GRAPH COLORING AND THE ESTIMATION OF SPARSE JACOBIAN MATRICES WITH SEGMENTED COLUMNS*

TROND STEIHAUG† AND A. K. M. SHAHADAT HOSSAIN†

Abstract. It is well known that a sparse *Jacobian matrix* can be estimated by fewer function evaluations than the number of columns by using the CPR technique. An example shows that if the rows of the matrix are partitioned into two blocks then fewer function evaluations are needed. In this paper we show the relationship between estimating the Jacobian matrix by grouping together both rows and columns and the graph coloring problem. We give an easy implementation of the *element isolation* principle.

Key words. Jacobian matrix, graph coloring, row partitioning

AMS subject classifications. 65K05, 65H10

1. Introduction.

A number of problems in nonlinear optimization and differential equation give rise to the estimation of *Jacobian matrix* of a mapping $f : \mathbf{R}^n \to \mathbf{R}^m$. In this paper[1] we assume that if one or more components of f at x are needed the m-vector $f(x)$ has to be computed.

A forward estimate of column j of the Jacobian matrix A of a differentiable function f at x is

$$a_j = \frac{\partial}{\partial x_j} f(x) \cong \frac{1}{\varepsilon}[f(x+\varepsilon e_j) - f(x)], \quad 1 \le j \le n \tag{1}$$

where e_j is the j-th coordinate vector and ε is a positive increment. When the Jacobian matrix A is sparse, and the sparsity pattern is known, Curtis, Powell and Reid [1] observed that they could take advantage of this knowledge to reduce the number of function evaluations by approximating only the nonzeros of the Jacobian matrix. They suggest a method, the CPR technique, where the columns are partitioned into groups such that no two columns in the same group have nonzeros in the same row position. If columns j and k are in the same group then

$$(a_j + a_k)_i = a_{ij} \text{ or } (a_j + a_k)_i = a_{ik}, \qquad i = 1, 2, \cdots, m \tag{2}$$

and

$$a_j + a_k \approx \frac{1}{\varepsilon}[f(x + \varepsilon(e_j + e_k)) - f(x)], \tag{3}$$

where a_j and a_k are the j-th and k-th columns of the Jacobian matrix at x. Thus all columns in a particular group can be estimated with one extra function evaluation when $f(x)$ is known.

Coleman and Moré [2] address the problem from a graph theoretic view point. It has been observed that the problem [2] of partitioning the columns of the Jacobian matrix can be posed as the well known graph theoretic problem - the coloring problem. This combinatorial approach permits a deeper insight into the problem so that efficient algorithms could be found. They propose 3 different ordering algorithms LFO, SLO and IDO. Numerical results in [2] attest the fact that the CPR algorithm preceded by ordering of the columns often requires fewer function evaluations.

An often-cited example [2] shows that it may be advantageous to consider a more general problem where segments of columns are grouped together. In this paper, we propose a scheme to estimate sparse Jacobian matrices. In this approach both columns and rows are grouped together and the resulting segments of columns are partitioned in an effort to reduce the number of function evaluation. We also give a graph theoretic characterization of the scheme.

* This work was supported by VISTA - a research cooperation between the Norwegian Academy of Science and Den norske stats oljeselskap a.s. (Statoil).

† Department of Informatics, University of Bergen, Høyteknologisenteret, N-5020 Bergen , Norway (Trond.Steihaug@ii.uib.no and shahadat@ii.uib.no)

1 Revised May 1997. Revision of [7]

The segmented column approach described in this paper can be combined with a block iterative method similar to the techniques in Coleman [8] and Dennis and Steihaug [9]. An example of solving unconstrained optimization problem using approximate Hessian information can be found in [10] and an example of solving ordinary differential equations is given in [11].

Rest of the paper is organized as follows. In Section 2 basic concepts regarding *partitioning* are introduced. Two motivating examples are analyzed. We also discuss the molecular conformation problem that arise in molecular cluster statics. The computation of the gradient of this problem can be formulated as the computation of the Jacobian of a function in partially separable form and can be computed quite efficiently by using the technique proposed in this paper.

In Section 3 the concept of *independence* of columns is generalized to apply to *segmented columns* resulting from a row partitioning. The relationship between the estimation of the Jacobian matrix using the segmented columns and coloring the *block independent graph* is discussed. In Section 4 we show that the segmented column approach of this paper is a generalization of the *Variable Isolation* and the *Element Isolation* principles of Newsam and Ramsdell [3]. In Section 5 we give a simple procedure to construct an *expanded matrix* from a given matrix. This enables us to use coloring routines implemented in [5] to find a partition of the segmented columns. In Section 6 we present computational results. Finally, in Section 7 we give some directions for further research.

2. Row partitioning. As has been pointed out earlier, CPR technique is based on grouping together of columns such that each group can be estimated with one extra function evaluation. In this section we show that there are matrices for which computing segmented columns is more economical in terms of number of function evaluations.

A *Graph* $G = (V, E)$ is a set V of *vertices* and a set E of *edges*. An edge $e \in E$ is denoted by an unordered pair $\{u, v\}$ which connects two vertices u and v, $u, v \in V$. A graph G is said to be a *complete* graph if there is an edge between every pair of distinct vertices.

DEFINITION 2.1. A *p-coloring* of the vertices of G is a function $\Phi : V \to \{1, 2, \cdots, p\}$ such that $\{u, v\} \in E \Rightarrow \Phi(u) \neq \Phi(v)$.

The chromatic number $\chi(G)$ of G is the smallest p for which it has a *p-coloring*. An *optimal coloring* is a *p-coloring* with $p = \chi(G)$.

DEFINITION 2.2. Let $A \in \mathbf{R}^{m\times n}$ and let the columns of A be $\{a_1, a_2, \cdots, a_n\}$. The *intersection graph* of the columns of A is denoted by $G(A) = (V, E)$ where $V = \{a_1, a_2, \cdots, a_n\}$ and $\{a_i, a_j\} \in E$ if and only if $i \neq j$ and columns a_i and a_j have a nonzero in the same row position.

To estimate the Jacobian matrix we partition the columns in groups so that the columns in the same group can be estimated with one extra function evaluation. If we form the groups such that each column belongs to exactly one group then we have a *consistent partition* [2] of the columns.

The following result is given in [2] and it follows from the definition of intersection graph $G(A)$.

THEOREM 2.3. *Φ is a coloring of $G(A)$ if and only if Φ induces a consistent partition of the columns of A.* □

Let the rows of A be denoted by $\{r_1, r_2, \cdots, r_m\}$.

DEFINITION 2.4. A *row partition* of the matrix A is a mapping Π from $\{r_1, r_2, \cdots, r_m\}$ onto $\{1, 2, .., k\}$ for some $k \leq m$.

A graph interpretation of row partition is possible.

	1	2	3	4	5	6
1	X		X	X		
2		X			X	X
3	X		X			
4	X			X		X
5		X				X
6			X			

FIG. 1. *Matrix A.*

A *subgraph* $G_1 = (V_1, E_1)$ of a graph $G = (V, E)$ is a graph such that $V_1 \subset V$ and $E_1 \subset (V_1 \times V_1) \cap E$.

Given a nonempty $V_1 \subset V$, the subgraph $G_1 = (V_1, E_1)$ is induced by V_1 if $E_1 = \{\{u,v\} \in E : u, v \in V_1\}$. Note that the structure of the graph $G(A)$ is not sensitive to the row and column permutation of the matrix A.

For a given row partition Π we can re-index the rows such that they have consecutive indices. A group of rows in Π is said to form a *block*. An example row partition of A in Fig. 1 is represented by the blocks A_1, A_2, A_3, where $A_1 = \{r_1, r_2, r_3\}$, $A_2 = \{r_4\}$ and $A_3 = \{r_5, r_6\}$.

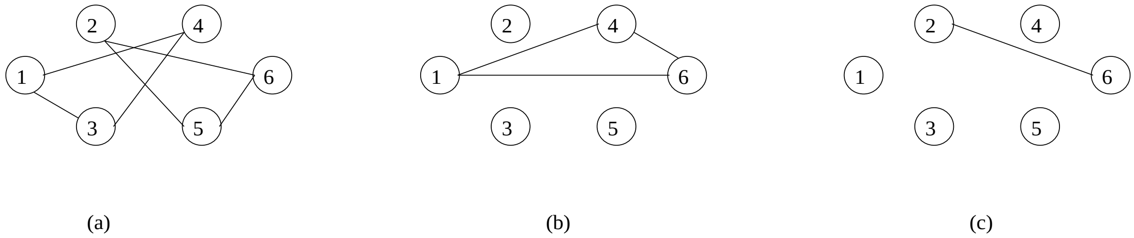

FIG. 2. *Intersection graphs (a) $G(A_1)$, (b) $G(A_2)$ and (c) $G(A_3)$*

Let $G(A_i) = (V, E_i)$, $i = 1, 2, \cdots, k$, be the intersection graph of A_i. Then $G(A) = (V, \bigcup_{i=1}^{k} E_i)$. Observe that the intersection graph corresponding to a *block* of the matrix A is a subgraph of $G(A)$, but in general this subgraph is not induced by the vertex set. It is possible to estimate A by estimating each block $A_i, i = 1, 2, \cdots, k$ separately i.e., by finding a partition of the columns of A_i and ignoring other blocks until all the nonzeros are determined. Figure 2 shows the intersection graphs $G(A_1), G(A_2)$ and $G(A_3)$ of the example in Figure 1.

2.1. Observation. Let $\gamma(A)$ be the number of groups in an optimal partition of the columns of A and let $\gamma(A_i)$ be the number of groups in an optimal partition of the columns of A_i. Since we assume that whenever one or more nonzeros of a column is needed whole column has to be computed, the total number of column computations for row partitioned matrix is $\sum_{i=1}^{k} \gamma(A_i)$. We want to know the relationship between $\gamma(A)$ and $\sum_{i=1}^{k} \gamma(A_i)$. Following examples illustrate that A can be determined with less than $\gamma(A)$ function evaluations by determining each A_i separately. In both the examples, there is a significant reduction in number of function evaluations. In the second example, we have a reduction from $O(n)$ to $O(log(n))$.

Example 1 (Eisenstat, see [2]). Let $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ where $A_1 = [D_1 D_2]$ and $A_2 = \begin{bmatrix} Z & 0 \\ D_3 & B \end{bmatrix}$, $D_1, D_2, D_3 \in \mathbf{R}^{\frac{n}{2} \times \frac{n}{2}}$ are non-singular diagonal matrices. The diagonal entries of $B \in \mathbf{R}^{\frac{n}{2} \times \frac{n}{2}}$ are zeros while off-diagonal entries are nonzeros, $Z \in \mathbf{R}^{1 \times \frac{n}{2}}$consists entirely of nonzeros and $0 \in \mathbf{R}^{1 \times \frac{n}{2}}$ is a zero vector. Further, $n \geq 6$ and n is an even number.

The columns of A_1 can be partitioned into two groups: columns of D_1 in one group and columns of D_2 in another group. This partitioning is optimal because there are only two nonzeros in each of the rows.

Each row of A_2 contains $\frac{n}{2}$ nonzeros and it can be shown that $\gamma(A_2) = \frac{n}{2}$. On the other hand, it is clear that $\gamma(A) = n$ since $G(A)$ is a complete graph on n vertices. Thus matrix A can be determined by $\frac{n}{2} + 2$ function evaluations if we partition the rows into two blocks as above. Fig. 3 shows row partitioned Eisenstat example matrix and the associated intersection graphs.

Example 2 (Exponential). Matrices in this class have several interesting features regarding their construction, sparsity and the chromatic number of the associated intersection graphs. For a positive integer $k \geq 2$, number of columns is an exponential function of k and hence the name *Exponential*. Although they are very sparse, their intersection graphs are complete. This implies that the chromatic number is equal to the number of vertices. This type of structures arise in the *molecular conformation problems* which is discussed at the end of this section. As the columns of A are identified with the vertices of $G(A)$, in the discussion below we use these terms interchangeably.

The central idea in the construction of an instance of exponential example is best described in terms of graphs: define complete bipartite graphs recursively, i.e., if G_{k-1}^1 and G_{k-1}^2 are complete bipartite subgraphs on $2^{(k-1)}$ vertices, define a new complete bipartite graph G_k on 2^k vertices such that $V_k = V_{k-1}^1 \cup V_{k-1}^2$ and $E_k = \{\{u,v\} : u \in V_{k-1}^1, v \in V_{k-1}^2\}$. For $k = 1$, G_0 is a graph with a single vertex. We note here that a row of A containing 2 nonzero entries is represented by an edge in $G(A)$.

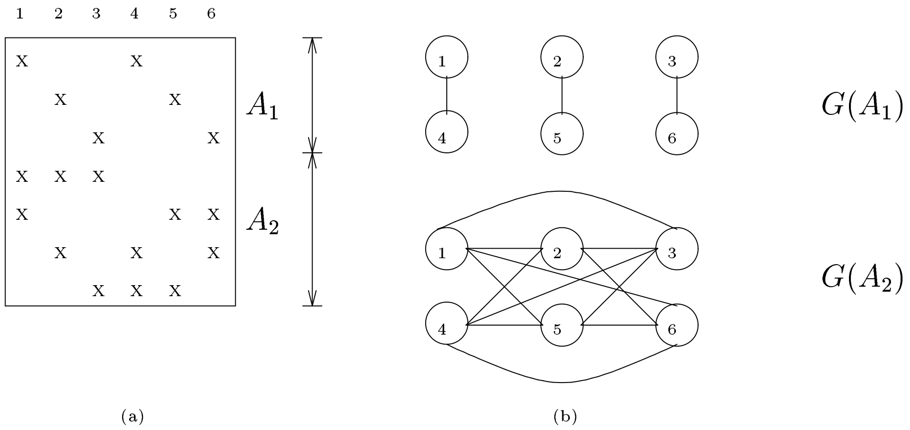

FIG. 3. *(a) Eisenstat matrix A when $n = 6$. (b) Intersection graphs $G(A_1)$ and $G(A_2)$*

Let $k \geq 2$ be given. Then the matrix A contains $n = 2^k$ columns and $m = 2^{k-1}(2^k - 1)$ rows. Without loss of generality we assume that the vertices and columns are numbered $1, 2, \cdots, n$. At step $i, i = 1, 2, \cdots, k$, block A_i is defined where $G(A_i)$ is the union of 2^{k-i} disjoint complete bipartite subgraphs.

For vertices u and v let $u_k u_{k-1} \cdots u_1$ and $v_k v_{k-1} \cdots v_1$ denote the binary representation of $u - 1$ and $v - 1$, respectively. Then $G(A_i)$ contains all edges $\{u, v\}$ such that $u_i \neq v_i$ but $u_j = v_j$ for $j > i$. Accordingly, A_i contains nonzeros in columns u and v corresponding to the edge $\{u, v\}$, one row per edge.

If the row partition is derived from the construction such that A_i constitutes the i'th block then there are k blocks in the partition. To evaluate the nonzero entries of A_i it suffices to compute

$$f(x + \epsilon \textstyle\sum_{u_i=0} e_u) \text{ and } f(x + \epsilon \textstyle\sum_{v_i \neq 0} e_v)$$

This implies that the total number of extra function evaluations needed for the computation of A is $2k$. However, since $G(A)$ is a complete graph, at least 2^k extra function evaluations are necessary to compute A without row partitioning. We like to point out that the above construction procedure can be generalized to define matrices where basic unit of construction in $G(A)$ is a clique on a constant number of vertices [6].

In Fig. 4 structure of the matrix and the graph is shown for $k = 3$. Rows of the matrix A are partitioned into blocks, A_1, A_2 and A_3. The subgraphs $G(A_1), G(A_2)$ and $G(A_3)$ are shown in Fig. 4(b).

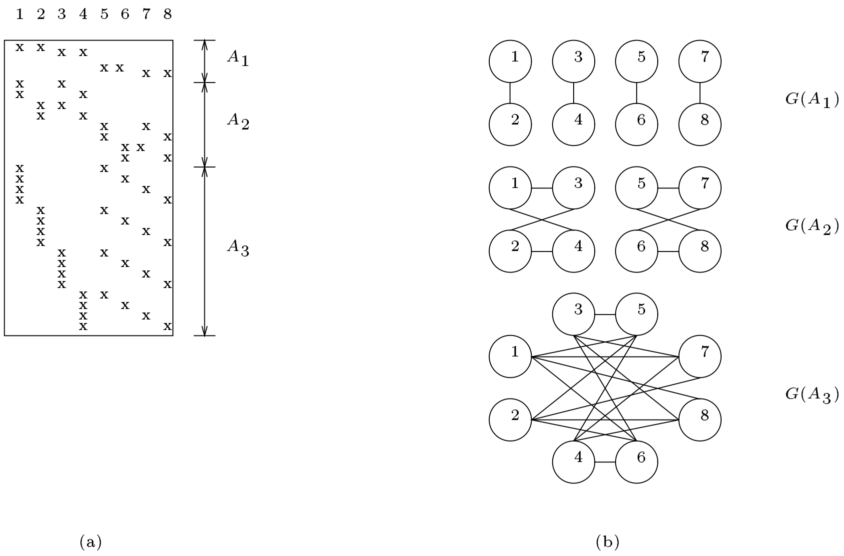

FIG. 4. *(a) Exponential matrix A when $n = 8$. (b) Intersection graphs $G(A_1)$, $G(A_2)$, and $G(A_3)$.*

2.2. Molecular Conformation Problem. The molecular conformation problem in cluster statics is concerned with the determination of the minimum energy configuration of a cluster of atoms or molecules (see [13]). This problem is formulated as a unconstrained minimization problem where the objective function can be represented in a partially separable (see [12]) form. The gradient of a partially separable function can be computed efficiently.

Given the positions $p_1, p_2, \cdots, p_n$ of n molecules in $\mathbf{R}^d$, the energy potential is defined as

$$\sum_{j=1}^{n} \sum_{i=1}^{j-1} \phi(\|p_j - p_i\|_2), \tag{4}$$

where $\phi : \mathbf{R} \to \mathbf{R}$ is the potential function between pairs of molecules. We assume that $\phi(r)$ is differentiable for $r \geq 0$. The molecular conformation problem can be posed as finding the $p = (p_1, p_2, \cdots, p_n) \in \mathbf{R}^{nd}$ such that the potential (4) is minimized.

Let f be a mapping from $\mathbf{R}^n \to \mathbf{R}$. The function f is called *partially separable* if it can be written in the form

$$f(x) = \sum_{i=1}^{m} f_i(x) = F(x)^T e, \tag{5}$$

where $e \in \mathbf{R}^m$ is a vector of unity and

$$F(x) = (f_1(x) f_2(x) \cdots f_m(x))^T \tag{6}$$

and each component function f_i depends on only $n_i << n$ variables. The gradient of the function f is given by

$$\nabla f(x) = F'(x)^T e, \tag{7}$$

where $F'(x)$ is the Jacobian matrix of F at x.

Consider the molecular conformation problem in plane i.e., $d = 2$. Let the coordinates for position p_i be x_{2i-1} and x_{2i}. Then we have $2n$ independent variables $x_1, \cdots, x_{2n}$. It follows that ϕ can be written as a function of 4 variables $v_{ij}(x_{2i-1}, x_{2i}, x_{2j-1}, x_{2j})$ for positions p_i and p_j. Let

$$V_2(x) = \sum_{j=1}^{n} \sum_{i=1}^{j-1} v_{ij}(x), \tag{8}$$

Clearly, (8) is in partially separable form. Then the gradient of V_2 is

$$\nabla V_2(x) = \tilde{V}'(x)^T e, \tilde{V}(x) = (v_{21}(x)\, v_{31}(x)\, v_{32}(x)\, \cdots\, v_{nd\, nd-1}(x))^T \tag{9}$$

where $\tilde{V}'(x)$ is the Jacobian matrix of $\tilde{V}$ at x. It is easily seen that each row of the Jacobian matrix $\tilde{V}'(x)$ contains 4 nonzero elements corresponding to the variables $x_{2i-1}, x_{2i}, x_{2j-1}, x_{2j}$.

Although $\tilde{V}'(x)$ is sparse (only 4 nonzeros per row), it may not be possible to exploit the sparsity directly since corresponding intersection graph is a complete graph. By exploiting the structure, we will see that the molecular conformation problem is related to the exponential example.

Let $G = (V, E)$ be a graph and let $e \in E$ with $e = \{u, v\}$. The *contraction* of e in G is the operation that removes e and the vertices u and v, and adds a new vertex w to the graph such that w is connected with exactly those vertices that were connected with u or v in G. An edge contraction in G is the contraction of some edge $e \in G$. Note that a sequence of edge contractions may lead to multiple edges between a pair of vertices in the resulting graph. In such a case we ignore the multiple edges and treat them as a single edge.

Consider the Jacobian matrix $\tilde{V}'(x)$. For simplicity we assume that $n = 2^k d$, $k \geq 2$. Corresponding to the component function v_{ij} there is a row in $\tilde{V}'(x)$ containing nonzero elements in columns $x_{2i-1}, x_{2i}, x_{2j-1}, x_{2j}$. In the intersection graph $G(\tilde{V}'(x))$ vertices $x_{2i-1}, x_{2i}, x_{2j-1}, x_{2j}$ constitutes a clique. For each such clique we contract edges $\{x_{2i-1}, x_{2i}\}$ and $\{x_{2j-1}, x_{2j}\}$. This corresponds to the columns $\{x_{2i-1}, x_{2i}\}$ and $\{x_{2j-1}, x_{2j}\}$

being “collapsed” into columns p_i and p_j respectively. Note that the columns x_{2i-1} and x_{2i} for $i = 1, 2, \cdots, n$ have the same sparsity structure and the collapsing is done in a structural sense. If we reorder the rows, it is clear that the above construction gives the exponential example of the last section. To actually compute $\tilde{V}'(x)$ we need 4 function evaluations for each block instead of 2. This construction carries over to three or more dimensions.

If the number of molecules n satisfies

$$\frac{n}{\lceil log_2 n \rceil} > 2d \tag{10}$$

then the number of function evaluations required to compute the Jacobian matrix is strictly less than n.

3. The Graph problem. Let Π be a row partition of A producing blocks $A_1, A_2, \cdots, A_k$, where block $A_i \in \mathbf{R}^{m_i \times n}$.

DEFINITION 3.1. Segment of column j in block A_i, denoted by $\mathbf{a}_{ij} \in \mathbf{R}^{m_i}$, is called *segmented column*

DEFINITION 3.2. Two segmented columns $\mathbf{a}_{ij}$ and $\mathbf{a}_{iq}$ are *structurally independent* if they do not have nonzeros in the same row position.

Segmented columns $\mathbf{a}_{ij}$ and $\mathbf{a}_{pq}$, $i \neq p$ and $j \neq q$, are structurally independent if

- $\mathbf{a}_{ij}$ and $\mathbf{a}_{iq}$ are structurally independent and
- $\mathbf{a}_{pj}$ and $\mathbf{a}_{pq}$ are structurally independent.

For simplicity, we say that the segmented columns $\mathbf{a}_{ij}$ and $\mathbf{a}_{pq}$ are independent when they are structurally independent.

In Fig. 5 blocks A_i and A_p and the corresponding segmented columns $\mathbf{a}_{ij}$, $\mathbf{a}_{iq}$, $\mathbf{a}_{pj}$ and $\mathbf{a}_{pq}$ are shown. Segmented columns originating from the same column are always independent. Thus $\mathbf{a}_{ij}$ and $\mathbf{a}_{pj}$ are independent. Similarly, $\mathbf{a}_{iq}$ and $\mathbf{a}_{pq}$ are independent. Since $\mathbf{a}_{ij}$ and $\mathbf{a}_{iq}$ contain nonzero in the same row, they are dependent. Then $\mathbf{a}_{ij}$ and $\mathbf{a}_{pq}$ are dependent. Similarly, $\mathbf{a}_{pj}$ and $\mathbf{a}_{iq}$ are dependent.

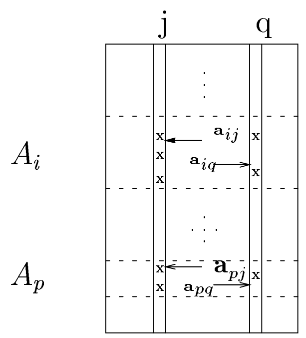

FIG. 5. *Blocks A_i and A_p of A.*

We are now ready to state our result regarding the computation of row partitioned matrix A.

LEMMA 3.3. *A group of independent segmented columns can be estimated with one extra function evaluation.*

Proof. Let rows of A be partitioned into k blocks $A_1, A_2, \cdots A_k$. Let B be a set of independent segmented columns. Then for each block A_i,

$$\sum_{j:\mathbf{a}_{ij} \in B} \mathbf{a}_{ij} \approx \frac{1}{\varepsilon} \left(f(x + \varepsilon \sum_{q:\exists p, \mathbf{a}_{pq} \in B} e_q) - f(x) \right)_i = y_i \in \mathbf{R}^{m_i}$$

where e_q is the q-th coordinate vector in $\mathbf{R}^n$. Since the segmented columns in B are independent, for a given block index p for which there exists a segmented column $\mathbf{a}_{pq} \in B$, each element of y_p is an estimate of an unique element of $\mathbf{a}_{pj}$ for some column index j. Therefore, all the segmented columns of B may be estimated with one extra function evaluation. $\square$

If the segmented columns can be grouped such that each segmented column is included in exactly one group and the segmented columns in the same group are independent then we have a *consistent* partition of the segmented columns.

We now discuss the relationship between determining a consistent partition of the segmented columns and coloring a graph.

DEFINITION 3.4. Given a matrix A and a row partition Π of A, the *block independent graph* is denoted by $G_\Pi(A) = (V, E)$, where $V = \{\mathbf{a}_{ij} : 1 \leq i \leq k, 1 \leq j \leq n\}$ and the edge $\{\mathbf{a}_{ij}, \mathbf{a}_{pq}\} \in E$ if and only if segmented columns $\mathbf{a}_{ij}$ and $\mathbf{a}_{pq}$ are not independent.

This leads to following the analogous (see [2]) result.

THEOREM 3.5. *Φ is a coloring of $G_\Pi(A)$ if and only if Φ induces a consistent partition of the segmented columns.*

Proof. Consider a consistent partition of the segmented columns and let the number of groups of segmented columns in this partition be p. Vertices corresponding to a group of independent segmented columns can be colored with the same color since there cannot be edges between these vertices by the definition of independence. On the other hand, given a p-coloring of $G_\Pi(A)$, segmented columns corresponding to a particular color are independent and can be estimated with one extra function evaluation by Lemma 3.3. Hence the coloring induces a consistent partition. □

If every block corresponding to the row partition Π is further subdivided into smaller blocks then we have a *refinement* Π' of the partition .

THEOREM 3.6. *For any refinement Π' of Π, $\chi(G_{\Pi'}(A)) \leq \chi(G_\Pi(A))$.*

Proof. Let $G_{\Pi'}(A)$ and $G_\Pi(A)$ be the block independent graphs of A corresponding to the row partitions Π' and Π respectively. Consider a coloring of $G_\Pi(A)$. Let A_i be a block in A obtained from the row partition Π and let $A_{i_1}, A_{i_2}, \cdots, A_{i_{k'}}$ be the blocks obtained from A_i under the refinement Π' of Π. If the vertices corresponding to the segmented columns $\mathbf{a}_{ij}$ and $\mathbf{a}_{iq}$ in $G_\Pi(A)$ are colored with the same color, then $\mathbf{a}_{ij}$ and $\mathbf{a}_{iq}$ are independent. This implies that the vertices corresponding to the segmented columns $\overline{\mathbf{a}}_{i_l j}, \overline{\mathbf{a}}_{i_l q}$, $l = 1, \cdots, k'$ in $G_{\Pi'}(A)$ do not contain any edge among themselves. Hence we can color these vertices in $G_{\Pi'}(A)$ with the same color. Thus if the optimal coloring of $G_\Pi(A)$ requires $\chi(G_\Pi(A))$ colors we can use this coloring to color the graph $G'_\Pi(A)$. This establishes the theorem. □

One consequence of Theorem 3.6 is that it provides an upper bound on the number of colors required for the graph $G_\Pi(A)$ in a minimal coloring.

COROLLARY 3.7. $\chi(G_\Pi(A)) \leq \chi(G(A))$.

4. Element Isolation. Newsam and Ramsdell [3] have introduced the concept of *variable isolation(VI)* and *element isolation(EI)* for the estimation of the Jacobian matrices. The *variable isolation graph* of A is simply the column intersection graph $G(A)$. On the other hand, construction of the element isolation graph is based on individual nonzero elements of A. We now show that the segmented column approach generalizes the VI and the EI principle of Newsam and Ramsdell [3].

Let α_{ij} be element ij in the Jacobian matrix. Note that $\alpha_{ij} = a_{ij}$ when number of blocks $k = m$.

DEFINITION 4.1. α_{ij} is *isolated* from α_{pq} if and only if $\alpha_{iq} = \alpha_{pj} = 0$ or $j = q$.

The EI principle is stated as follows.

DEFINITION 4.2. *The* EI *(Element Isolation) principle.* Let C_k be a set of double subscripts. The $|C_k|$ elements $\alpha_{ij}, ij \in C_k$, may be estimated in one difference

$$\frac{1}{\varepsilon}\left(\mathbf{f}(\mathbf{x} + \varepsilon \sum_{j | \exists i, ij \in C_k} \mathbf{e}_j) - \mathbf{f}(\mathbf{x})\right)$$

if $\forall ij, pq \in C_k, \alpha_{ij}$ is isolated from α_{pq}.

DEFINITION 4.3. The matrix A is *p-EI estimable* if and only if there are p sets of double subscripts, as in Definition 4.2, such that all the elements of the matrix A is uniquely determined by p differences given in Definition 4.2.

DEFINITION 4.4. The *Element Isolation graph* associated with the matrix A is $G_{EI} = (V, E)$, where $V = \{\alpha_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ and

$$E = \{\{\alpha_{ij}, \alpha_{pq}\} : \alpha_{ij} \text{ is not isolated from } \alpha_{pq}\}.$$

It follows from Definition 4.4 that the Element Isolation graph associated with the matrix A is isomorphic to $G_\Pi(A)$ when the number of blocks obtained from the row partition Π is m. This is stated in the following lemma.

LEMMA 4.5. *Given a Jacobian matrix $A \in \mathbf{R}^{m\times n}$ and a row partition Π where there are m blocks, $G_\Pi(A)$ is isomorphic to $G_{EI}(A)$.*

THEOREM 4.6. *The chromatic number of the Element Isoalation graph is minimal over all possible row partition Π.*

Proof. Let Π be any row partition. Then the partition Π' where there are m blocks, is a refinement. Hence, from Corollary 3.7

$$\chi(G_{EI}(A)) = \chi(G_{\Pi'}(A)) \leq \chi(G_\Pi(A)).$$

□

The EI partition is the "finest" (maximum number of blocks = number of rows) partition and can be obtained trivially. Theorem 4.6 shows that the EI partition is optimal.

Now consider a row partition Π where the number of blocks is less than the number of rows. Then the number of vertices and edges in the graph $G(A_\Pi)$ cannot be more than the number of vertices and edges in the graph $G(A_{EI})$. It is therefore natural to find a row partition such that the associated graph is "small" (in the sense that it has fewer vertices and edges) and that it requires no extra colors. To put it more precisely: find a row partition Π with fewest blocks such that $\chi(G_\Pi(A)) = \chi(G_{EI}(A))$.

Since obtaining a minimal coloring for $G_\Pi(A)$ in polynomial time is unlikely, we must look for heuristics. A clique is a lower bound on the number of colors in a coloring. Each row of A induces a clique in $G_\Pi(A)$ and the maximum number of nonzeros in any row provides a good lower bound on the chromatic number. A simple heuristic may be based on observing this lower bound while we search for a "coarse" row partition. We do not address this problem in this paper.

Coleman and Moré [2] have shown that obtaining an optimal partitioning of the columns of an arbitrary matrix and general graph coloring problem have an equivalent time complexity. It suggests that obtaining an optimal partitioning of the segmented columns is at least as hard as the aforementioned problems. Newsam and Ramsdell [3] have shown that deciding whether a Jacobian matrix is *p-EI-estimable* is an *NP-complete* (see Garey and Johnson [4]) problem; *p-EI-estimability* means that the Jacobian matrix can be estimated with p function evaluations. As we have stated earlier, the *p-EI-estimability* is a special case of the determination of the Jacobian matrices by segmented columns. The *NP-completeness* of the latter problem follows by restricting ourselves only to those problem instances where each row constitutes a block.

5. Matrix formulation. In Section 3 we have shown that a group of independent segmented columns can be estimated with one extra function evaluation. In this section we present a heuristic for the computation of Jacobian matrices using segmented columns. In this vein we define a larger matrix, the *Expanded matrix*, A_Π corresponding to matrix A and row partition Π such that $G(A_\Pi)$ is isomorphic to $G_\Pi(A)$. This enables us to use standard code on A_Π to estimate A. We also show how to construct $G_\Pi(A)$.

Let Π be a row partition of the matrix A that divides the matrix into blocks $A_1, A_2, \cdots A_k$ (See Fig. 7(a)). Denote the intersection graph corresponding to A_i by $G(A_i), i = 1, 2, \cdots k$. The construction of A_Π involves two phases. In the first phase, blocks A_i, $i = 1, 2, \cdots, k$ are placed successively in left to right fashion (see Fig. 7(b)) such that each segmented column is mapped to a unique column of A_Π. Here a segmented column is one which has at least one nonzero. We need not consider segmented columns where all the entries are zero. In other words, for every segmented column of A, a column is created in A_Π where all the entries are zero except that the segmented column is copied in the matching row positions. This situation is illustrated in the top part of Fig. 7(b).

In the second phase of construction, dependency information among segmented columns of A are incorporated into the matrix A_Π. Recall the definition of independent (dependent) segmented columns. Consider segmented columns $\mathbf{a}_{ij}$ and $\mathbf{a}_{iq}$ in A_i. If there are nonzeros in the same row position in $\mathbf{a}_{ij}$ and $\mathbf{a}_{iq}$ then they are dependent which implies that $\mathbf{a}_{ij}$ is also dependent on segmented columns $\mathbf{a}_{pq}$ for all $p \neq i$. Consequently, $\mathbf{a}_{ij}$ cannot be grouped together with any of the segmented columns $\mathbf{a}_{pq}$. Similarly, $\mathbf{a}_{iq}$ is dependent on segmented columns $\mathbf{a}_{pj}$ for all $p \neq i$. To enforce this restriction we simply introduce two new rows in A_Π, one containing nonzeros in the column positions mapped by the segmented columns $\mathbf{a}_{ij}$ and $\mathbf{a}_{pq}$ and the other containing nonzeros in the column positions mapped by the segmented columns $\mathbf{a}_{iq}$ and $\mathbf{a}_{pj}$ for all $p \neq i$. This is done for every pair of dependent segmented columns in A_i, $i = 1, 2, \cdots, k$ (see Fig. 7(b)). To see this dependency restriction in terms of graph theory consider vertices $\mathbf{a}_{ij}$ and $\mathbf{a}_{iq}$ in $G(A_i)$. For each such edge $\{\mathbf{a}_{ij}, \mathbf{a}_{iq}\}$ in $G(A_i)$ we connect vertex $\mathbf{a}_{ij}$ with the vertices $\mathbf{a}_{pq}$ from $G(A_p)$ for $p \neq i$. Similarly, vertex $\mathbf{a}_{iq}$ is connected with the vertices $\mathbf{a}_{pj}$ from $G(A_p)$ for $p \neq i$. This situation is illustrated in Fig. 6.

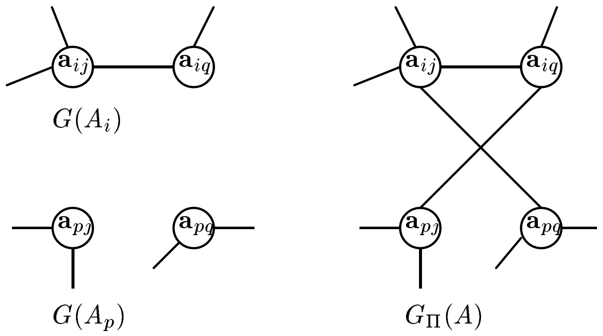

FIG. 6. *Graph $G_\Pi(A)$ before and after the insertion of edges due to the edge $\{\mathbf{a}_{ij}, \mathbf{a}_{iq}\}$ in $G(A_i)$.*

In Fig. 7(a) the matrix is partitioned into k blocks denoted by $A_1, A_2, \cdots, A_k$. In Fig. 7(b) placement of each of the blocks $A_1, A_2, \cdots, A_k$ is shown. Partitioned columns a_{ij} and a_{iq} are dependent and hence two rows containing nonzero in the appropriate columns are introduced.

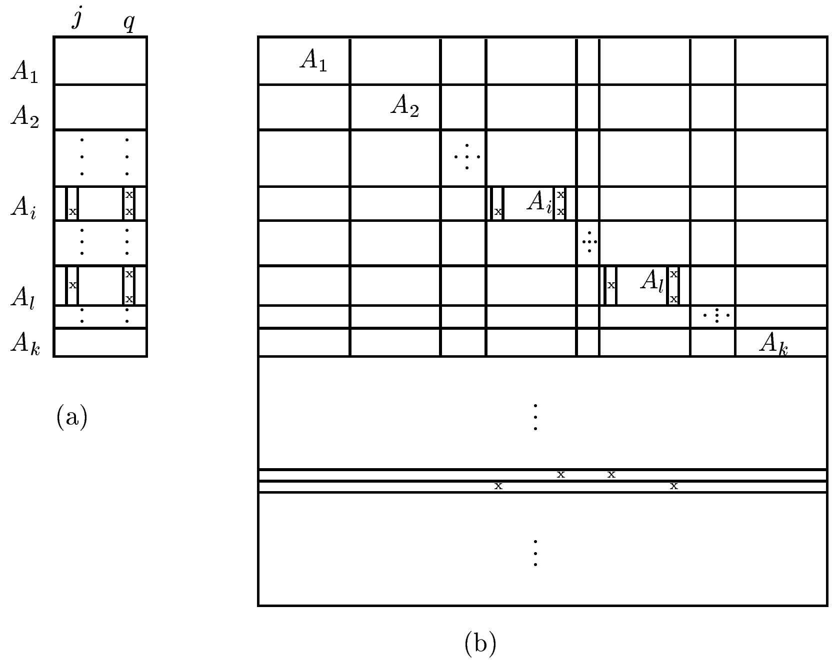

FIG. 7. *(a): Matrix A is partitioned into k blocks (b): Expanded matrix corresponding to the partition.*

THEOREM 5.1. *$G_\Pi(A)$ and $G(A_\Pi)$ are isomorphic.*

Proof. Let Π be a row partition of the matrix A. Then the vertices $\mathbf{a}_{ij}$ and $\mathbf{a}_{pq}$ in $G_\Pi(A)$ are connected by an edge if and only if the corresponding segmented columns are not independent. It implies that there must be at least one row in the block A_i or in the block A_p of the matrix A containing two nonzeros in the columns j and q. The edges in $G_\Pi(A)$ comes from two cases; one that connects vertices corresponding to segmented columns in the same block and the other that connects vertices corresponding to segmented columns from two different blocks. By the definition of $G_\Pi(A)$ and the definition of A_Π, each vertex of

TABLE 1
Exact coloring test results

Matrix	$G(A)$				$G_\Pi(A)$			
	Nodes	*Edges*	*Lower bound*	*Optimal*	*Nodes*	*Edges*	*Lower bound*	*Optimal*
ash219	85	219	4	4	438	2205	4	4
abb313	176	3206	10	10	1557	65390	6*	10*
ash331	104	331	6	6	662	4185	3	4
will199	199	960	7	7	701	7065	6	7
ash608	188	608	6	6	1216	7844	3	4
ash958	292	958	6	6	1916	12506	3	5*

$G_\Pi(A)$ corresponds to a nonzero segmented column of A and each nonzero segmented column is mapped to a unique column of A_Π. Thus if the edge $\{\mathbf{a}_{ij}, \mathbf{a}_{pq}\}$ connects vertices from different blocks, i.e., $i \neq p$ then there is an extra row in A_Π containing nonzeros in the column positions mapped by segmented columns $\mathbf{a}_{ij}$ and $\mathbf{a}_{pq}$. Thus, $\{\mathbf{a}_{ij}, \mathbf{a}_{pq}\}$ is an edge of $G_\Pi(A)$ if and only if there is an edge between vertices in $G(A_\Pi)$ corresponding to two unique columns in A_Π. $\square$

The size of the expanded matrix can be easily obtained from its construction. For a k block partition the number of columns $n' \leq n * k$ and number of rows $m' \leq m + (k-1)\sum_{i=1}^{m} \rho_i(\rho_i - 1)$ where ρ_i is the number of nonzero in the i'th row of A. In practice, however, the numbers n' and m' are smaller due to the presence of segmented columns containing zeros only.

One consequence of Theorem 5.1 and Lemma 4.5 is that if Π has m blocks then the resulting graph $G(A_\Pi)$ is isomorphic to $G_{EI}(A)$. Thus the expanded matrix procedure of this section provides an easy implementation of the EI-principle. This helps us to use existing software [5] to compute the Jacobian matrix.

6. Computational Results.

In this section we report the results from computational experiments for the segmented column approach. We test "exact" and heuristic coloring on $G(A)$, $G_\Pi(A)$ and $G(A_\Pi)$.

6.1. Test Suit.

The test suit consists of matrices from the Harwell-Boeing test matrix collection SMTAPE [16]. We also report test results for the Molecular Conformation problem.

For heuristic coloring we have used ordering algorithms LFO, IDO and SLO as implemented in *minpack* [5] subroutine *dsm*. Subroutine *dsm* first orders the vertices of $G(A)$ by each of the ordering algorithms followed by a greedy technique to find a coloring of the vertices and chooses the coloring with fewer colors. The test matrices A are chosen for which

$$maxgrp > \rho_{max}$$

where *maxgrp* is the number of colors reported by *dsm* on $G(A)$ and ρ_{max} is the maximum number of nonzero elements in any row of A. This is because ρ_{max} is a lower bound for the chromatic number of $G_\Pi(A)$ for any row partition Π.

6.2. Exact coloring.

Here we describe some experimental results using an "exact" coloring algorithm. In our experiments we construct the graph $G_\Pi(A)$ for a given matrix A and a row partition Π.

The coloring used is based on the DSATUR [14] algorithm. In Table 1 we present the following information:

- Nodes: The number of nodes in $G(A)$ and $G_\Pi(A)$
- Edges: The number of edges in $G(A)$ and $G_\Pi(A)$
- Lower bound: Size of the largest clique.
- Optimal: Value of the chromatic number of the graph.

The row partition Π is one where each row constitutes a block. Out of 6 examples we have investigated, 3 of them showed a reduction in chromatic number from $G(A)$ to $G_\Pi(A)$. The coloring routine did not terminate on examples "ash958" and "abb313". This is indicated by the asterisk in "Optimal" and/or "Lower bound" entry for the graph $G_\Pi(A)$. On these examples, value of the largest clique could not be confirmed. Whenever the coloring routine terminated it terminated quickly.

TABLE 2
Complexity of the expanded matrix.

Test suit	*Matrix*	A			A_Π			
		n	m	nnz	n'	m'	nnz'	k
Molecular Conformation	1	4	6	12	8	18	36	2
	2	8	28	56	24	140	280	3
	3	16	120	240	64	840	1680	4
	4	32	496	992	160	4464	8928	5
	5	64	2016	4032	384	22176	44352	6
	6	128	8128	16256	896	105664	211328	7
Harwell	ash219	85	219	438	438	2205	4410	219
	abb313	176	313	1557	1557	62497	125925	313
	ash331	104	331	662	662	4185	8370	331
	will199	199	199	701	701	6304	12911	199
	ash608	188	608	1216	1216	7844	15688	608
	ash958	292	958	1916	1916	12506	25012	958

The "exact" coloring results indicate that $G_\Pi(A)$ can be colored with strictly fewer colors than that is required to color $G(A)$. Also we noticed that, the time to find optimal coloring (for the instances where the algorithm terminated) is small compared to the time to verify the optimality.

It has often been observed [5] that heuristic coloring algorithms are effective on graph $G(A)$. This would suggest that we look for heuristics that are particularly suitable for $G_\Pi(A)$. Alternatively, several modifications to "exact" coloring schemes are possible. We may terminate the algorithm when a "satisfactory" coloring is achieved. Another possibility is to incorporate heuristics in an "exact" scheme. Mehrotra and Trick [15] discuss this type of heuristic algorithms with an integer linear programming (ILP) formulation of the coloring problem.

6.3. Heuristic coloring. In our implementation we first construct the expanded matrix A_Π and then use *dsm* to partition its columns. The time complexity of *dsm* is proportional to

$$\sum_{i=1}^{m} \rho_i^2$$

where ρ_i is the number of nonzero in row r_i. The first m rows of A_Π are identical to those of A and the number of extra rows is bounded by:

$$(k-1)\sum_{i=1}^{m} \rho_i(\rho_i - 1)$$

As each extra row contains exactly 2 nonzeros, number of operations required by *dsm* on A_Π is proportional to

$$\sum_{i=1}^{m} \rho_i^2 + 4(k-1)\sum_{i=1}^{m} \rho_i(\rho_i - 1) = O(k\sum_{i=1}^{m} \rho_i^2)$$

For the Harwell-Boeing test matrices we have chosen a row partition Π where each row constitutes a block. This is an easy implementation of the EI-principle.

In Table 2, n and n' denote the number of columns, m and m' denote the number of rows, nnz and nnz' denote the number of nonzeros of A and A_Π, respectively. The number of blocks is denoted by k.

The number of nonzero elements grows with a factor of $2k-1$ for the Molecular Conformation example and the time complexity in *dsm* grows with approximately $2k$ for A to A_Π.

Table 3 gives the number of colors required by *dsm* [5] on test matrices. The maximum number of nonzeros in any row is denoted by ρ_{max}. The value of ρ_{max} is same for A and A_Π. *Mingroup* is set to

$$max\{maxclq,\ \rho_{max}\}$$

TABLE 3
Coloring by *dsm*

Test suit	*Matrix*	$G(A)$			$G(A_\Pi)$	
		ρ_{max}	*Mingroup*	*Maxgroup*	*Mingroup*	*Max group*
Molecular Conformation	1	2	4	4	3	4
	2	2	8	8	4	6
	3	2	16	16	5	8
	4	2	32	32	6	10
	5	2	64	64	7	12
	6	2	128	128	8	14
Harwell	ash219	2	3	4	2	4
	abb313	6	10	10	6	10
	ash331	2	6	6	2	6
	will199	6	7	7	6	7
	sh608	2	5	6	2	6
	ash958	2	6	6	2	6

where *maxclq* is the largest clique found in LFO and IDO ordering. *Maxgroup* denotes the number of groups (colors) obtained by *dsm*. The number of colors computed by *dsm* remained same for A and A_Π on the Harwell-Boeing test matrices. This is evident by comparing *Maxgroup* columns in Table 3.

On the Eisenstat example matrices experimental results agree with our observations in Section 2. These results are not shown in Table 3.

For the Molecular Conformation problem we have a predefined partition of the rows. On this example *dsm* yielded a minimal partition of the segmented columns. This clearly shows that on problems with special structure, segmented column approach is worth considering.

7. Final Remarks. We have proposed a graph coloring based technique to compute sparse Jacobian matrices. Our segmented column approach generalizes the CPR and the EI technique.

The usefulness of this technique depends on several factors. We have shown that for any row partition Π of A, the chromatic number of $G_\Pi(A)$ can be no greater than that of $G(A)$. On 3 of the Harwell-Boeing test matrices the "exact" coloring algorithm computes a coloring for $G_\Pi(A)$ that has strictly fewer color than in the coloring for $G(A)$. We have not been able to obtain a better coloring for $G(A_\Pi)$ compared with $G(A)$ using the ordering heuristics LFO, IDO and SLO on Harwell-Boeing test matrices. However, we have shown that special structures as in the Molecular Conformation problems can be exploited using segmented column approach.

For large problems it is often impractical to use "exact" coloring. An interesting further work in this regard would be to design heuristics based on a scheme where it is not required to find an optimal coloring. Such a scheme should take into account the special structure of $G_\Pi(A)$. Another interesting work is concerned with finding row partitions. The graph $G(A_{EI})$ is likely to be "large". It is advantageous to find a row partition Π with fewest blocks such that $\chi(G(A_\Pi)) = \chi(G(A_{EI}))$.

REFERENCES

[1] A. R. Curtis, M. J. D. Powell and J. K. Reid, *On the estimation of sparse Jacobian matrices,* J. Inst. Math. Appl., 13 (1974), pp. 117–119.
[2] T. F. Coleman and J. J. Moré, *Estimation of sparse Jacobian matrices and graph coloring problems,* SIAM J. Numer. Anal., 20 (1983), pp. 187–209.
[3] G. N. Newsam and J. D. Ramsdell, *Estimation of Jacobian matrices,* SIAM J. Alg. Disc. Meth., 4 (1983), pp. 404-418.
[4] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H.Freeman, San Francisco, 1979.
[5] T. F. Coleman, B. S. Garbow and J. J. Moré, *Software for estimating sparse Jacobian matrices,* ACM Trans. Math. Software, 10 (1984), pp. 329-345.
[6] A. K. M. Shahadat Hossain, *Estimation of sparse Jacobian Matrices by using row and column partitioning and the graph coloring approach*, Cand.Scient thesis, Department of Informatics, University of Bergen, Norway, 1992.

[7] Trond Steihaug and A. K. M. Shahadat Hossain, *Graph coloring and the estimation of sparse Jacobian matrices using row and column partitioning* , Report 72, Department of Informatics, University of Bergen, 1992.
[8] T. F. Coleman, *Large Sparse Numerical Optimization*, Lecture Notes in Computer Science 165, Springer-Verlag, Berlin.
[9] J. E. Dennis, JR. and Trond Steihaug, *On the Successive Projections Approach to Least-Squares Problems*, SIAM J. Numer. Anal., 23 (1986), pp. 717–733.
[10] Trond Steihaug, *The Conjugate Gradient Method and Trust Regions in Large Scale Optimization*, SIAM J. Numer. Anal., 20 (1983), pp. 626–637.
[11] Trond Steihaug and Arne Wolfbrandt, *An attempt to avoid Exact Jacobian and Nonlinear Equations in the Numerical Solution of Stiff Differential Equations*, Math. Comp., 33 (1979), pp. 521–534.
[12] J. J. Moré and Jhijun Wu, *Global smoothing and continuation for large-scale molecular optimization*, Preprint MCS-P539-1095, Argonne National Laboratory, Argonne, Illinois, 1995.
[13] A. Griewank and P. L. Toint, *Numerical experiments with partially separable optimization problems*, in Numerical Analysis: Proceedings Dundee 1983, D.F. Griffiths, ed., Lecture Notes in Mathematics 1066, Springer-Verlag, 1984.
[14] Daniel Brélaz, *New Methods to Color the Vertices of a Graph*, Communication of the ACM, 22(1979), pp. 251–256.
[15] Anuz Mehrotra and Michael A. Trick, *A Column Generation Approach for Graph Coloring*, INFORMS Journal on Computing, 8(1996), pp. 344–354.
[16] Iain Duff, Roger G. Grimes, and John G. Lewis, *Users' guide for the Harwell-Boeing sparse matrix collection (Release I)*, Technical Report TR/PA/92/86, CERFACS, October 1992.