$$

and since $A^T A$ is symmetric and positive definite, $R^T R$ is therefore (at least mathematically) the Cholesky decomposition of $A^T A$.

Let $P_1$ and $P_2$ be $m$ by $m$ and $n$ by $n$ permutation matrices respectively. We may then write

$$(P_2^T A^T P_1^T)(P_1 A P_2) = P_2^T A^T A P_2 = \overline{A}^T \overline{A}$$

which shows that row permutations of $A$ have no effect on the nonzero structure of $\overline{A}^T \overline{A}$. However, column permutations of $A$ correspond to symmetric row and column permutations of $A^T A$, and it is well-known that the choice of $P_2$ can drastically affect the sparsity of the Cholesky factor $\overline{R}^T$ of $\overline{A}^T \overline{A}$. Reliable algorithms are available for finding a $P_2$ which yields a sparse $\overline{R}$ [5], so if doing so is our sole objective, the ordering problem is already solved.

However, it is known that for a given column permutation $P_2$ of $A$, the cost of transforming $AP_2$ to upper trapezoidal form using Givens rotations depends very much on the row permutation $P_1$ [4]. Unfortunately, it is not obvious how to find a "good" row permutation cheaply.

In this paper, we give some results about the relationship between row and column permutations. A heuristic algorithm which is based on these results is proposed, and some experiments based on the algorithm are provided.

## 2. Basic Graph-theoretic terminology

An *undirected graph* $G=(X,E)$ consists of a finite set $X$ of *nodes* together with a set $E$ of *edges* which are unordered pairs of distinct nodes. A graph $\overline{G}=(\overline{X},\overline{E})$ is a *subgraph* of $G$ if $\overline{X}\subseteq X$ and $\overline{E}\subseteq E$. For any non-empty subset $Y$ of $X$, the *section subgraph* $G(Y)$ is the subgraph $(Y,E(Y))$ of $G$, where $E(Y)$ is the set $\{\{x,y\}\in E \mid x,y\in Y\}$.

Two nodes $x$ and $y$ in $G$ are said to be *adjacent* if $\{x,y\}\in E$. For $Y\subseteq X$, the *adjacent set* of $Y$ is defined as

$$Adj(Y)=\{x\in X-Y \mid \{x,y\}\in E \text{ for some } y\in Y\} .$$

if $Y=\{y\}$, we shall write $Adj(y)$ rather than $Adj(\{y\})$.

A set $C\subseteq X$ is a *clique* of $G$ if the nodes in $C$ are pairwise adjacent; i.e., if $x,y\in C$, then $\{x,y\}\in E$. The section subgraph $G(C)$ is called a *complete subgraph*.

For distinct nodes $x$ and $y$ in $G$, a *path* from $x$ to $y$ of length $l$ is an ordered set of distinct nodes $(v_0,v_1,\ldots,v_l)$ where $v_0=x$, $v_l=y$ and $\{v_{i-1},v_i\}\in E$, $1\leqslant i\leqslant l$. A graph $G$ is said to be *connected* if there is at least one path connecting every pair of distinct nodes in $G$. If $G$ is *disconnected*, it consists of two or more connected subgraphs called *components*. The *distance* $d(x,y)$ between any two nodes $x$ and $y$ in a connected graph $G$ is the length of the shortest path connecting them.

Let $T\subseteq X$ and $y\notin T$. The node $y$ is said to be *reachable from* a node $x$ *through* $T$ if there exists a path $(x,v_1,v_2,\ldots,v_k,y)$ from $x$ to $y$ such that $v_i\in T$ for $1\leqslant i\leqslant k$. The *reachable set* of $x$ through $T$, denoted by $Reach_G(x,T)$, is then defined to be the set of nodes $y\in X-T$ such that $y$ is reachable from $x$ through $T$. Note that the paths may be only of length one, and $T$ may be empty. The definition can also be extended to $Reach_G(Y,T)$, where $Y$ is any subset of $X$ [5].

A *partitioning* $\Phi$ of the node set $X$ of a graph $G$ is defined as $\Phi=\{X_1,X_2,\ldots,X_p\}$ where for $i\neq j$, $X_i\cap X_j=\emptyset$ and $\bigcup_{i=1}^{p}X_i=X$.

Let $G=(X,E)$ be a connected graph. A non-empty subset $C$ of $X$ is a *separator* of $G$ if the section graph $G(X-C)$ consists of *two or more* components. We denote the components by $G(C_1)$, $G(C_2)$, ..., $G(C_k)$, $k\geqslant 2$. The set $C$ is called a *width-l separator* of $G$ if for $x\in C_i$, $y\in C_j$, $i\neq j$, the distance $d(x,y)$ in $G$ is *greater* than $l$. Furthermore, if no proper subset of $C$ is a width-$l$ separator of $G$, then $C$ is called a *minimal width-l separator* of $G$. In this paper we are concerned with the case $l=2$. An example of a width-$l$ separator is given in Figure 2.1 for $l=2$.

Let $G_1=(X_1,E_1)$ and $G_2=(X_2,E_2)$ be two graphs. The *union* of $G_1$ and $G_2$, denoted by $G_1\bigcup G_2$, is the graph $(X_1\bigcup X_2,E_1\bigcup E_2)$. An example is given in Figure 2.2

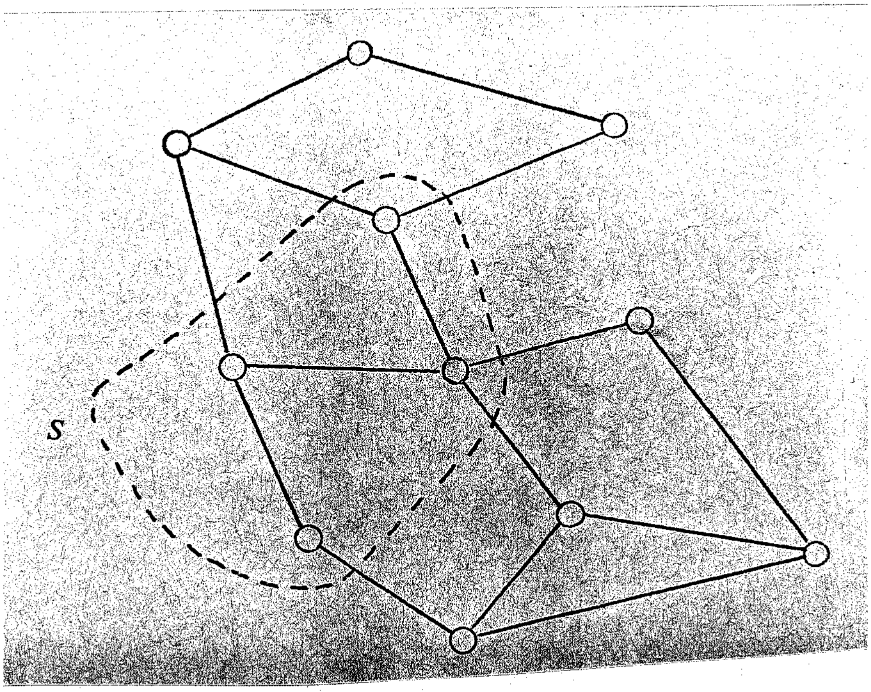


Figure 2.1 An example of a minimal width-2 separator ($S$).

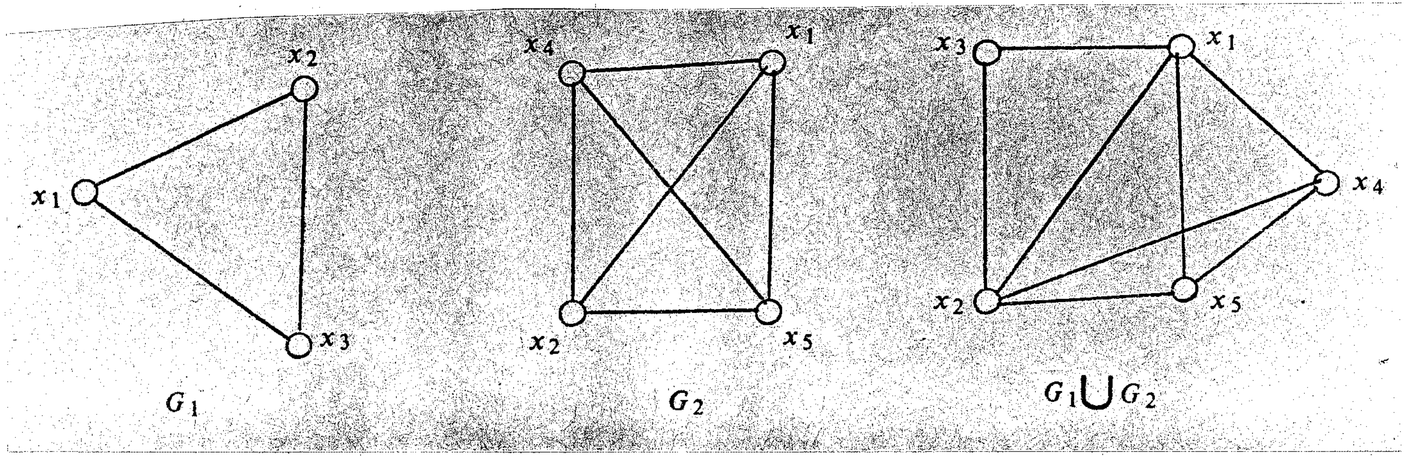


Figure 2.2 An example of the union of two graphs.

For a graph $G = (X,E)$ with $|X| = n$, an *ordering* (or *labelling*) of $G$ is a bijective mapping

$$\alpha : \{1, 2, \ldots, n\} \twoheadrightarrow X \ .$$

The node and edge sets of a labelled graph $G^{\alpha}$ are denoted by $X^{\alpha}$ and $E^{\alpha}$ respectively.

## 3. Row elimination using Givens rotations

Our first observation is well-known: when two sparse rows $x^T$ and $y^T$ are operated on by a rotation, so as to annihilate the leading element of $y^T$ say, the structure of the remaining parts of the transformed $\bar{x}^T$ and $\bar{y}^T$ is the union of those of $x^T$ and $y^T$. This is illustrated in Figure 3.1, where the rotation has been chosen to annihilate $y_1$.

$$\begin{bmatrix} c & s \\ s & -c \end{bmatrix} \left[\begin{array}{ccccccccccc} \times & 0 & 0 & 0 & \times & 0 & \times & 0 & 0 & 0 & \times \\ \times & 0 & \times & 0 & \times & 0 & 0 & \times & 0 & 0 & 0 \end{array}\right] \begin{array}{c} x^T \\ y^T \end{array}$$

$$= \left[\begin{array}{ccccccccccc} \times & 0 & \times & 0 & \times & 0 & \times & \times & 0 & 0 & \times \\ 0 & 0 & \times & 0 & \times & 0 & \times & \times & 0 & 0 & \times \end{array}\right] \begin{array}{c} \bar{x}^T \\ \bar{y}^T \end{array}$$

Figure 3.1 Example showing fill-in that occurs
when two rows are transformed by a rotation.

In the sequel, the row being used to annihilate an element, $x^T$ in Figure 3.1, will be called the *pivot row*. As a simple way of measuring the arithmetic cost in such operations, we count the number of nonzeros in the *transformed* pivot row, which is 6 in $\bar{x}^T$ in Figure 3.1.

Now consider using such rotations to reduce a sparse matrix $A$ to upper trapezoidal form $\begin{bmatrix} R \\ O \end{bmatrix}$. Recall from section 1 that our viewpoint is that the computation begins with an "empty" $R^0=O$, and the sequence of matrices $R^1$, $R^2$, ..., $R^{m-1}$, $R^m=R$ is computed, where $R^k$ is obtained from $R^{k-1}$ by rotating in the $k-th$ row of $A$. This process is illustrated in Figure 3.2.

We include a cost even for a row whose rotation into $R$ simply amounts to transferring it into $R$. We do so for simplicity, and because in most cases, time proportional to the number of nonzeros in the row will be expended, even if it is not done so in performing arithmetic. In any case, the error introduced in our cost is at most $O(min(|A|,|R|))$, and problems where such a term dominates the execution time bound are probably too small or special to be of much practical significance. Here and elsewhere, $|M|$ denotes the number of nonzeros in $M$ when $M$ is a vector or matrix, and the number of elements in $M$ when $M$ is a set.

We now consider the elimination of a row in more detail, using the example in Figure 3.3. We assume that the first $k$ rows of $A$ have been processed to generate $R^k$. The $k-th$ row of $A$ is denoted by $a^k$, and its elements are $a_i^k, 1 \leqslant i \leqslant n$. In Figure 3.3, nonzero elements of $R^{k-1}$ are denoted by $\times$, nonzeros introduced into $R^k$ and $a^k$ due to the elimination of $a^k$ are denoted by $+$, and all elements involved in the elimination of $a^k$ are circled. Of course elements in $a^k$ denoted by $\oplus$ are themselves ultimately annihilated.

$$A = \begin{bmatrix} \times & \times & 0 \\ \times & \times & 0 \\ 0 & 0 & \times \\ 0 & 0 & \times \\ \times & 0 & \times \end{bmatrix}$$

$$R^0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad R^1 = \begin{bmatrix} \times & \times & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad R^2 = \begin{bmatrix} \times & \times & 0 \\ 0 & \times & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$cost = 0 \qquad\qquad cost = 2 \qquad\qquad cost = 2$$

$$R^3 = \begin{bmatrix} \times & \times & 0 \\ 0 & \times & 0 \\ 0 & 0 & \times \end{bmatrix} \quad R^4 = \begin{bmatrix} \times & \times & 0 \\ 0 & \times & 0 \\ 0 & 0 & \times \end{bmatrix} \quad R^5 = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

$$cost = 1 \qquad\qquad cost = 1 \qquad\qquad cost = 6$$

Figure 3.2 A matrix $A$ and the structure of the sequence of matrices $R^0$, $R^1$, ..., $R^5$, along with cost of rotating in each row of $A$. Nonzeros are denoted by $\times$.

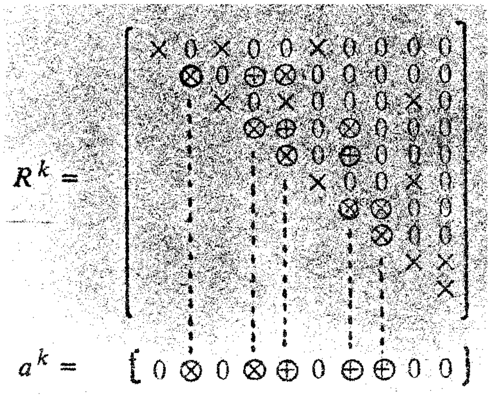


Figure 3.3 A sparse upper triangular $R^k$ where circled elements are involved in the elimination of $a^k$. Elements introduced into $R^k$ and $a^k$ due to the elimination of $a^k$ are denoted by $\oplus$.

We call the increasing sequence of row indices involved in the elimination of $a^k$ its *elimination sequence*, which we denote by $\Xi^k=\{\xi_1^k,\xi_2^k,\ldots,\xi_{\mu_k}^k\}$. In Figure 3.3, $\Xi^k=\{2,4,5,7,8\}$. Obviously, $\xi_1^k$ is the column subscript of the first nonzero in $a^k$, and $\xi_{i+1}^k$ is the column subscript of the first off-diagonal nonzero in row $\xi_i^k$ of $R^k$. The sequence terminates for one of two reasons: a) the pivot row has no off-diagonal nonzeros, as in the example of Figure 3.3, or b) an empty row is encountered at some point in the elimination, as occurred several times in the example of Figure 3.2. In case a), we say $\Xi^k$ is *maximal*.

Since $m\geqslant n$, and often $m\gg n$, there will usually be many maximal elimination sequences. Our objective is therefore to develop some conditions under which $\mu_k=|\,\Xi^k|$ can be limited.

The following lemma is an immediate consequence of the elimination process, assuming exact cancellation does not occur. We denote by $M_{ij}$ the $(i,j)$-element of $M$ when $M$ is a matrix.

Lemma 3.1

Let $s$ and $t$ be consecutive members of $\Xi^k$. Then

a) $R^k_{st}\neq 0$,

b) if $t-s>1$, then $R^k_{sj}=0$, $s<j<t$,

and c) if $R^k_{sj}\neq 0$, then $R^k_{tj}\neq 0$, for $j\geqslant t$.

□

It is useful to distinguish between two types of nonzeros in $R^k$. A nonzero $R^k_{ij}$ is a *non-fill element* if there is a row $a^l$, $l\leqslant k$, of $A$ such that $\min\{q\mid a^l_q\neq 0\}=i$ and $a^l_j\neq 0$. Obviously $R^k_{ij}\neq 0$ regardless of whether any elimination sequences occurred in computing $R^k$. If no such row of $A$ exists, then $R^k_{ij}$ is called a *fill element*, and its existence must be due to $i$ having appeared in some elimination sequence $\Xi^l=\{\xi^l_1,\xi^l_2,\ldots,\xi^l_{\mu_l}\}$, $l\leqslant k$, with $i>\xi^l_1$. Note that fill elements may turn into non-fill elements as $k$ increases.

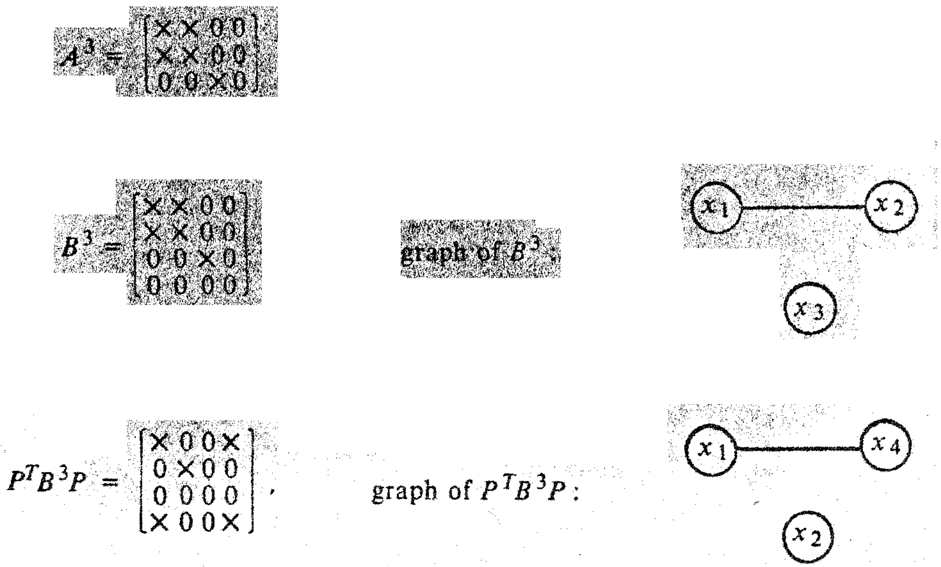
## 4. A graph-based interpretation

Our objective in this section is to obtain some information about the cost of eliminating the $k-th$ row of $A$, and in doing so, to gain some insight into obtaining good row and column orderings. Obviously the cost of eliminating row $k$ of $A$ depends only on the first $k$ rows of $A$, so it is helpful to define the matrix $A^k$ by

$$A^k = \begin{bmatrix} a^1 \\ a^2 \\ \cdot \\ \cdot \\ a^k \end{bmatrix},$$

where $a^j$ is the $j-th$ row of $A$. Let $B^k$ denote the $n$ by $n$ symmetric matrix $(A^k)^T A^k$.

The labelled graph of $B^k$, denoted by $G^k=(X^k,E^k)$, is an undirected graph having $|X^k| \leqslant n$ nodes, labelled as implied by $B^k$ (column ordering of $A^k$), with $\{x_i,x_j\} \in E^k$ if and only if $B_{ij}^k \neq 0$. Note that for *any* $n$ by $n$ permutation matrix $P \neq I$, the *unlabelled* graphs of $B^k$ and $P^T B^k P$ are identical; symmetrically permuting $B^k$ corresponds to a relabelling of the underlying graph. In this section we assume that the column ordering of $A$ (labelling of $G^k$, $1 \leqslant k \leqslant m$) is *fixed*, and we denote the node having label $i$ by $x_i$. An example illustrating these ideas is given in Figure 4.1.



$$A^3 = \begin{bmatrix} \times & \times & 0 & 0 \\ \times & \times & 0 & 0 \\ 0 & 0 & \times & 0 \end{bmatrix}$$

$$B^3 = \begin{bmatrix} \times & \times & 0 & 0 \\ \times & \times & 0 & 0 \\ 0 & 0 & \times & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$P^T B^3 P = \begin{bmatrix} \times & 0 & 0 & \times \\ 0 & \times & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \times & 0 & 0 & \times \end{bmatrix},$$

Figure 4.1 An example of $A^k$, $B^k$, $P^T B^k P$ and labelled graphs of $B^k$ and $P^T B^k P$. Here $k=3$.

Clearly the graph of $A^TA$, denoted by $G=(X,E)$, is given by $G^m=(X^m,E^m)$. When $A$ is of full rank, $A^TA$ is symmetric and positive definite, and $R^T$ is therefore the Cholesky factor of $A^TA$. The following result characterizes the nonzero structure of $R$ using the graph $G$ [3].

Lemma 4.1

For $j>i$, $R_{ij}\neq 0$ if and only if $x_j \in Reach_G(x_i, \{x_1,\ldots,x_{i-1}\})$.

□

The next lemma follows directly from Lemma 4.1 and its proof is omitted.

Lemma 4.2

Let $G$ be a disconnected graph having connected components $G(C_l)$, $l=1,2,\ldots,r$, and let $x_i$ and $x_j$, $j>i$, be in different components. Then $R_{ij}=0$.

□

A special case of Lemma 4.2 is when a component of $G$ consists of an isolated node, say $x_l$, corresponding to a row and column of $A^TA$ having all zeros except its diagonal element. From Lemma 4.2, $R_{lj}=0$, for $j>l$. In other words, the presence of $x_l$ in $G$ has no effect on the application of Lemma 4.1 or the determination of the nonzero structure of $R$.

Consider the sequence of symmetric matrices $B^k=(A^k)^TA^k$. Since some of the $A^k$ may have null columns, $B^k$ may be only (structurally) *positive semi-definite*, which is manifested in $R^k$ as null rows corresponding to null rows of $B^k$. However, our discussion in the previous paragraph shows that except for the null rows of $R^k$, it has the same structure as it would have if all diagonal elements of $R^k$ had been nonzero.

Thus, we can determine the nonzero structure of $R^k$ from $G^k$. Nodes corresponding to null rows in $B^k$ are deleted. Lemma 4.1 applies as before, but involves nodes actually present in $G^k$.

It is important to note that all our discussions, and Lemma 4.1, assume that we are working with $B^k$ and that *no* cancellation occurs during the computation. Thus the structure of $R^k$ determined represents the "worst case" situation with respect to fill-in. Moreover, it is assumed that *all* elimination sequences are maximal.

We now define the graph $G^k$ in a more refined way. Define the $n$ by $n$, rank-one matrix

$$Y_j=(a^j)^Ta^j,$$

and denote its corresponding symmetric graph by $G_j=(X_j,E_j)$. Note that $G_j$ is a complete graph.

Using the fact that

$$B^k = (A^k)^T A^k = \sum_{j=1}^{k} Y_j ,$$

we can express the graph $G^k$ as the union of $k$ complete graphs:

$$G^k = \bigcup_{j=1}^{k} G_j = (\bigcup_{j=1}^{k} X_j, \bigcup_{j=1}^{k} E_j) .$$

For future reference, we state the following observation as a lemma.

Lemma 4.3

$$G^k = G^{k-1} \bigcup G_k = (X^{k-1} \bigcup X_k, E^{k-1} \bigcup E_k) .$$

□

Now for each $G^k$, let $\Omega^k = \{C_1^k, C_2^k, \ldots, C_{\gamma_k}^k\}$ be its *component partitioning*; that is, the partitioning of its node set induced by the connected components $G^k(C_j^k)$ of the graph $G^k$. Thus, for $1 \leqslant i,j \leqslant \gamma_k$ and $i \neq j$, $C_i^k \bigcap C_j^k = \varnothing$ and $\bigcup_{l=1}^{\gamma_k} C_l^k = X^k$. The following lemma is obvious from the way in which the graphs $G^k$ are defined.

Lemma 4.4

For $1 \leqslant k_1 < k_2 \leqslant m$, if $C_i^{k_1} \in \Omega^{k_1}$, then there exists $j$ such that $C_i^{k_1} \subseteq C_j^{k_2}$ and $C_j^{k_2} \in \Omega^{k_2}$.

□

That is, the size of the component sets $C_i^k$ is non-decreasing as $k$ increases. An example is given in Figure 4.2, where the graphs correspond to those of the matrix $A$ given in Figure 3.2.

It should be emphasized that the sequence of component partitionings $\Omega^k$ depends *only* on the ordering of the rows of $A$. It is independent of the column ordering. The effect of permuting the columns of $A$ is just a relabelling of the nodes in $G^k$.

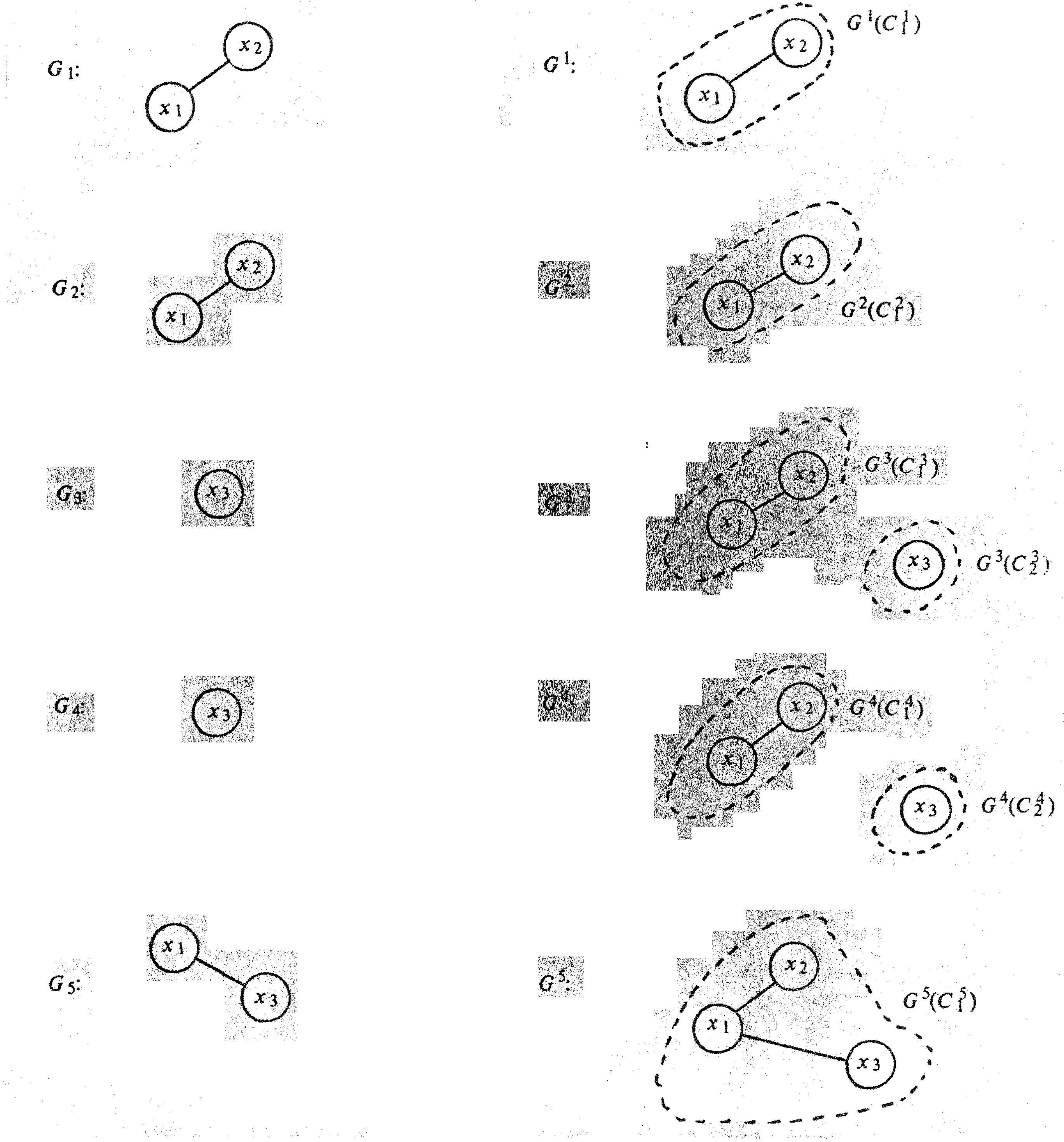


Figure 4.2 The graphs of the rows of the matrix $A$ in Figure 3.2, and the component partitionings of their unions.

Lemma 4.5 and Theorem 4.6 characterize the elimination sequences introduced in section 3 in terms of the graphs $G^k$.

Lemma 4.5

a) $\xi_1^k = \min\{i \mid x_i \in X_k\}$.

b) For $2 \leqslant j \leqslant \mu_k$, $\xi_j^k = \min\{i \mid x_i \in Reach_{G^k}(x_{\xi_{j-1}^k}, T_{\xi_{j-1}^k})\}$, where $T_{\xi_{j-1}^k} = \{x_i \in X^k \mid i < \xi_{j-1}^k\}$.

Proof:

This follows from the definition of $\Xi^k$ and Lemma 4.1.

□

Theorem 4.6

Let $\Omega^k = \{C_1^k, C_2^k, \ldots, C_{\gamma_k}^k\}$ be the component partitioning of $G^k$ induced by the row ordering of $A$. Denote the component containing $G_k$ by $G^k(C_{\sigma_k}^k)$. Then $x_p \in C_{\sigma_k}^k$, for all $p \in \Xi^k$.

Proof:

The proof is by contradiction. Since $x_{\xi_1^k} \in C_{\sigma_k}^k$, there must exist two consecutive members $s$ and $t$ of $\Xi^k$ such that $x_s \in C_{\sigma_k}^k$ and $x_t \in C_l^k$, for some $l \neq \sigma_k$. From Lemma 3.1, $R_{st}^k \neq 0$. Furthermore, $R_{st}^k$ must be a fill element, since otherwise there would exist a row $a^j$, $j \leqslant k$, such that $a_s^j$ is the first nonzero in $a^j$ and $a_t^j \neq 0$. Using the graph interpretation, this implies that there would exist an $X_j$, $j \leqslant k$, such that $x_s \in X_j \bigcap C_{\sigma_k}^k$ and $x_t \in X_j \bigcap C_l^k$, which contradicts the definition of $\Omega^k$.

Thus a necessary *prerequisite* for $\Xi^k$ to violate the theorem is that there exists a row having such a fill element.

We now want to show that no such row exists. Suppose for a contradiction that some do, and let $(r,s)$ be a subsequence of the elimination sequence $\Xi^l$ that *creates* the first such element $R_{st}^l$, where $x_r, x_s \in C_{\sigma_l}^l$ and $x_t \in C_j^l$, $j \neq \sigma_l$ and $l \leqslant k$. Now in order for $R_{st}^l$ to be created, there must exist elements $R_{rs}^l \neq 0$ and $R_{rt}^l \neq 0$. However $R_{rt}^l$ cannot be a fill element because row $s$ is the first row having a fill element $R_{st}^l$ so that $x_s \in C_{\sigma_l}^l$ and $x_t \in C_j^l$, $j \neq \sigma_l$. Thus $R_{rt}^l$ must be a non-fill element, which implies that there is an $X_q$, $q \leqslant l$, such that $x_r \in X_q \bigcap C_{\sigma_l}^l$ and $x_t \in X_q \bigcap C_j^l$, which again contradicts the definition of $\Omega^l$.

□

Theorem 4.6 is important because it illustrates the significance of the component partitioning $\Omega^k$. It shows that the set of nodes that are involved in the elimination of row $k$, $\{x_{\xi_1^k}, x_{\xi_2^k}, \ldots, x_{\xi_{\mu_k}^k}\}$, is limited to the component $G^k(C^k_{\sigma_k})$ whose node set $C^k_{\sigma_k}$ contains $X_k$. Note that the cost of eliminating a row depends in part on the *length* of its elimination sequence. Obviously, we want to find a row and column orderings which allow the component $G^k(C^k_{\sigma_k})$ to be kept small for as large a $k$ as possible.

The following results are consequences of Theorem 4.6.

Theorem 4.7

The cost of eliminating row $k$ is bounded by

$$½|C^k_{\sigma_k}|(|C^k_{\sigma_k}|+1).$$

Proof:

The bound is obtained by assuming $\Xi^k$ is maximal and each row in the elimination sequence has nonzeros in all positions in $C^k_{\sigma_k}$ which are to the left of the diagonal in $R$.

□

Corollary 4.8

Let $\delta_k = |\{x_l \in C^k_{\sigma_k} \mid l \geqslant \xi_1^k\}|$. Then the cost of eliminating row $k$ is bounded by $½\delta_k(\delta_k+1)$.

□

Theorem 4.7 says we want to keep $C^k_{\sigma_k}$ small and Corollary 4.8 says that regardless of whether $C^k_{\sigma_k}$ is small, we want to arrange that the leading column subscript of row $k$ be as large as possible.

## 5. Automatic width-2 nested dissection ordering

Let $G=(X,E)$ be the unlabelled graph of $B=A^TA$. For simplicity, we assume $G$ is connected. Let $S$ be a width-2 separator in $G$, whose removal disconnects the graph into two or more components (say 2). Denote the node sets of the components by $C_1$ and $C_2$. Lemma 5.1 follows directly from the definition of $S$.

Lemma 5.1

Let $K$ be any clique in $G$. Then either $K \subseteq C_1 \bigcup S$ or $K \subseteq C_2 \bigcup S$.

□

We recall from section 4 that $G$ can be written as the union of the graphs $G_k=(X_k,E_k)$, where each is a complete subgraph of $G$. The following lemmas characterize the cliques $X_k$ in the partitioning $\{S,C_1,C_2\}$ of $X$.

Lemma 5.2

Let $X_i \subseteq C_1 \bigcup S$ and $X_j \subseteq C_2 \bigcup S$, $i \neq j$. If $X_i \bigcap C_1 \neq \emptyset$ and $X_j \bigcap C_2 \neq \emptyset$, then $X_i \bigcap X_j = \emptyset$.

Proof:

If $X_i \bigcap S = \emptyset$ or $X_j \bigcap S = \emptyset$, then the result follows immediately from Lemma 5.1. Assume $X_i \bigcap S \neq \emptyset$ and $X_j \bigcap S \neq \emptyset$. Let $x_i \in X_i \bigcap C_1$ and $x_j \in X_j \bigcap C_2$. If $X_i \bigcap X_j \neq \emptyset$, there would exist $y \in X_i \bigcap X_j \subseteq S$ such that $x_i, x_j \in Adj(y)$, which contradicts the definition of $S$.

□

Lemma 5.3

If $S$ is a minimal width-2 separator, then there is at least one complete graph $G_k$ in the section graph $G(S)$.

Proof:

Since $S$ is a minimal width-2 separator, there must exist $u \in C_1$, $v \in C_2$, and $x,y \in S$ $(x \neq y)$ such that $(u,x,y,v)$ is a path in $G$. Otherwise either $S-\{x\}$ or $S-\{y\}$ is also a width-2 separator which contradicts the fact that $S$ is minimal. Now $G$ is the union of the complete graphs $G_i$, so there must exist one, say $G_k$, such that $x,y \in X_k$ and $\{x,y\} \in E_k$. By Lemma 5.1, either $X_k \subseteq C_1 \bigcup S$ or $X_k \subseteq C_2 \bigcup S$. Assume the first alternative. If $X_k \bigcap C_1 \neq \emptyset$, then there exists $w \in X_k \bigcap C_1$ and a path $(w,y,v)$ in $G$, contradicting the fact that $S$ is a width-2 separator. Thus $X_k \subseteq S$, and $G_k$ is a subgraph of $G(S)$.

□

Lemma 5.4

If $S$ is a minimal width-2 separator, then every edge in $G(S)$ belongs to at least one complete subgraph $G_k$ of $G(S)$.

Proof:

This follows from the fact that $G$ is a union of the complete graphs $G_i$ and $S$ is a minimal width-2 separator.

□

Corollary 5.5

If $S$ is a minimal width-2 separator, then $G(S)$ is the union of one or more complete graphs $G_k$.

Proof:

This follows from Lemmas 5.3 and 5.4.

□

Lemma 5.2 and Corollary 5.5 are important because they provide some insight into how the columns and rows of $A$ should be ordered. We now assume that $S$ is minimal. If the nodes of $S$ are numbered *after* those of $C_1$ and $C_2$, it follows directly from Lemma 4.1 that $\{x,y\}$ is not a fill edge for $x \in C_1$ and $y \in C_2$. Thus this *dissection* technique induces a labelling $\alpha$ for $G$ and hence a column permutation $P_2$ for $A$) such that the upper triangular matrix $R$ suffers *low* fill-in.

Apart from providing a good column permutation for $A$, this dissection techique also *induces* a good row ordering for $A$ in a natural way. Note that if $x \in C_1$, $y \in C_2$ and $z \in S$, then $\alpha^{-1}(x) < \alpha^{-1}(y) < \alpha^{-1}(z)$. Denote the nodes associated with the rows of the matrix $AP_2$ by $X_1^\alpha$, $X_2^\alpha$, ..., $X_m^\alpha$, and their correspnding graphs by $G_1^\alpha$, $G_2^\alpha$, ..., $G_m^\alpha$.

The results in section 4 imply that for any column permutation, say $P_2$, the cost of transforming the matrix $AP_2$ to upper trapezoidal form depends on the order in which the rows are processed. Let $\Phi_S$ be the set of rows of $AP_2$ such that their associated nodes are in $S$; and $\Phi_{C_i}$, $i = 1, 2$, be the set of rows of $AP_2$ such that their associated nodes are in $C_i \bigcup Adj(C_i)$, where $Adj(C_i) \subset S$.

Clearly $\Phi_S \bigcup \Phi_{C_1} \bigcup \Phi_{C_2}$ is the set of all rows in $AP_2$, due to Lemma 5.2 and Corollary 5.5. Since $Adj(S) \subseteq C_1 \bigcup C_2$, if the rows in $\Phi_S$ are processed first, then when the rows in $\Phi_{C_1}$ and $\Phi_{C_2}$ are processed, the component $G^k(C^k_{\sigma_k})$ in $G^k = \bigcup_{i=1}^{k} G_i^\alpha$, $k > |\Phi_S|$, would become as large as $G^k$. In view of the results in section 4, this is undesirable. Thus one would like to process the rows in $\Phi_S$ as late as possible. Assuming we do that, it follows from Lemma 5.2 and Corollary 5.5 that for $i = 1, 2$, the component $G^k(C^k_{\sigma_k})$ is at most $C_i \bigcup Adj(C_i)$ when the rows in $\Phi_{C_i}$ are processed. This row ordering can be obtained by arranging the rows in $AP_2$ so that the leading column subscripts in the rows of $AP_2$ are in non-decreasing order.

This dissection technique can of course be applied recursively, yielding a *width-2 nested dissection*, which is similar to the nested dissection described in [2]. However, care has to be taken when choosing width-2 separators. Consider the graph $G = (X, E)$ in Figure 5.1.

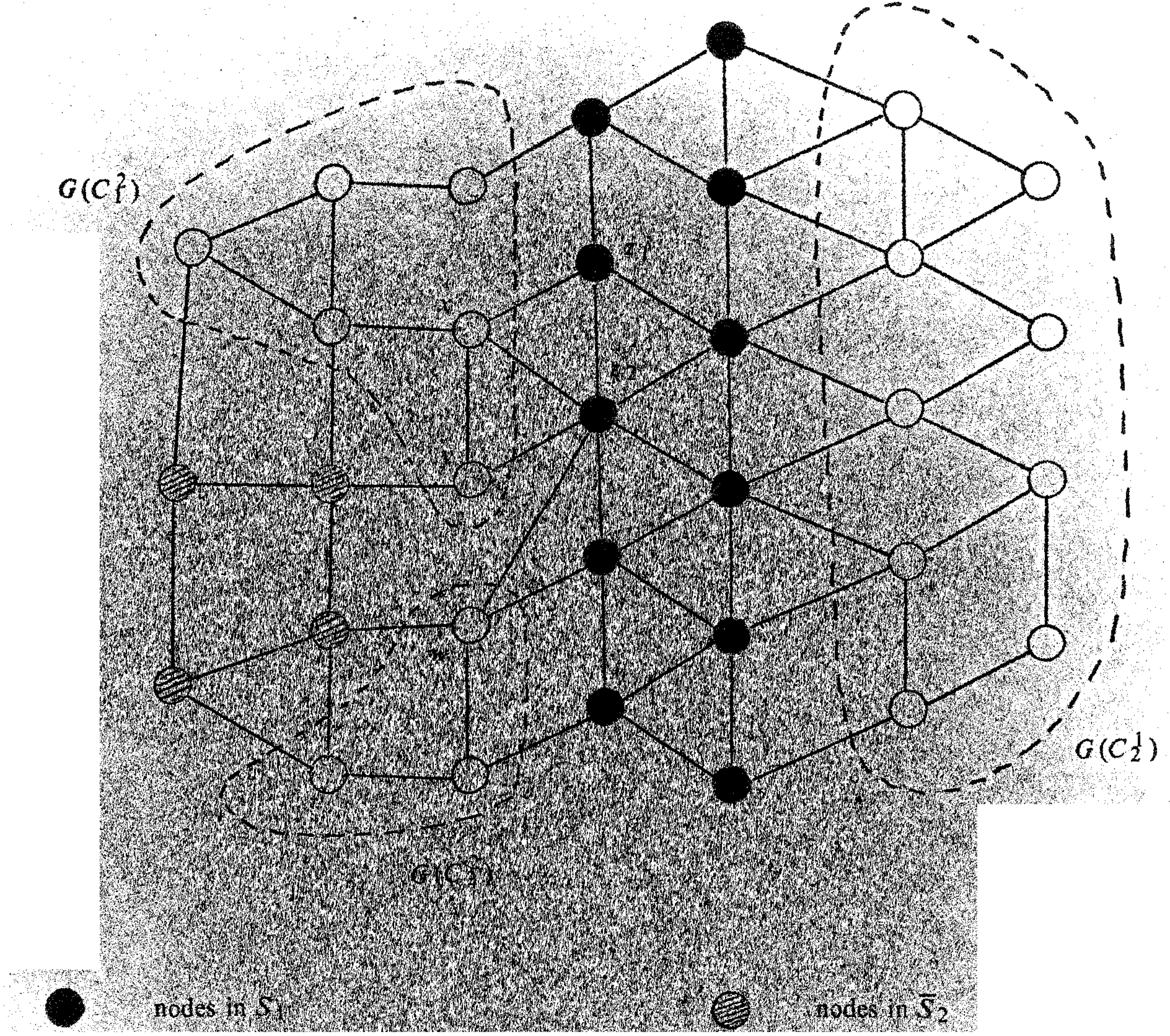


Figure 5.1 An example illustrating the choice of width-2 separators.

Let $S_1$ be the set of nodes that are darkened. Clearly $S_1$ is a minimal width-2 separator in $G$. Denote the components of $G(X-S)$ by $G(C_1^1)$ and $G(C_2^1)$. Now we apply the dissection technique again to the graph $G(C_1^1)$ and finds a minimal width-2 separator $\bar{S}_2$ for it ($\bar{S}_2$ contains nodes that are shaded). Denote the components in $G(C_1^1-\bar{S}_2)$ by $G(C_1^2)$ and $G(C_2^2)$. It is clear that even though $\bar{S}_2$ is a minimal width-2 separator in $G(C_1^1)$, $S_1 \bigcup \bar{S}_2$ is *not* a minimal width-2 separator in $G$ with respect to the components $G(C_1^2)$, $G(C_2^2)$ and $G(C_2^1)$, since there is a path of length 2 from $x$ to $w$. The problem is due to the fact that $Adj(C_1^1) \subset S_1$. In this case, we cannot guarantee that there would not exist $u \in C_1^2$ and $v \in C_2^2$ such that $S_1 \bigcap Adj(u) \bigcap Adj(v) \neq \emptyset$.

To solve this problem, we do the following. Instead of choosing a width-2 separator from $G(C_1^1)$, it is chosen from the graph $G(C_1^1 \bigcup Adj(C_1^1))$. As an example, $S_2 = \overline{S}_2 \bigcup \{x,y,z_1,z_2\}$ is a minimal width-2 separator in $G(C_1^1 \bigcup Adj(C_1^1))$ in Figure 5.1, and $S_1 \bigcup S_2 = S_1 \bigcup \overline{S}_2 \bigcup \{x,y\}$ is a minimal width-2 separator in $G$. The following theorem shows that the separator constructed in this way is always a minimal width-2 separator in $G$.

Theorem 5.6

Let $S_1^1$ be a minimal width-2 separator in a connected graph $G=(X,E)$ and denote the components in $G(X-S_1^1)$ by $G(C_1^1)$, $G(C_2^1)$, ..., $G(C_p^1)$. Let $S_k^2$ be a minimal width-2 separator in $G(C_k^1 \bigcup Adj(C_k^1))$ and denote the components in $G(C_k^1 - S_k^2)$ by $G(C_1^2)$, $G(C_2^2)$, ..., $G(C_q^2)$. If $x \in C_i^2$ and $y \in C_j^2$, $i \neq j$, then the distance between $x$ and $y$ in $G$ is greater that 2. That is, $S_1^1 \bigcup S_k^2$ is a width-2 separator in $G$. Furthermore, $S_1^1 \bigcup S_k^2$ is minimal.

Proof:

If $x \in C_i^2$ and $y \in C_j^2$, $i \neq j$, then $Adj(x) \subseteq C_i^2 \bigcup Adj(C_i^2)$ and $Adj(y) \subseteq C_j^2 \bigcup Adj(C_j^2)$. Note that $G(C_i^2 \bigcup Adj(C_i^2) - S_k^2)$ and $G(C_j^2 \bigcup Adj(C_j^2) - S_k^2)$ are components in $G(C_k^1 \bigcup Adj(C_k^1))$, and $S_k^2$ is a minimal width-2 separator in $G(C_k^1 \bigcup Adj(C_k^1))$. Thus, $C_i^2 \bigcup Adj(C_i^2)$ and $C_j^2 \bigcup Adj(C_j^2)$ must be disjoint, implying that $Adj(x) \bigcap Adj(y) = \emptyset$. Now $S_1^1 \bigcup S_k^2$ is minimal because $S_1^1$ and $S_k^2$ are minimal, and $G$ is a union of complete graphs.

□

We now define a *width-2 nested dissection partitioning* formally. Let $G=(X,E)$ be the unlabelled graph of $A^TA$. Let $Y^0 = X$, and for $m=0,1,2,\ldots,h$ until $Y^{h+1} = \emptyset$, do the following:

a) Determine the connected components of $Y^m$ and label them $Y_1^m$, $Y_2^m$, ..., $Y_{r_m}^m$.

b) For $j=1,2,\ldots,r_m$, choose $\overline{S}_j^m \subseteq Y_j^m \bigcup Adj(Y_j^m)$ such that $\overline{S}_j^m$ is a minimal width-2 separator of $G(Y_j^m \bigcup Adj(Y_j^m))$ and set $S_j^m = \overline{S}_j^m \bigcap Y_j^m$, or else is equal to $Y_j^m$.

c) Define $S^m = \bigcup_{j=1}^{r_m} S_j^m$ and $Y^{m+1} = Y^m - S^m$.

The partitioning $\Phi = \{S_j^m \subseteq X, 1 \leqslant j \leqslant r_m, 0 \leqslant m \leqslant h\}$ is a width-2 nested dissection partitioning of $G$.

An ordering $\alpha$ of $X$ is said to be a *width-2 nested dissection ordering* with respect to $\Phi = \{S_j^m\}$ if for $x \in S_j^m$ and $y \in Y_j^m - S_j^m$, $\alpha^{-1}(x) > \alpha^{-1}(y)$. An example of a width-2 nested dissection ordering on the partitioning of Figure 5.2 is shown in Figure 5.3.

The proof of the following theorem is similar to that of George and Liu [2]. It provides a bound on the number of nonzeros in the upper triangular martix $R$.

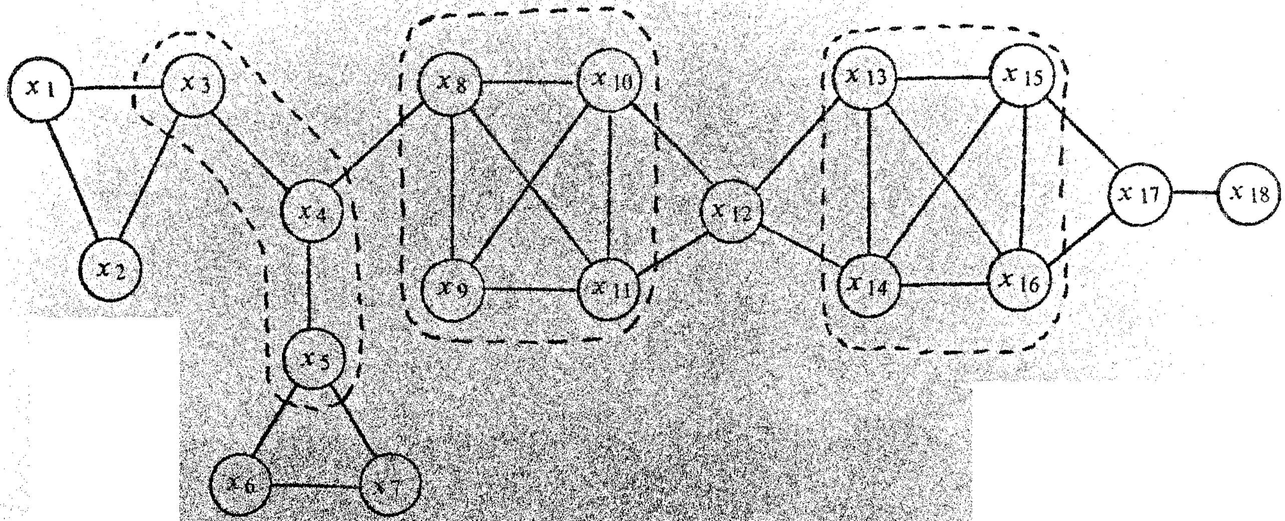
Theorem 5.7

Let $\Phi=\{S_j^m\}$ be a width-2 nested dissection partitioning on $G$ and $\alpha$ be any width-2 nested dissection ordering with respect to $\Phi$. Then the number of nonzeros in $R$ is bounded by

$$\sum_{m=0}^{h} \sum_{j=1}^{r_m} | S_j^m | \{| Adj(Y_j^m)| +(| S_j^m | -1)/2\} .$$

□

The bound given in Theorem 5.7 can be used as a guideline in determining width-2 nested dissection orderings with small fill. Clearly, small minimal width-2 separators which disconnect the graph into two or more components of approximately equal size should be used.



$R^0 = X$

$R_1^1 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$

$R_2^1 = \{x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}\}$

$R_1^2 = \{x_1, x_2\}$

$R_2^2 = \{x_6, x_7\}$

$R_3^2 = \{x_{12}\}$

$R_4^2 = \{x_{17}, x_{18}\}$

$S^0 = \{x_8, x_9, x_{10}, x_{11}\}$

$S_1^1 = \{x_3, x_4, x_5\}$

$S_2^1 = \{x_{13}, x_{14}, x_{15}, x_{16}\}$

$S_j^2 = R_j^2,\ j = 1, 2, 3, 4$

Figure 5.2 A width-2 nested dissection partitioning $\{S_j^m\}$ of a graph of $A^TA$.

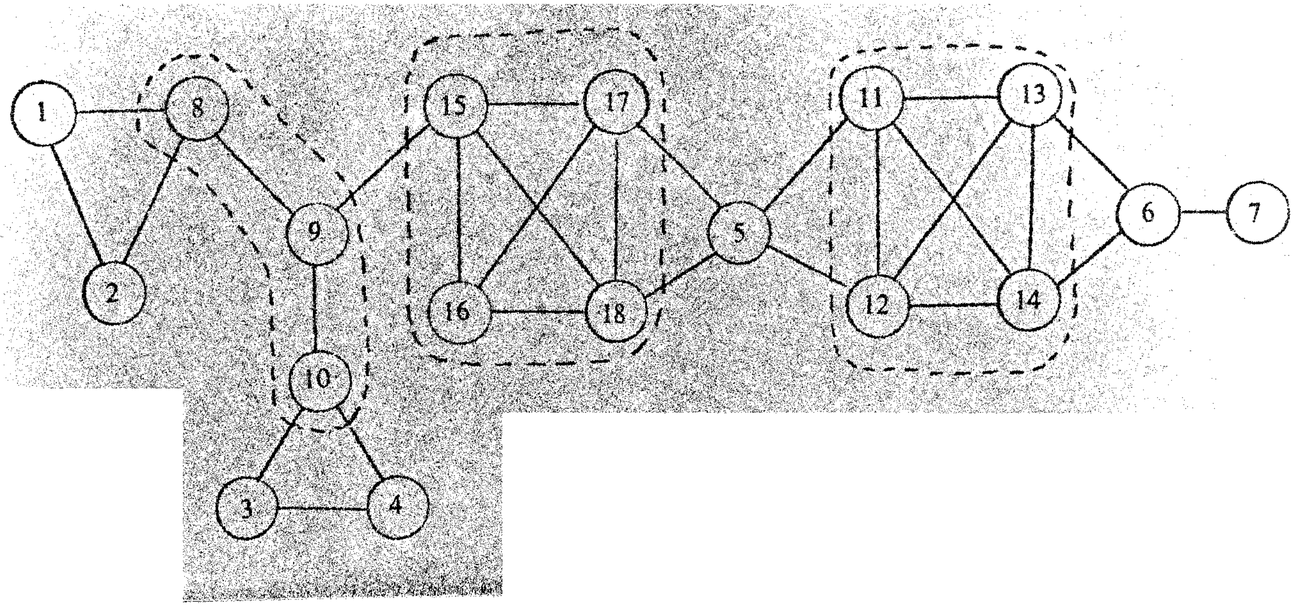


Figure 5.3 A width-2 nested dissection ordering on the width-2 nested dissection partitioning of Figure 5.2.

## 6. Analysis of a model problem

In this section, we analyze a model problem which is typical of those which arise in the natural factor formulation of finite element methods [1]. Consider an $n$ by $n$ grid which consists of $(n-1)^2$ small squares. An example is given in Figure 6.1 with $n=3$.

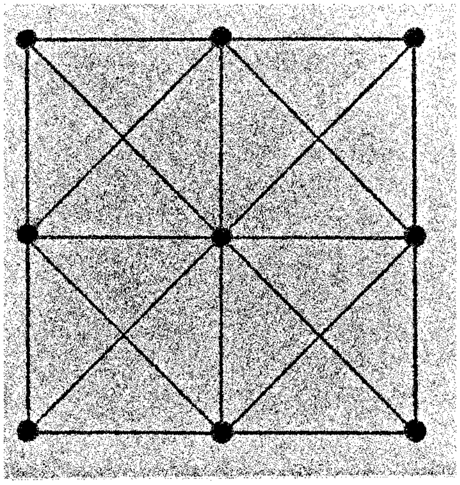
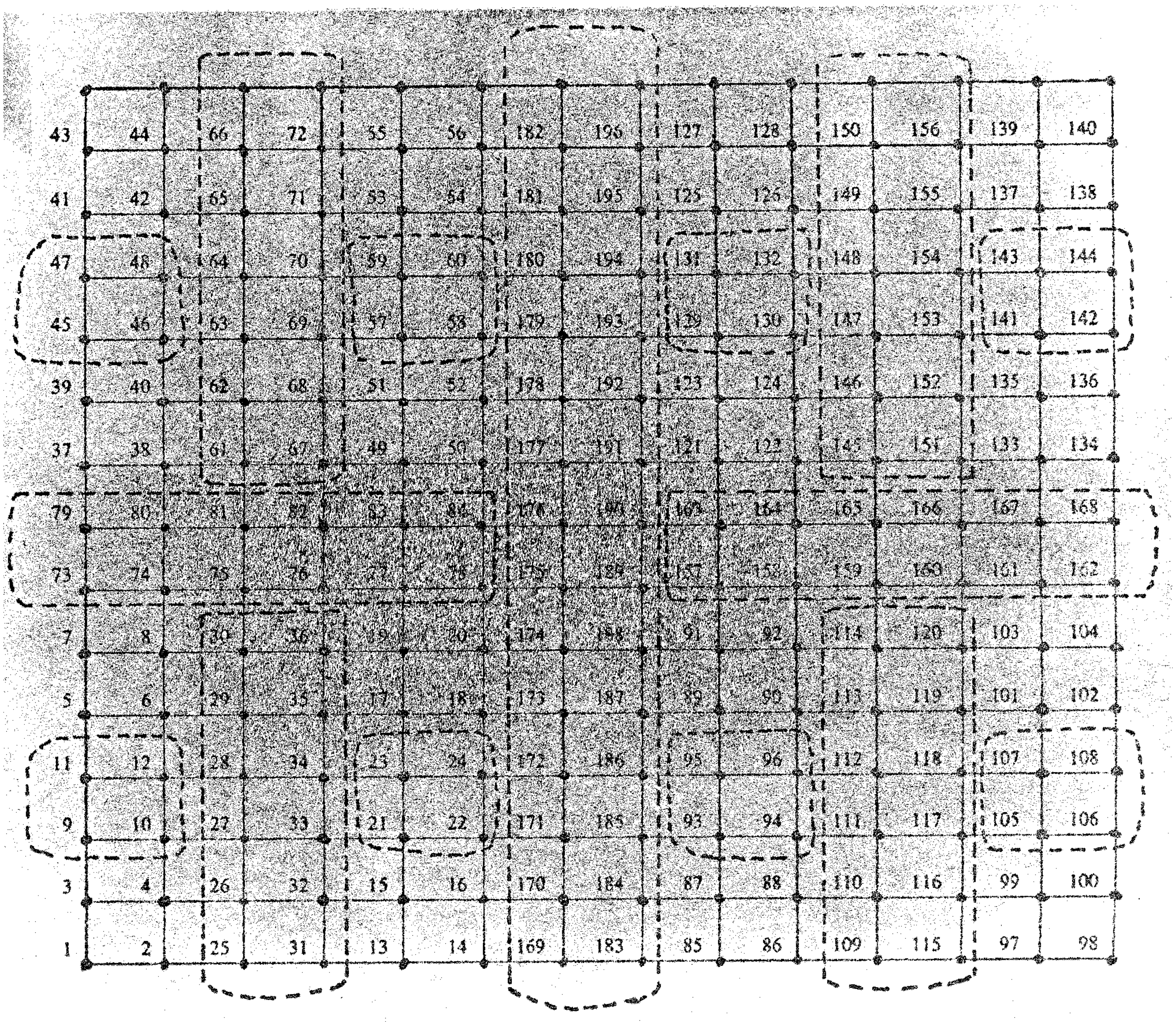


Figure 6.1 A 3 by 3 finite element grid.

Associated with each of the $n^2$ grid points (or nodes) is a variable, and associated with each small square is a set of four observations (or equations) involving the four variables at the corners of the square. This gives rise to a large sparse overdetermined system of linear equations. The unknowns are the variables at the grid points. Such a system is usually solved in the least squares sense. Denote the observation matrix by $A$.

Suppose $\alpha$ is a width-2 nested dissection ordering on the $n$ by $n$ grid. An example is given in Figure 6.2 with $n=14$. Thus we have a column permutation $P_2$ for the observation matrix $A$. We assume that the rows of $AP_2$ are sorted so that the leading column subscripts are in non-decreasing order.

Note that in this case, a minimal width-2 separator is a 2 by $p$ subgrid of the original grid, where $p \leqslant n$. Here and elsewhere, a *separator-line* refers to a grid-line in the separator that has $p$ nodes. Thus every separator has 2 separator-lines. It is assumed that the nodes in a separator are labelled separator-line by separator-line (see Figure 6.2). Thus the labellings of the nodes on one separator-line (called the *first separator-line*) will be less than those of the nodes on the other separator-line (called the *second separator-line*).



| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | 44 | 66 | 72 | 55 | 56 | 182 | 196 | 127 | 128 | 150 | 156 | 139 | 140 |
| 41 | 42 | 65 | 71 | 53 | 54 | 181 | 195 | 125 | 126 | 149 | 155 | 137 | 138 |
| 47 | 48 | 64 | 70 | 59 | 60 | 180 | 194 | 131 | 132 | 148 | 154 | 143 | 144 |
| 45 | 46 | 63 | 69 | 57 | 58 | 179 | 193 | 129 | 130 | 147 | 153 | 141 | 142 |
| 39 | 40 | 62 | 68 | 51 | 52 | 178 | 192 | 123 | 124 | 146 | 152 | 135 | 136 |
| 37 | 38 | 61 | 67 | 49 | 50 | 177 | 191 | 121 | 122 | 145 | 151 | 133 | 134 |
| 79 | 80 | 81 | 82 | 83 | 84 | 176 | 190 | 163 | 164 | 165 | 166 | 167 | 168 |
| 73 | 74 | 75 | 76 | 77 | 78 | 175 | 189 | 157 | 158 | 159 | 160 | 161 | 162 |
| 7 | 8 | 30 | 36 | 19 | 20 | 174 | 188 | 91 | 92 | 114 | 120 | 103 | 104 |
| 5 | 6 | 29 | 35 | 17 | 18 | 173 | 187 | 89 | 90 | 113 | 119 | 101 | 102 |
| 11 | 12 | 28 | 34 | 23 | 24 | 172 | 186 | 95 | 96 | 112 | 118 | 107 | 108 |
| 9 | 10 | 27 | 33 | 21 | 22 | 171 | 185 | 93 | 94 | 111 | 117 | 105 | 106 |
| 3 | 4 | 26 | 32 | 15 | 16 | 170 | 184 | 87 | 88 | 110 | 116 | 99 | 100 |
| 1 | 2 | 25 | 31 | 13 | 14 | 169 | 183 | 85 | 86 | 109 | 115 | 97 | 98 |

Figure 6.2 A width-2 nested dissection ordering on a 14 by 14 grid.

Note that for each small square in the grid, the four associated equations have the same nonzero structure. We assume that their elimination sequences are maximal, and thus they should be the same if they are reduced together. Hence for simplicity, we will consider only one of the four equations for each small square. Now we want to determine a bound for the number of nonzeros in $R$ and the cost of computing it.

In order to simplify the analysis, we introduce a so-called *bordered n by n grid* [5] which is an $n$ by $n$ grid where one or more sides of this grid are bordered by an additional grid-line. Some examples of bordered 3 by 3 grids are given in Figure 6.3.

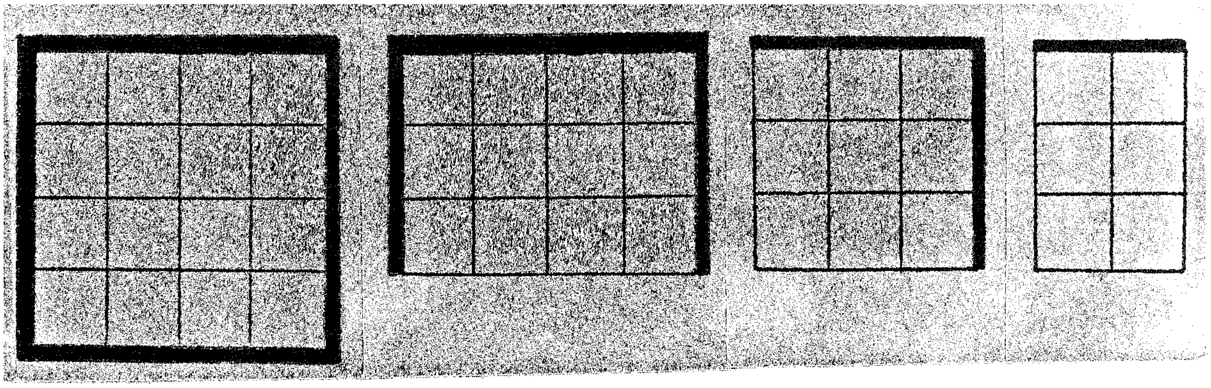
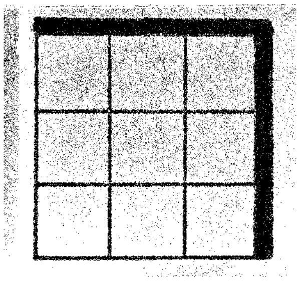
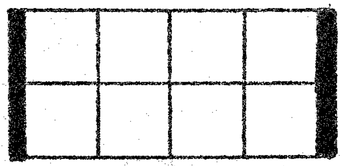


Figure 6.3 Examples of bordered 3 by 3 grids.

Note that when a grid is bordered along two sides, we assume that it has the following form

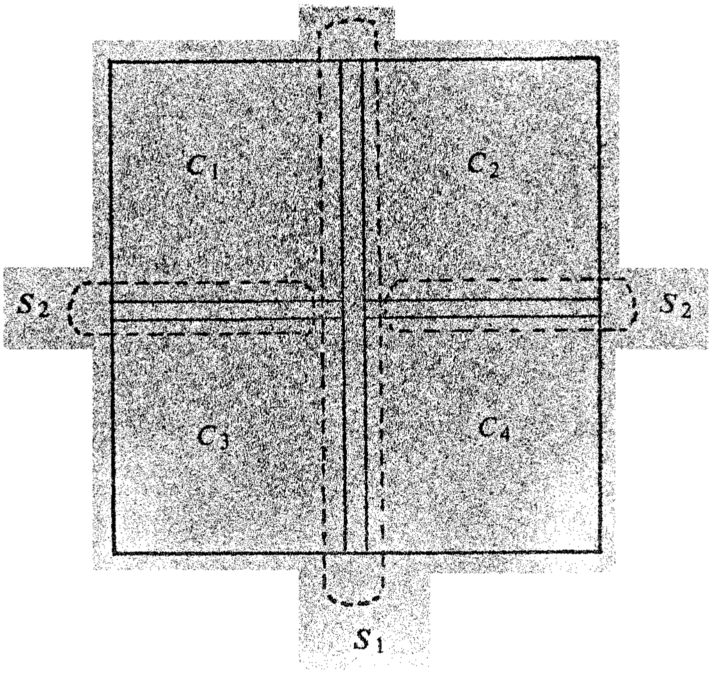


rather than



Let $\zeta(n,i)$ be the number of nonzeros in the upper triangular matrix $R$ associated with an $n$ by $n$ grid ordered by a width-2 nested dissection algorithm, where the grid is bordered along $i$ sides. Let $\theta(n,i)$ be the number of operations required to compute $R$. Clearly, $\zeta(n,0)$ and $\theta(n,0)$ are the quantities we wish to determine.

Now consider the unbordered $n$ by $n$ grid. If we apply the dissection technique twice, we will obtain the following.



($S_1$, $S_2$ are the separators.)

It should be noted that if $x \in S_1$, $y \in S_2$ and $z \in C_i$ for $1 \leqslant i \leqslant 4$, then $\alpha^{-1}(x) > \alpha^{-1}(y) > \alpha^{-1}(z)$. Moreover, as we have noted in Section 5, the equations associated with the small squares in $C_i \bigcup Adj(C_i)$ will be reduced first, while those in $S_1$ and $S_2$ will be reduced last. To make the discussion less tedious, we denote the first and second separator-lines in $S_i$ by $S_{i1}$ and $S_{i2}$ respectively. Let $\bar{j}_i = \min\{l \mid x_l \in S_{i2}\}$. Now note that $C_i$ is approximately a $\frac{1}{2}n$ by $\frac{1}{2}n$ grid which is bordered along 2 sides. Thus this dissection technique is based on the idea of "divide and conquer". Using this fact, we then obtain the following recurrence equations:

$$\zeta(n,0) = 4\zeta(\tfrac{1}{2}n,2) + \sum_j \zeta_{1j} + 2\sum_j \zeta_{2j} \cdot$$

and $$\theta(n,0) = 4\theta(\tfrac{1}{2}n,2) + \sum_k \theta_{1k} + 2\sum_k \theta_{2k} \cdot$$

Denote by $x_j$ the node in the grid with label $j$. Here $\zeta_{ij}$, $i=1,2$, is the number of nonzeros in row $j$ of $R$ with $x_j \in S_i$; and $\theta_{ik}$, $i=1,2$, is the number of operations required to reduce the $k-th$ equation in $S_i$.

To determine $\zeta_{ij}$, we recall from Lemma 4.1 that the off-diagonal nonzeros in row $j$ of $R$ are given by $Reach_{G^\alpha}(x_j, \{x_k \mid k<j\})$, and note that $\alpha$ is a width-2 nested dissection ordering. Using these facts, one can show that if $x_j \in S_1$, then the nonzeros in row $j$ of $R$ are $R_{jl}$ for $l \geqslant j$ and $x_l \in S_1$. This is illustrated by an example in Figure 6.4 in which the darkened nodes are in $Reach_{G^\alpha}(x_j, \{x_k \mid k<j\})$. Nodes which are shaded have labels less than $j$.

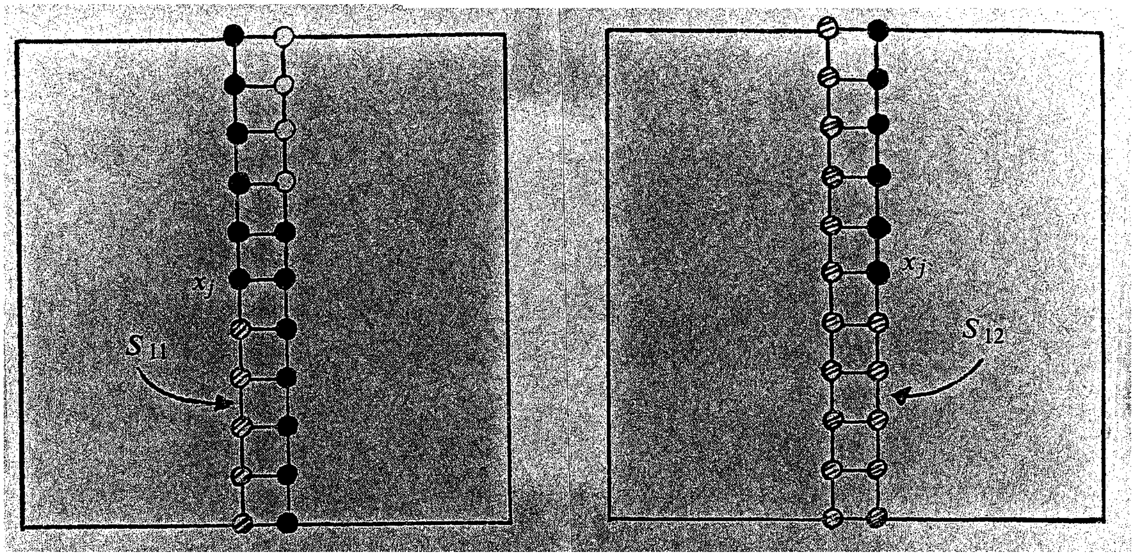


Figure 6.4 Darkened nodes are reachable from $x_j \in S_1$ through $\{x_k \mid k < j\}$. They represent the nonzeros in row $j$ of $R$. Shaded nodes have labels less than $j$. Unmarked nodes have labels greater than $j$ and are not reachable from $x_j$.

Then,

$$\zeta_{1j} \approx \begin{cases} n & \text{if} x_j \in S_{11}, \\ n-(j-\bar{j}_1) & \text{if} x_j \in S_{12}. \end{cases}$$

Similarly, one can show that if $x_j \in S_2$, then $R_{jl} \neq 0$ for $l \geqslant j$, where $x_j$ belongs to $S_2$ or $x_j$ belongs to one of the separator-lines in $S_1$ that is closest to $S_2$ (see Figure 6.5).

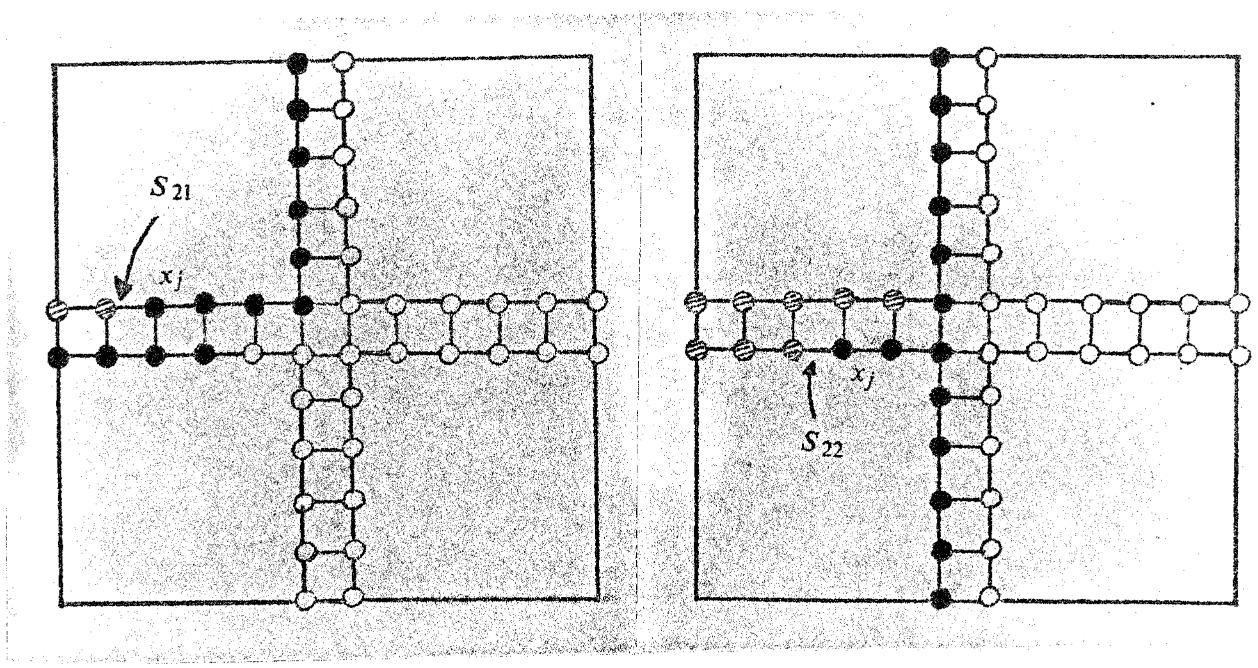


Figure 6.5 Same as Figure 6.4, except that $x_j \in S_2$.

Thus,

$$\zeta_{2j} \approx \begin{cases} n & \text{if } x_j \in S_{21}, \\ \frac{3}{2}n - (j - \overline{j}_2) & \text{if } x_j \in S_{22}. \end{cases}$$

Hence,

$$\zeta(n,0) = 4\zeta(\tfrac{1}{2}n, 2) + \frac{15}{4}n^2 + O(n).$$

We now derive $\theta_{ik}$. Note that the column and row orderings have imposed a specific order in which the equations in a separator should be reduced. Let $\Psi_{S_2}$ be the set of equations associated with the small squares in $S_2$. Clearly, the equation which has the smallest leading column subscript in $\Psi_{S_2}$ is the one which is associated with the small square at one end of $S_2$. Hence this equation will be the first one to be reduced. The leading column subscripts of the equations increase as we move from this small square to the one at the other end of $S_2$. Note that at this stage, none of the equations in $S_1$ would have been reduced. Now consider the $k-th$ equation in $S_2$. If $\xi_1^k$ is its leading column subscript, then using Lemma 4.5, the corresponding elimination sequence $\Xi^k$ will include $\{l \geqslant \xi_1^k | x_l \in S_2\}$ and $\{l \geqslant \xi_1^k | x_l$ is a node on the separator-line in $S_1$ that is closest to $S_2\}$. This is illustrated in Figure 6.6 where darkened nodes are in the elimination sequence. The labels of shaded nodes are less than $\xi_1^k$, thus these shaded nodes are not in the elimination sequence. Here $| \Xi^k | \approx 2n - k$, so

$$\theta_{2k} \approx \sum_{l=1}^{2n-k} l = 2n^2 - 2nk + \tfrac{1}{2}k^2 + O(n) .$$

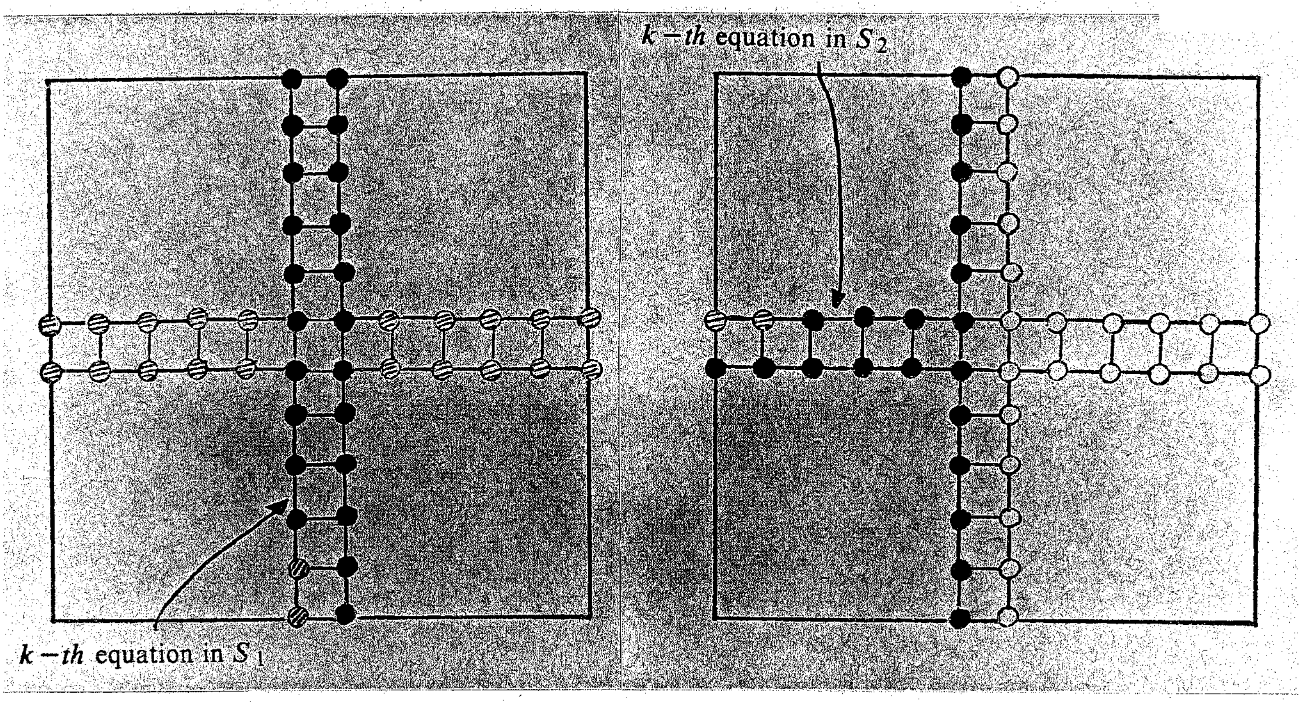


Figure 6.6 Darkened nodes are in the elimination sequence of an equation. Shaded nodes have labels less than $\xi_1^k$, they are not in the elimination sequence.

Now using a similar argument, it is easy to show that for the $k-th$ equation in $S_1$, $| \Xi^k | \approx 2n - k$, so

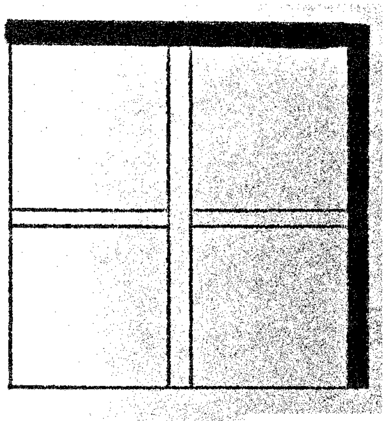
$$\theta_{1k} \approx \sum_{l=1}^{2n-k} l = 2n^2 - 2nk + \tfrac{1}{2}k^2 + O(n) .$$

Since there are approximately $n$ and $\tfrac{1}{2}n$ equations in $S_1$ and $S_2$ respectively, we then have

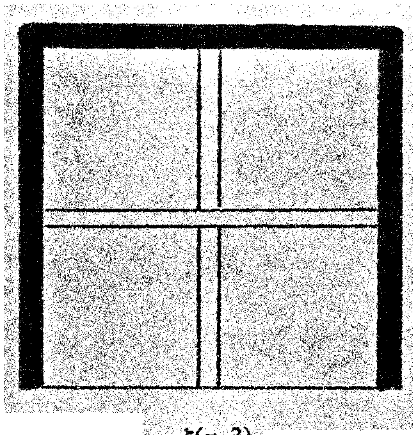
$$\theta(n, 0) = 4\theta(\tfrac{1}{2}n, 2) + \frac{65}{24}n^3 + O(n^2) .$$

Using the same technique and arguments, we can derive the recurrence equations for $\zeta(n,i)$ and $\theta(n,i)$, $i=2,3,4$.

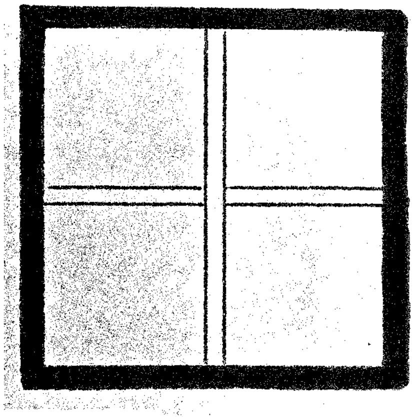
figure



$\zeta(n,2)$
$\theta(n,2)$



$\zeta(n,3)$
$\theta(n,3)$



$\zeta(n,4)$
$\theta(n,4)$

$$\zeta(n,2)=\zeta(\tfrac{1}{2}n,2)+2\zeta(\tfrac{1}{2}n,3)+\zeta(\tfrac{1}{2}n,4)+\frac{15}{2}n^2+O(n),$$

$$\zeta(n,3)=2\zeta(\tfrac{1}{2}n,3)+2\zeta(\tfrac{1}{2}n,4)+10n^2+O(n),$$

$$\zeta(n,4)=4\zeta(\tfrac{1}{2}n,4)+\frac{51}{4}n^2+O(n);$$

and

$$\theta(n,2)=\theta(\tfrac{1}{2}n,2)+2\theta(\tfrac{1}{2}n,3)+\theta(\tfrac{1}{2}n,4)+\frac{121}{12}n^3+O(n^2),$$

$$\theta(n,3)=2\theta(\tfrac{1}{2}n,3)+2\theta(\tfrac{1}{2}n,4)+\frac{187}{12}n^3+O(n^2),$$

$$\theta(n,4)=4\theta(\tfrac{1}{2}n,4)+\frac{533}{24}n^3+O(n^2).$$

The following lemmas can be proved by induction and are useful when we determine a closed form for $\zeta(n,0)$ and $\theta(n,0)$.

Lemma 6.1 [5]

a) Let $f(n) = f(\frac{1}{2}n) + kn^2\log_2 n + O(n)$.
Then $f(n) = \frac{4}{3}kn^2\log_2 n + O(n^2)$.

b) Let $g(n) = 2g(\frac{1}{2}n) + kn^2\log_2 n + O(n^2)$.
Then $g(n) = 2kn^2\log_2 n + O(n^2)$.

c) Let $h(n) = 4h(\frac{1}{2}n) + kn^2 + O(n)$.
Then $h(n) = kn^2\log_2 n + O(n^2)$.

□

Lemma 6.2 [5]

a) Let $f(n) = f(\frac{1}{2}n) + kn^3 + O(n^2\log_2 n)$.
Then $f(n) = \frac{8}{7}kn^3 + O(n^2\log_2 n)$.

b) Let $g(n) = 2g(\frac{1}{2}n) + kn^3 + O(n^2\log_2 n)$.
Then $g(n) = \frac{4}{3}kn^3 + O(n^2\log_2 n)$.

c) Let $h(n) = 4h(\frac{1}{2}n) + kn^3 + O(n^2)$.
Then $h(n) = 2kn^3 + O(n^2\log_2 n)$.

□

Now using the recurrence equations and Lemmas 6.1 and 6.2, we obtain the following results.

Theorem 6.3

Consider an $n$ by $n$ grid, and the associated least squares problem as described at the beginning of this section, with row and column orderings induced by a width-2 nested dissection labelling of the grid. Then

a) the number of nonzeros in the upper triangular matrix $R$ is bounded by

$$\frac{51}{4} n^2 \log_2 n + O(n^2) ,$$

b) the number of operations required to reduce the equations is bounded by

$$4 \times \frac{1405}{84} n^3 + O(n^2 \log_2 n) .$$

□

The factor 4 in the second part of Theorem 6.3 is due to there being four equations associated with each small square. Note that in the analysis, we consider only one of those four equations for each small square. Furthermore, the bound for the operation count may be too large. This is because in the derivation, it is assumed that if the elimination sequence of an equation is $\Xi = \{\xi_1, \xi_2, \ldots, \xi_\mu\}$, then it is maximal, and $R_{ij} \neq 0$ for $i,j \in \Xi$ and $j \geqslant i$. Of course, this may not always be true.

The following shows some numerical experiments on $n$ by $n$ grids which were carried out on an IBM 4341, using a modification of SPARSPAK ([4], [5]).

| n | no. of columns | no. of rows | storage $S$ | $\frac{S}{n^2 \log_2 n}$ | reduction time $\theta$ (sec) | $\frac{\theta}{n^3}$ |
|---|---|---|---|---|---|---|
| 10 | 100 | 324 | 2223 | 6.692 | 2.110 | .00211 |
| 12 | 144 | 484 | 3419 | 6.623 | 3.777 | .00219 |
| 14 | 196 | 676 | 5058 | 6.778 | 5.957 | .00217 |
| 16 | 256 | 900 | 7189 | 7.021 | 9.103 | .00222 |
| 18 | 324 | 1156 | 9805 | 7.257 | 13.227 | .00226 |
| 20 | 400 | 1444 | 12679 | 7.332 | 17.760 | .00222 |
| 22 | 484 | 1764 | 16076 | 7.448 | 24.160 | .00227 |

These results confirm that the reduction time and storage requirement are $O(n^3)$ and $O(n^2 \log_2 n)$ respectively. Note that "storage $S$" refers to the storage required to store the nonzeros in $R$ and the structure of $R$ (i.e., pointers, subscripts, etc.). Further implementation details can be found in [4].

## 7. References

[1] J.H. Argyris and O.E. Brønlund, "The natural factor formulation of the stiffness matrix displacement method", *Computer Methods in Applied Mech. and Eng.* 5 (1975), 97-119.

[2] A. George and J.W.H. Liu, "An automatic nested dissection algorithm for irregular finite element problems", *SIAM J. Numer. Anal.* 15 (1978), 1053-1069.

[3] A. George and J.W.H. Liu, "A minimal storage implementation of the minimum degree algorithm", *SIAM J. Numer. Anal.* 17 (1980), 282-299.

[4] A. George and M.T. Heath, "Solution of sparse linear least squares problems using Givens rotations", *Linear Algebra and its Appl.* 34 (1980), 69-83.

[5] A. George and J.W.H. Liu, *Computer solution of large sparse positive definite systems*, Prentice-Hall Inc., Englewood Cliffs, N.J., (1981).

[6] G.H. Golub, "Numerical methods for solving linear least squares problems", *Numer. Math.* 7 (1965), 206-216.