

A Hybrid Algorithm for Solving Sparse Nonlinear Systems of Equations*

By J. E. Dennis, Jr. and Guangye Li**

Abstract. This paper presents a hybrid algorithm for solving sparse nonlinear systems of equations. The algorithm is based on dividing the columns of the Jacobian into two parts and using different algorithms on each part. The hybrid algorithm incorporates advantages of both component algorithms by exploiting the special structure of the Jacobian to obtain a good approximation to the Jacobian, using as little effort as possible. A Kantorovich-type analysis and a locally q -superlinear convergence result for this algorithm are given.

1. Introduction. Consider the nonlinear system of equations

$$(1.1) \quad F(x) = 0,$$

where $F: R^n \rightarrow R^n$ is continuously differentiable on an open convex set $D \subset R^n$, and the Jacobian matrix $F'(x)$ is sparse. To solve the system, we use the iteration

$$(1.2) \quad \bar{x} = x - B^{-1}F(x),$$

where x is the current iterate, \bar{x} is the new iterate, and B is an approximation to $F'(x)$, which has the same sparsity as the Jacobian.

Suppose we have finished the current iteration. Then the information we have is $x, \bar{x}, F(x), F(\bar{x}), B$. The purpose of this paper is to find a matrix \bar{B} which is a good approximation to $F'(\bar{x})$ and to economize on the number of function evaluations required for this approximation.

In 1970 Schubert [11] gave a sparse modification of Broyden's [1] update. Broyden [2] also gave this algorithm independently. In order to present Schubert's algorithm, we introduce the following notation concerning the sparsity pattern of the Jacobian:

Definition 1.1. For $j = 1, 2, \dots, n$ define the subspace $Z_j \subset R^n$ determined by the sparsity pattern of the j th row of the Jacobian:

$$Z_j \equiv \{v \in R^n : e_i^T v = 0 \text{ for all } i \text{ such that } [F'(x)]_{ji} = 0 \text{ for all } x \in R^n\},$$

Received August 4, 1986; revised November 17, 1986 and January 8, 1987.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 65H10.

Key words and phrases. Finite difference, Jacobian, q -superlinear convergence, Kantorovich type analysis, sparsity, nonlinear system of equations.

*This research was partially supported by contracts and grants: DOE DE-AS05-82ER1-13016 and AFOSR 85-0243, and it was also partially supported by Jilin University, People's Republic of China. This work forms a portion of Dr. Li's Ph.D. dissertation in Mathematical Sciences, Rice University, written under the direction of Professor Dennis.

***Permanent address:* Computer Center, Jilin University, People's Republic of China.

where e_i is the i th column of the $n \times n$ identity matrix. Define the set of matrices Z that preserve the sparsity pattern of the Jacobian:

$$Z \equiv \{A \in L(R^n) : A^T e_j \in Z_j \text{ for } j = 1, 2, \dots, n\}.$$

Definition 1.2. For $j = 1, 2, \dots, n$ define the projection operator, $D_j \in L(R^n)$, that maps R^n onto Z_j :

$$D_j \equiv \text{diag}(d_{j1}, d_{j2}, \dots, d_{jn}),$$

where

$$d_{ji} = \begin{cases} 1, & \text{if } e_i \in Z_j, \\ 0, & \text{otherwise.} \end{cases}$$

For a scalar $\alpha \in R$, define the pseudoreciprocal

$$\alpha^+ = \begin{cases} \alpha^{-1}, & \text{if } \alpha \neq 0, \\ 0, & \text{if } \alpha = 0. \end{cases}$$

Now Schubert's update is formulated as follows:

$$(1.3) \quad \bar{B} = B + \sum_{j=1}^n ([s]_j^T [s]_j)^+ e_j e_j^T (y - Bs) [s]_j^T,$$

where $[s]_j = D_j s$, $s = \bar{x} - x$ and $y = F(\bar{x}) - F(x)$.

The advantage of Schubert's algorithm is that at each iteration only one function evaluation is required, and it is q -superlinearly convergent (see Marwil [8]). However, it usually requires more iterations than finite difference algorithms (see Li [7]).

Curtis, Powell, and Reid [4] proposed a finite difference algorithm, called the CPR algorithm, which is based on a partition of the columns of the Jacobian. Coleman and Moré [3] associate the partition problem with a graph coloring problem and gave some partitioning algorithms which can make the number of function evaluations needed to approximate the Jacobian by the CPR algorithm optimal or nearly optimal.

Following Coleman and Moré, we give some definitions concerning a partition of the columns of the Jacobian.

Definition 1.3. A partition of the columns of a matrix B is a division of the columns into groups c_1, c_2, \dots, c_p such that each column belongs to one and only one group.

Definition 1.4. A partition of the columns of a matrix B is consistent with the direct determination of B if, whenever b_{ij} is a nonzero element of B , then the group containing column j has no other column with a nonzero element in row i .

The CPR algorithm can be formulated as follows: For a given consistent partition of the columns of the Jacobian, which divides the set $\{1, \dots, n\}$ into p subsets c_1, \dots, c_p (for convenience, c_i , $i = 1, 2, \dots, p$, denotes both the sets of the columns and the sets of the indices of these columns), obtain vectors d_1, d_2, \dots, d_p such that B is determined uniquely by the equations

$$(1.4) \quad Bd_i = F(x + d_i) - F(x) \equiv y_i, \quad i = 1, 2, \dots, p.$$

Notice that for the CPR algorithm the number of function evaluations at each iteration is $p + 1$. Since the partition of the columns of the Jacobian plays an

important role in the CPR algorithm, we call the CPR algorithm based on Coleman and Moré's algorithms the CPR-CM algorithm.

The advantage of the CPR algorithm is that it usually requires fewer iterations than Schubert's algorithm. However, it requires more function evaluations at each iteration than Schubert's algorithm (see Li [7]).

In [7] we proposed an algorithm called the secant/finite difference (SFD) algorithm, which is also based on a consistent partition of the columns of the Jacobian. However, it uses the information we already have at every iterative step more efficiently than the CPR algorithm. Let

$$(1.5) \quad d_i = \sum_{j \in c_i} s_j e_j, \quad i = 1, 2, \dots, p,$$

$$(1.6) \quad g_i = \sum_{j=1}^i d_j, \quad g_0 = 0,$$

and

$$(1.7) \quad y_i = F(\bar{x} - g_{i-1}) - F(\bar{x} - g_i), \quad i = 1, 2, \dots, p,$$

where $s_i = \bar{x}_i - x_i$ denotes the i th component of s . The SFD algorithm can be formulated as follows: If $s_j \neq 0$ for some $j \in c_i$, then the j th column of \bar{B} is determined uniquely by the equations

$$\bar{B}d_i = y_i.$$

If $s_j = 0$, then the j th column of \bar{B} is equal to the j th column of B .

Since

$$(1.8) \quad \begin{aligned} y_1 &= F(\bar{x} - g_0) - F(\bar{x} - g_1) = F(\bar{x}) - F(\bar{x} - g_1), \\ y_p &= F(\bar{x} - g_{p-1}) - F(\bar{x} - g_p) = F(\bar{x} - g_{p-1}) - F(x), \end{aligned}$$

the number of function evaluations required by the SFD algorithm at each iteration is one less than that required by the CPR-CM algorithm.

As an example, consider

$$(1.9) \quad \begin{bmatrix} \times & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \times & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \times & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \times & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & \times & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 & \times & 0 \\ \times & \times & \times & 0 & 0 & 0 & 0 & \times \end{bmatrix}.$$

The partition $c_1 = \{1\}$, $c_2 = \{2\}$, $c_3 = \{3\}$, $c_4 = \{4, 5, 6, 7\}$ is an optimal consistent partition of the columns of the Jacobian. For this problem, the CPR-CM algorithm and the SFD algorithm require five and four function evaluations at each iteration, respectively.

In this paper we propose a hybrid algorithm for solving nonlinear systems of equations which is a combination of the SFD algorithm and Schubert's algorithm (including Broyden's algorithm). For some problems, this algorithm can reduce the number of function evaluations required at each iteration to fewer than that for the

SFD algorithm by exploiting the special structure of the Jacobian. For example, in (1.9) the number of function evaluations is two.

The hybrid algorithm and its properties are given in Section 2. A Kantorovich-type analysis for this algorithm is given in Section 3. A q -superlinear convergence result is given in Section 4.

Throughout the paper, $L(R^n)$ denotes the linear space of all real $n \times n$ matrices, $\|\cdot\|_F$ the Frobenius norm of a matrix, and $\|\cdot\|$ the l_2 -vector norm.

2. The Hybrid Algorithm and Its Properties. Consider the example (1.9). The first three columns of the matrix are denser than the other columns, and this makes p , the number of the groups in the partition, at least 4. The hybrid algorithm divides the columns of the Jacobian into two parts, and uses different algorithms on each part.

We say a group of the columns of a matrix has ‘good sparsity’ if the columns in this group have few nonzeros in the same row position. Otherwise, we say the group of the columns has ‘bad sparsity’.

Suppose the columns of the Jacobian can be divided into two groups—the good sparsity group c and the bad sparsity group c_1 . For convenience, we use c and c_1 to denote both the groups of the columns of a matrix and the sets of the indices of these columns. Then,

$$c \cup c_1 = \{1, \dots, n\}.$$

For any matrix $A \in L(R^n)$, let

$$A_1 = A \sum_{j \in c_1} e_j e_j^T, \quad A_2 = A \sum_{j \in c} e_j e_j^T.$$

Then $A = A_1 + A_2$. The main idea of the hybrid algorithm is to use Schubert’s update (including Broyden’s update) on B_1 , and to use the SFD algorithm on B_2 , where $B = B_1 + B_2$.

In practice, there are many ways to choose c and c_1 . For example, we can first partition the columns by using a CPR-CM procedure. Then, if we can afford m evaluations of F at each iteration, we can keep the columns of the $m - 1$ largest groups of the partition for c and put all the remaining columns into c_1 .

ALGORITHM 2.1. Given a consistent partition of B_2 , which divides c into $p - 1$ subsets c_2, c_3, \dots, c_p , and given an $x^0 \in R^n$ and a nonsingular matrix B^0 with the same sparsity as the Jacobian, for each $k \geq 0$ do the following:

- (1) Solve $B^k s_N^k = -F(x^k)$.
- (2) Choose x^{k+1} by $x^{k+1} = x^k + s_N^k$ or by a global strategy such as a trust-region method. Let $s^k = x^{k+1} - x^k$.
- (3) Check for convergence.
- (4) Update B_1^k by Schubert’s update to get B_1^{k+1} , and update B_2^k by the SFD algorithm to get B_2^{k+1} .
- (5) Set

$$B^{k+1} = B_1^{k+1} + B_2^{k+1}.$$

Let d_i, g_i and y_i be defined as in (1.5)–(1.7), and let

$$(2.1) \quad J_i = \int_0^1 F'(\bar{x} - g_i + t(g_i - g_{i-1})) dt, \quad i = 1, 2, \dots, p.$$

Then,

$$(2.2) \quad J_i d_i = y_i, \quad i = 1, 2, \dots, p,$$

and the update of Algorithm 2.1 can be formulated as

$$(2.3) \quad \begin{cases} \bar{B}_1 = B_1 + \sum_{i=1}^n ([d_1]_i^T [d_1]_i)^+ e_i e_i^T (y_1 - B_1 d_1) [d_1]_i^T, \\ \bar{B}_2 = B_2 + \sum_{i=2}^p \sum_{j \in c_i} s_j^+ s_j (J_i - B_2) e_j e_j^T, \\ \bar{B} = \bar{B}_1 + \bar{B}_2. \end{cases}$$

We now give some of the properties of \bar{B} obtained from (2.3).

LEMMA 2.1. \bar{B} satisfies the secant equations

$$(2.4) \quad \bar{B} d_i = y_i, \quad i = 1, \dots, p,$$

and (2.4) implies that

$$(2.5) \quad \bar{B} s = F(\bar{x}) - F(x) = y.$$

LEMMA 2.2. \bar{B} is the unique solution to

$$(2.6) \quad \min\{\|\hat{B} - B\|_F: \hat{B} d_i = y_i, \quad i = 1, \dots, p, \text{ and } \hat{B} \in Z\}.$$

The proof of this lemma is similar to that for Schubert's algorithm given by Reid [10] and Marwil [8].

THEOREM 2.3. If $A \in L(R^n)$ has the same sparsity as the Jacobian, then

$$(2.7) \quad \begin{aligned} \|\bar{B}_1 - A_1\|_F^2 &= \|B_1 - A_1\|_F^2 - \frac{1}{\|s\|^2} \|(B_1 - A_1)s\|^2 \\ &+ \sum_{i=1}^n ([d_1]_i^T [d_1]_i)^+ [e_i^T (y_1 - A d_1)]^2. \end{aligned}$$

Proof. Let $\bar{E}_1 = \bar{B}_1 - A_1$ and $E_1 = B_1 - A_1$. From (2.3) we have

$$(2.8) \quad e_i^T \bar{B}_1 = e_i^T B_1 + ([d_1]_i^T [d_1]_i)^+ e_i^T (y_1 - B_1 d_1) [d_1]_i^T.$$

Subtracting $e_i^T A_1$ from both sides of (2.8), and noticing that $e_i^T B_1 d_1 = e_i^T B_1 [d_1]_i$ and that $e_i^T A_1 d_1 = e_i^T A_1 [d_1]_i$, we obtain

$$(2.9) \quad \begin{aligned} e_i^T \bar{E}_1 &= e_i^T E_1 + ([d_1]_i^T [d_1]_i)^+ e_i^T (y_1 - B_1 d_1) [d_1]_i^T \\ &= e_i^T E_1 (I - ([d_1]_i^T [d_1]_i)^+ [d_1]_i [d_1]_i^T) \\ &+ ([d_1]_i^T [d_1]_i)^+ e_i^T (y_1 - A_1 d_1) [d_1]_i^T. \end{aligned}$$

Since $([d_1]_i^T [d_1]_i)^+ e_i^T (y_1 - A_1 d_1)$ is a scalar, the first and second terms on the right of (2.9) are perpendicular to each other, and we have

$$\begin{aligned} \|e_i^T \bar{E}_1\|^2 &= \|e_i^T E_1 (I - ([d_1]_i^T [d_1]_i)^+ [d_1]_i [d_1]_i^T)\|^2 + ([d_1]_i^T [d_1]_i)^+ |e_i^T (y_1 - A_1 d_1)|^2 \\ &= \|e_i^T E_1\|^2 - ([d_1]_i^T [d_1]_i)^+ |e_i^T E_1 [d_1]_i|^2 + ([d_1]_i^T [d_1]_i)^+ |e_i^T (y_1 - A_1 d_1)|^2 \\ &\leq \|e_i^T E_1\|^2 - \frac{1}{\|s\|^2} |e_i^T E_1 d_1|^2 + ([d_1]_i^T [d_1]_i)^+ |e_i^T (y_1 - A_1 d_1)|^2. \end{aligned}$$

Therefore,

$$\begin{aligned}
\|\bar{B}_1 - A_1\|_F^2 &= \sum_{i=1}^n \|e_i^T \bar{E}_1\|^2 \\
&\leq \|B_1 - A_1\|_F^2 - \frac{1}{\|s\|^2} \|(B_1 - A_1)d_1\|^2 \\
&\quad + \sum_{i=1}^n ([d_1]_i^T [d_1]_i)^+ [e_i^T (y_1 - A_1 d_1)]^2 \\
&= \|B_1 - A_1\|_F^2 - \frac{1}{\|s\|^2} \|(B_1 - A_1)s\|^2 \\
&\quad + \sum_{i=1}^n ([d_1]_i^T [d_1]_i)^+ [e_i^T (y_1 - A_1 d_1)]^2. \quad \square
\end{aligned}$$

THEOREM 2.4. *If $A \in L(R^n)$ has the same sparsity as the Jacobian, then*

$$\begin{aligned}
(2.10) \quad \|\bar{B}_2 - A_2\|_F^2 &\leq \|B_2 - A_2\|_F^2 - \frac{1}{\|s\|^2} \|(B_2 - A_2)s\|^2 \\
&\quad + \sum_{i=2}^p \sum_{j \in c_i} s_j^+ s_j \|(J_i - A)e_j\|^2.
\end{aligned}$$

Proof. Let $\bar{E}_2 = \bar{B}_2 - A_2$ and $E_2 = B_2 - A_2$. It follows from (2.3) that if $j \in c_i$, $i = 2, \dots, p$, then

$$(2.11) \quad \bar{B}_2 e_j = B_2 e_j + s_j^+ s_j (J_i - B_2) e_j.$$

Subtracting $A_2 e_j$ from both sides of (2.11), we obtain

$$\bar{E}_2 e_j = (1 - s_j^+ s_j) E_2 e_j + s_j^+ s_j (J_i - A_2) e_j.$$

Since $(1 - s_j^+ s_j) s_j^+ s_j = 0$, we have

$$\begin{aligned}
\|\bar{E}_2 e_j\|^2 &= (1 - s_j^+ s_j) \|E_2 e_j\|^2 + s_j^+ s_j \|(J_i - A_2) e_j\|^2 \\
&= \|E_2 e_j\|^2 - s_j^+ s_j \|E_2 e_j\|^2 + s_j^+ s_j \|(J_i - A_2) e_j\|^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
(2.12) \quad \|\bar{E}_2\|_F^2 &= \sum_{j \in c} \|\bar{E}_2 e_j\|^2 \\
&= \|E_2\|_F^2 - \sum_{j \in c} s_j^+ s_j \|E_2 e_j\|^2 + \sum_{i=2}^p \sum_{j \in c_i} s_j^+ s_j \|(J_i - A) e_j\|^2.
\end{aligned}$$

In addition,

$$\begin{aligned}
\sum_{j \in c} s_j^+ s_j \|E_2 e_j\|^2 &= \left\| E_2 \sum_{j \in c} s_j^+ s_j e_j e_j^T \right\|_F^2 \\
&\geq \frac{\|E_2 \sum_{j \in c} s_j^+ s_j e_j e_j^T s\|^2}{\|s\|^2} = \frac{\|E_2 s\|^2}{\|s\|^2}.
\end{aligned}$$

Thus, (2.10) follows from (2.12). \square

3. A Kantorovich-Type Analysis. To study the convergence properties of Algorithm 2.1, we assume that F' satisfies the following Lipschitz condition in the domain D defined after (1.1): For every $i \in c$, there exists $\gamma_i > 0$ such that

$$(3.1) \quad \|(F'(x) - F'(y))e_i\| \leq \gamma_i \|x - y\| \quad \forall x, y \in D,$$

and there exists $\theta_i > 0$, $i = 1, 2, \dots, n$ such that

$$(3.2) \quad \|e_i^T(F'(x)_1 - F'(y)_1)\| \leq \theta_i \|x - y\| \quad \forall x, y \in D.$$

Let $\gamma = (\sum_{i \in c} \gamma_i^2)^{1/2}$, $\theta = (\sum_{i=1}^n \theta_i^2)^{1/2}$, $\alpha = (\gamma^2 + \theta^2)^{1/2}$. If F' satisfies this Lipschitz condition, then the following are true:

$$(3.3) \quad \|F'(x)_1 - F'(y)_1\|_F \leq \theta \|x - y\| \quad \forall x, y \in D,$$

$$(3.4) \quad \|F'(x)_2 - F'(y)_2\|_F \leq \gamma \|x - y\| \quad \forall x, y \in D,$$

and

$$(3.5) \quad \|F'(x) - F'(y)\|_F \leq \alpha \|x - y\| \quad \forall x, y \in D.$$

LEMMA 3.1. *Let F' satisfy (3.1) and (3.2), and let \bar{B} be generated by Algorithm 2.1. If $\bar{x} \in D$ and $\bar{x} - d_1 \in D$, then for any $z \in D$,*

$$(3.6) \quad \begin{aligned} \|\bar{B}_1 - F'(z)_1\|_F^2 &\leq \|B_1 - F'(z)_1\|_F^2 - \frac{1}{\|s\|^2} \|(B_1 - F'(z)_1)s\|^2 \\ &\quad + \theta^2 \left(\|\bar{x} - z\| + \frac{1}{2} \|d_1\| \right)^2. \end{aligned}$$

Proof. Substituting $F'(z)$ for A in (2.7), we obtain

$$(3.7) \quad \begin{aligned} \|\bar{B}_1 - F'(z)_1\|_F^2 &\leq \|B_1 - F'(z)_1\|_F^2 - \frac{1}{\|s\|^2} \|(B_1 - F'(z)_1)s\|^2 \\ &\quad + \sum_{i=1}^n ([d_1]_i^T [d_1]_i)^+ [e_i^T (y_1 - F'(z)d_1)]^2. \end{aligned}$$

By (2.1), (2.2), (3.3), and the Cauchy-Schwarz inequality, we have

$$(3.8) \quad \begin{aligned} &\sum_{i=1}^n ([d_1]_i^T [d_1]_i)^+ [e_i^T (y_1 - F'(z)d_1)]^2 \\ &= \sum_{i=1}^n ([d_1]_i^T [d_1]_i)^+ (e_i^T (J_1 - F'(z))_1 [d_1]_i)^2 \\ &\leq \sum_{i=1}^n ([d_1]_i^T [d_1]_i)^+ \|e_i^T (J_1 - F'(z))_1\|^2 \| [d_1]_i \|^2 \leq \sum_{i=1}^n \|e_i^T (J_1 - F'(z))_1\|^2 \\ &= \|(J_1 - F'(z))_1\|_F^2 = \left\| \int_0^1 (F'(\bar{x} - (1-t)d_1) - F'(z))_1 dt \right\|_F^2 \\ &\leq \theta^2 (\|\bar{x} - z\| + \frac{1}{2} \|d_1\|)^2. \end{aligned}$$

Then (3.6) follows from (3.7) and (3.8). \square

LEMMA 3.2. Let F' satisfy (3.1) and (3.2), and let \bar{B} be generated by Algorithm 2.1. If $\bar{x} \in D$ and $\{\bar{x} - g_i, i = 2, \dots, p\} \subset D$, then for any $z \in D$,

$$(3.9) \quad \|\bar{B}_2 - F'(z)_2\|_F^2 \leq \|B_2 - F'(z)_2\|_F^2 - \frac{1}{\|s\|^2} \|(B_2 - F'(z)_2)s\|^2 + \gamma^2(\|\bar{x} - z\| + \|s\|)^2.$$

Proof. Substituting $F'(z)$ for A in (2.10), we obtain

$$(3.10) \quad \|\bar{B}_2 - F'(z)_2\|_F^2 \leq \|B_2 - F'(z)_2\|_F^2 - \frac{1}{\|s\|^2} \|(B_2 - F'(z)_2)s\|^2 + \sum_{i=2}^p \sum_{j \in c_i} s_j^+ s_j \|(J_i - F'(z))e_j\|^2.$$

It follows from (2.1) and (3.1) that

$$(3.11) \quad \begin{aligned} \sum_{i=2}^p \sum_{j \in c_i} s_j^+ s_j \|(J_i - F'(z))e_j\|^2 &\leq \sum_{i=2}^p \sum_{j \in c_i} \|(J_i - F'(z))e_j\|^2 \\ &= \sum_{i=2}^p \sum_{j \in c_i} \left\| \int_0^1 (F'(\bar{x} - g_i + t(g_i - g_{i-1})) - F'(z)) dt e_j \right\|^2 \\ &\leq \sum_{i=2}^p \sum_{j \in c_i} \gamma_j^2 \left(\int_0^1 (\|\bar{x} - z\| + (1-t)\|g_i\| + t\|g_{i-1}\|) dt \right)^2 \\ &\leq \sum_{i=2}^p \sum_{j \in c_i} \gamma_j^2 (\|\bar{x} - z\| + \|s\|)^2 = \gamma^2 (\|\bar{x} - z\| + \|s\|)^2. \end{aligned}$$

Thus, (3.9) follows from (3.10) and (3.11). \square

Let

$$d_i^k = \sum_{j \in c_i} s_j^k e_j,$$

and

$$g_i^k = \sum_{j=1}^i d_j^k, \quad i = 1, 2, \dots, p, \quad g_0^k = 0.$$

We have the following estimate for B^{k+1} .

THEOREM 3.3. Let F' satisfy (3.1) and (3.2), and let $\{x^k\}$ and $\{B^k\}$ be generated by Algorithm 2.1. If $\{x^j\}_{j=0}^{k+1} \subset D$ and $\{x^{j+1} - g_j^j, i = 1, 2, \dots, p\}_{j=0}^k \subset D$, then

$$(3.12) \quad \|B^{k+1} - F'(x^{k+1})\|_F \leq \|B^0 - F'(x^0)\|_F + 2\alpha \sum_{i=0}^k \|x^{i+1} - x^i\|.$$

Proof. Substituting z for \bar{x} in (3.6) and (3.9), we have

$$\|\bar{B}_1 - F'(\bar{x})_1\|_F^2 \leq \|B_1 - F'(\bar{x})_1\|_F^2 + \left(\frac{\theta}{2}\|d_1\|\right)^2$$

and

$$\|\bar{B}_2 - F'(\bar{x})_2\|_F^2 \leq \|B_2 - F'(\bar{x})_2\|_F^2 + (\gamma\|s\|)^2.$$

Therefore,

$$\begin{aligned} \|\bar{B} - F'(\bar{x})\|_F^2 &= \|\bar{B}_1 - F'(\bar{x})_1\|_F^2 + \|\bar{B}_2 - F'(\bar{x})_2\|_F^2 \\ &\leq \|B - F'(\bar{x})\|_F^2 + (\theta^2 + \gamma^2)\|s\|^2 = \|B - F'(\bar{x})\|_F^2 + \alpha^2\|s\|^2. \end{aligned}$$

Then,

$$(3.13) \quad \|\bar{B} - F'(\bar{x})\|_F \leq \|B - F'(\bar{x})\|_F + \alpha\|\bar{x} - x\| \leq \|B - F'(x)\|_F + 2\alpha\|\bar{x} - x\|.$$

Thus, (3.12) follows from (3.13). \square

Inequality (3.12) allows us to obtain the following Kantorovich-type theorem for Algorithm 2.1.

THEOREM 3.4. *Assume that F' satisfies (3.1) and (3.2). Also assume that $x^0 \in D$ and $B^0 \in L(R^n)$ satisfy*

$$\|B^0 - F'(x^0)\|_F \leq \delta, \quad \|(B^0)^{-1}\|_F \leq \beta, \quad \|(B^0)^{-1}F(x^0)\| \leq \eta$$

and

$$h \equiv \frac{\alpha\beta\eta}{(1 - 3\beta\delta)^2} \leq \frac{1}{10}, \quad \beta\delta < \frac{1}{3}.$$

If $\bar{S}(x^0, 2t^*) \equiv \{x: \|x - x^0\| \leq 2t^*\} \subset D$, where

$$t^* = \frac{1 - 3\beta\delta}{5\alpha\beta}(1 - (1 - 10h)^{1/2}),$$

then $\{x^k\}$, generated by Algorithm 2.1 without any global strategy, converges to x^* , the unique root of $F(x)$ in $\bar{S}(x^0, \bar{t}) \cap D$, where

$$\bar{t} = \frac{1 - \beta\delta}{\alpha\beta} \left(1 + \left(1 - \frac{2\alpha\beta\eta}{(1 - \beta\delta)^2} \right)^{1/2} \right).$$

Proof. Consider the scalar iteration

$$(3.14) \quad t_{k+1} - t_k = \frac{2\beta}{2 - \beta\delta} f(t_k), \quad t_0 = 0, \quad k = 0, 1, 2, \dots,$$

where

$$(3.15) \quad f(t) = \frac{5}{2}\alpha t^2 - \frac{1 - 3\beta\delta}{\beta}t + \frac{\eta}{\beta}.$$

It is easy to show that $\{t_k\}$ satisfies the difference equation

$$(3.16) \quad t_{k+1} - t_k = \frac{\beta}{1 - \varphi} [\alpha(t_k - t_{k-1}) + 2\alpha t_{k-1} + \delta](t_k - t_{k-1}),$$

where $\varphi = (3 + \beta\delta)/5 < 2/3$. Note that (3.16) is equivalent to

$$(3.17) \quad t_{k+1} - t_k = \frac{\beta}{1 - \varphi} [\alpha(t_k + t_{k-1}) + \delta](t_k - t_{k-1}).$$

From (3.17) we see that $\{t_k\}$ is a monotonically increasing sequence. Since t^* is the smaller of the two roots of (3.15), from (3.14) we have

$$\begin{aligned} t^* - t_{k+1} &= t^* - t_k - \frac{2\beta}{2 - \beta\delta} f(t_k) \\ &= \frac{2\beta}{2 - \beta\delta} \left[f(t^*) - f(t_k) - f'(t_k)(t^* - t_k) + \left(f'(t_k) + \frac{2 - \beta\delta}{2\beta} \right) (t^* - t_k) \right] \\ &= \frac{2\beta}{2 - \beta\delta} \left[\frac{5}{2}\alpha(t^* + t_k) + \frac{5}{2}\delta \right] (t^* - t_k). \end{aligned}$$

Noticing that $t^* > 0 = t_0$, by induction we have

$$t_k \leq t^*, \quad k = 0, 1, 2, \dots$$

Therefore, from (3.14),

$$\lim_{k \rightarrow \infty} t_k = t^*.$$

Now, by induction, we will prove that

$$(3.18) \quad \|x^{k+1} - x^k\| \leq t_{k+1} - t_k, \quad k = 1, 2, \dots,$$

$$(3.19) \quad \{x^k\} \subset \bar{S}(x^0, t^*),$$

$$(3.20) \quad \{x^{k+1} - g_i^k, i = 1, 2, \dots, p\} \subset \bar{S}(x^0, 2t^*),$$

and

$$(3.21) \quad \|(B^k)^{-1}\| \leq \frac{\beta}{1 - \varphi} \leq 3\beta, \quad k = 1, 2, \dots$$

For $k = 0$, we have

$$\|x^1 - x^0\| \leq \eta \leq \frac{2}{2 - \beta\delta} \eta = t_1 - t_0 \leq t^*.$$

Thus,

$$\|x^1 - g_i^0 - x^0\| \leq \|x^1 - x^0\| + \|g_i^0\| \leq 2\|x^1 - x^0\| \leq 2t^*.$$

Suppose (3.18)–(3.21) holds for $k = 0, 1, \dots, m-1$. Then,

$$\|x^m - x^0\| \leq \sum_{i=0}^{m-1} (t_{i+1} - t_i) = t_m \leq t^*.$$

Therefore, $x^m \in \bar{S}(x^0, t^*)$, and

$$\{x^m - g_i^{m-1}, i = 1, \dots, p\} \subset \bar{S}(x^0, 2t^*).$$

By Theorem 3.3,

$$\begin{aligned} & \|(B^0)^{-1}(B^m - B^0)\| \\ & \leq \|(B^0)^{-1}\|_F (\|B^m - F'(x^m)\|_F + \|F'(x^m) - F'(x^0)\|_F + \|F'(x^0) - B^0\|_F) \\ & \leq \beta \left(3\alpha \sum_{i=0}^{m-1} \|x^{i+1} - x^i\| + 2\delta \right) \leq \beta(3\alpha t^* + 2\delta) \leq \frac{3 + \beta\delta}{5} = \varphi. \end{aligned}$$

Thus, by Theorem 3.14 of Dennis and Schnabel [6, p. 45],

$$\|(B^m)^{-1}\| \leq \frac{\beta}{1 - \varphi} \leq 3\beta.$$

Therefore,

$$\begin{aligned} \|x^{m+1} - x^m\| & \leq \|(B^m)^{-1}\|_F \|F(x^m) - F(x^{m-1}) - B^{m-1}(x^m - x^{m-1})\| \\ & \leq \frac{\beta}{1 - \varphi} \left[\frac{\alpha}{2} \|x^m - x^{m-1}\| + 2\alpha \sum_{i=0}^{m-2} \|x^{i+1} - x^i\| + \delta \right] \|x^m - x^{m-1}\| \\ & \leq \frac{\beta}{1 - \varphi} [\alpha(t_m - t_{m-1}) + 2\alpha t_{m-1} + \delta](t_m - t_{m-1}) = t_{m+1} - t_m. \end{aligned}$$

This completes the induction step. From (3.18) it is easy to show that there is an $x^* \in D$ such that

$$\lim_{k \rightarrow \infty} x^k = x^*.$$

The uniqueness of x^* in $\bar{S}(x^0, \bar{t}) \cap D$ can be obtained from Theorem 12.6.4 of [9] by setting $A(x) \equiv B^0$. \square

4. Local Convergence Properties. To study the local convergence of our algorithm, we assume that $F: D \subset R^n \rightarrow R^n$ has the following property:

(4.1) There is an $x^* \in D$, such that $F(x^*) = 0$ and $F'(x^*)$ is nonsingular.

THEOREM 4.1. *Let F satisfy (4.1), and let F' satisfy (3.1) and (3.2). Also, let $\{x^k\}$ be generated by Algorithm 2.1 without any global strategy. Then there exist $\varepsilon, \delta > 0$ such that if $x^0 \in D$ and B^0 , a nonsingular $n \times n$ matrix, satisfy*

$$\|x^0 - x^*\| < \varepsilon, \quad \|B^0 - F'(x^*)\|_F \leq \delta,$$

then $\{x^k\}$ is well defined and converges q -superlinearly to x^* .

Proof. Notice that when ε and δ are small enough, we have that $h \leq 1/10$, $\beta\delta < 1/3$ and that $\bar{S}(x^0, 2t^*) \subset D$, where h, β and t^* are defined in Theorem 3.4. Therefore, by Theorem 3.4,

$$\{x^{k+1} - g_i^k, i = 1, 2, \dots, p\} \subset D.$$

Thus, substituting x^* for z in (3.6) and (3.9), we have

$$(4.2) \quad \begin{aligned} \|\bar{B}_1 - F'(x^*)\|_F^2 &\leq \|B_1 - F'(x^*)\|_F^2 - \frac{1}{\|s\|^2} \|(B_1 - F'(x^*)_1)s\|^2 \\ &\quad + \theta^2(\|\bar{x} - x^*\| + \|s\|)^2, \end{aligned}$$

and

$$(4.3) \quad \begin{aligned} \|\bar{B}_2 - F'(x^*)\|_F^2 &\leq \|B_2 - F'(x^*)\|_F^2 - \frac{1}{\|s\|^2} \|(B_2 - F'(x^*)_2)s\|^2 \\ &\quad + \gamma^2(\|\bar{x} - x^*\| + \|s\|)^2. \end{aligned}$$

Then,

$$\begin{aligned} \|\bar{B} - F'(x^*)\|_F^2 &= \|\bar{B}_1 - F'(x^*)\|_F^2 + \|\bar{B}_2 - F'(x^*)\|_F^2 \\ &\leq \|B - F'(x^*)\|_F^2 + \alpha^2(\|\bar{x} - x^*\| + \|s\|)^2 \\ &\leq \|B - F'(x^*)\|_F^2 + (3\alpha\sigma(x, \bar{x}))^2, \end{aligned}$$

where $\sigma(x, \bar{x}) = \max\{\|\bar{x} - x^*\|, \|x - x^*\|\}$. Therefore,

$$\|\bar{B} - F'(x^*)\|_F \leq \|B - F'(x^*)\|_F + 3\alpha\sigma(x, \bar{x}).$$

Thus, by Theorem 5.1 of Dennis and Moré [5], $\{x^k\}$ converges at least q -linearly to x^* .

By Theorem 3.1 of Dennis and Moré [5], to prove q -superlinear convergence, we need only prove that

$$(4.4) \quad \lim_{k \rightarrow \infty} \frac{\|(B^k - F'(x^*))s^k\|}{\|s^k\|} = 0.$$

Let $\bar{E} = \bar{B} - F'(x^*)$ and $E = B - F'(x^*)$. Then, it follows from (4.2) and (4.3) that

$$(4.5) \quad \|\bar{E}_1\|_F \leq \left(\|E_1\|_F^2 - \frac{\|E_1s\|^2}{\|s\|^2} \right)^{1/2} + 3\theta\sigma(x, \bar{x}),$$

and that

$$(4.6) \quad \|\bar{E}_2\|_F \leq \left(\|E_2\|_F^2 - \frac{\|E_2s\|^2}{\|s\|^2} \right)^{1/2} + 3\gamma\sigma(x, \bar{x}).$$

From (4.5) and (4.6), using the same argument for proving q -superlinear convergence of Broyden's algorithm (see Dennis and Moré [5]), we obtain

$$(4.7) \quad \lim_{k \rightarrow \infty} \frac{\|(B^k - F'(x^*))_1 s^k\|}{\|s^k\|} = 0,$$

and

$$(4.8) \quad \lim_{k \rightarrow \infty} \frac{\|(B^k - F'(x^*))_2 s^k\|}{\|s^k\|} = 0.$$

Notice that

$$\|(B^k - F'(x^*))s^k\| \leq \|(B^k - F'(x^*))_1 s^k\| + \|(B^k - F'(x^*))_2 s^k\|.$$

Thus, (4.4) follows from (4.7) and (4.8).

Acknowledgment. We would like to thank Professor Richard Tapia, Professor Walter Gautschi and the referees for helpful suggestions and corrections.

Department of Mathematical Sciences
Rice University
P.O. Box 1892
Houston, Texas 77251

Computer Science Department
Cornell University
Upson Hall
Ithaca, New York 14853

1. C. G. BROYDEN, "A class of methods for solving nonlinear simultaneous equations," *Math. Comp.*, v. 19, 1965, pp. 577-593.
2. C. G. BROYDEN, "The convergence of an algorithm for solving sparse nonlinear systems," *Math. Comp.*, v. 25, 1971, pp. 285-294.
3. T. F. COLEMAN & J. J. MORÉ, "Estimation of sparse Jacobian matrices and graph coloring problems," *SIAM J. Numer. Anal.*, v. 20, 1983, pp. 187-209.
4. A. R. CURTIS, M. J. D. POWELL & J. K. REID, "On the estimation of sparse Jacobian matrices," *J. Inst. Math. Appl.*, v. 13, 1974, pp. 117-119.
5. J. E. DENNIS, JR. & J. J. MORÉ, "Quasi-Newton methods, motivation and theory," *SIAM Rev.*, v. 19, 1977, pp. 46-89.
6. J. E. DENNIS, JR. & R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, N. J., 1983.
7. GUANGYE LI, *The Secant/Finite Difference Algorithm for Solving Sparse Nonlinear Systems of Equations*, Technical Report 86-1, Math Sciences Dept., Rice Univ., 1986.
8. E. MARWIL, "Convergence results for Schubert's method for solving sparse nonlinear equations," *SIAM J. Numer. Anal.*, v. 16, 1979, pp. 588-604.
9. J. M. ORTEGA & W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
10. J. K. REID, *Least Squares Solution of Sparse Systems of Non-linear Equations by a Modified Marquardt Algorithm*, Proceedings of the NATO Conf. at Cambridge, July 1972, North-Holland, Amsterdam, pp. 437-445.
11. L. K. SCHUBERT, "Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian," *Math. Comp.*, v. 24, 1970, pp. 27-30.