ON A GENERAL CONVERGENCE FOR BROYDEN LIKE UPDATE METHOD

RABINDRANATH SEN, RINI CHATTOPADHYAY and TRIPTI SAHA

Department of Applied Mathematics
Calcutta University
92, A.P.C. Road
Calcutta - 700009
India

(Received December 7, 1987 and in revised form November 20, 1990)

ABSTRACT. The role of Broyden's method as a powerful quasi-Newton method for solving unconstrained optimization problems or a system of nonlinear algebraic equations is well known. We offer here a general convergence criterion for a method akin to Broyden's method in \( \mathbb{R}^n \). The approach is different from those of other convergence proofs which are available only for the direct prediction methods.

KEY WORDS AND PHRASES. Unconstrained optimization, Broyden's method, partial ordering in \( \mathbb{R}^n \), M-matrix.

1980 AMS SUBJECT CLASSIFICATION CODE. 65K.

1. INTRODUCTION.

Let \( \mathcal{V}: \mathbb{D} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^1 \) be an \( \mathcal{F} \)-differentiable functional having an optimum \( x^* \in \text{int} (\mathbb{D}) \). Then the vector \( x^* \) at which the optimum of \( \mathcal{V} \) is realized, satisfies the equation

\[
\nabla \mathcal{V}(x) = \mathcal{F}(x) = 0
\]

In the above, \( \nabla = (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n})^T \). Our concern in this paper is iterative methods for the solution of simultaneous non-linear equations

\[
F(x) = 0; \ F: \mathbb{R}^m \rightarrow \mathbb{R}^m \quad (1.1)
\]

in the case when the complete computation of \( \mathcal{F}' \) is infeasible.

In the case, \( F = \mathcal{F} \), solving (1.1) means in effect finding the minimizer of \( \mathcal{V} \).

The algorithm under consideration takes the form

\[
x_{n+1} = x_n - \gamma_n H_n F(x_n)
\]

where \( H_n \) is generated by the method in such a way that the quasi-Newton equation

\[
H_{n+1} (F(x_{n+1}) - F(x_n)) = x_{n+1} - x_n \quad (1.3)
\]

is satisfied at each step.
The step length $\gamma_n$ is chosen to promote convergence. By analogy with the DFP-method (Davidson [1], Fletcher and Powell [2]) for unconstrained optimizations and by considering what is desirable when $f$ is linear, Broyden [3] in 1965 suggested an algorithm by which $H_{n+1}$ is obtained from $H_n$ by means of a rank-one update.

In the case of minimizing problems, Dixon [4] called a method perfect if $\gamma_n \to \gamma_n^*$ is obtained through line searches and a direct prediction method if $\gamma_n = 1$. The iterative procedures (1.2) and (1.3) may be described as follows:

Choose non-singular $H_0 \in R^{m \times m}$ and also $x_0 \in R^m$.

For $n = 0, 1, 2, \ldots$ let $s_n = -H_n F(x_n)$ (1.4) $\gamma_n$ being chosen such that

$$\frac{||F(x_{n+1})||}{||F(x_n)||} < \frac{||F(x_{n-1})||}{||F(x_{n-1})||}$$

(1.5) $x_{n+1} = x_n + s_n$

(1.6) $y_n = F(x_{n+1}) - F(x_n)$

If $y_n = 0$ then

$$H_{n+1} = H_n$$

(1.7) and choose $v_n \in R^m$ such that $v_n^T y_n = 1$ and $v_n^T s_n \neq 0$.

Let $H_{n+1} = H_n + (s_n - H_n y_n) v_n^T v_n$ (1.8)

Broyden's [3] method (sometimes called his first or good method) results from choosing $v_n = H_n^T s_n / H_n y_n$ in (1.8) and is defined for $y_n \neq 0$ only so long as $s_n^T H_n y_n \neq 0$.

Broyden's second or bad method results from choosing $v_n = y_n / y_n^T y_n$ where $y_n \neq 0$ and is defined so long as $y_n^T H_n y_n \neq 0$.

The convergence results that are available to date are proved for the direct prediction method [5]. Broyden has shown that his (first) method converges locally at least linearly on nonlinear problems and at least $R$-Superlinearly on linear problem [6].

Later Broyden et al [7] showed that both Broyden's good and bad methods converge locally at least $Q$-superlinearly.

More and Trangstein [8] subsequently proved that 'locally' could be replaced by "globally" when a modified form of Broyden's method is applied to linear systems of equations. On the other hand, Gay showed in 1979 [5] that Broyden's good and bad methods enjoy a finite termination property when applied to linear systems with a non-singular matrix. He has also proved that Broyden's good method enjoys local $2m$-step $Q$-quadratic convergence on non-linear systems.

Recently, Dennis and Walker [9] have made generalizations of their results [7], [10] and have put forward convergence theorems for least-change secant update methods. Decker et al, have considered Broyden's method for a class of problems having singular Jacobian at the root [11].

Our concern is to consider the method (1.2) which is not necessarily a direct prediction method. Our method is not perfect because in the case of minimization problems exact line searches have not been performed. However, the scalars do reduce $||F(x)||$ at each step. It may however be noted that perfect methods are not efficient
Our method can thus be viewed as an intermediate between direct prediction and perfect methods.

We have called our method Broyden-like because although we have used Broyden's good method we have taken $B_0$ to be an M-matrix [12] unlike Broyden [3]. We have used componentwise partial ordering in $\mathbb{R}^m$ to prove monotone convergence of the sequence $x_n$. We have asserted that under certain conditions, $H_n$ can be taken as non-negative matrices. In our analysis, $F$ is taken as an isotone operator [12]. Operators which are monotonically decomposable (MDO) [13] can also be brought within the scope of our convergence theorem. A local linear convergence is achieved. Experiments with numerical problems are also encouraging. Even where Newton's method has diverged, our method converges in a small number of iterations.

Section 2 contains mathematical preliminaries. Section 3 contains convergence results. Section 4 gives the algorithm. Numerical examples are presented in section 5 while we incorporate some 'discussion' in section 6.

2. MATHEMATICAL PRELIMINARIES.

DEFINITION 2.1. The componentwise partial ordering in $\mathbb{R}^m$ is defined as follows:

For $x, y \in \mathbb{R}^m$, $x = (x_1, x_2, \ldots, x_m)^T$, $y = (y_1, y_2, \ldots, y_m)^T$, $x \succ y$ if and only if

$x_i \geq y_i$, $i = 1, 2, \ldots, m$.

$x \succ y \iff x \succeq y$ and $x \neq y$.

DEFINITION 2.2. We define for any $x, y \in \mathbb{R}^m$, such that $x \preceq y$, the order interval $[x,y] = \{ u \in \mathbb{R}^m / x \preceq u \preceq y \}$.

DEFINITION 2.3. A mapping $F : D \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^m$ is isotone (antitone) [12] on $D$ if $Fx \preceq Fy$ ($Fx \succ Fy$) whenever $x \preceq y$, $x, y \in D$.

DEFINITION 2.4. We denote by $L(\mathbb{R}^m)$ the space of $m \times m$ matrices.

We introduce a partial ordering in $L(\mathbb{R}^m)$ which is compatible with the componentwise partial ordering in $\mathbb{R}^m$.

DEFINITION 2.5. A real $m \times m$ matrix $(a_{ij})$ with $a_{ij} \leq 0$, for all $i \neq j$ is defined to be an $M$-matrix [12] if $A$ is non-singular and $A^{-1} \succ 0$.

In what follows, $H_n = B_n^{-1}$, so that $B_n$ is a replacement for the Jacobian $J(x_n)$, and thus $B_n$ satisfies the quasi-Newton equation

$$B_n y_{n-1} = y_{n-1}, \quad n = 1, 2, \ldots$$

(2.1)

According to Broyden's good method, the update formula for $B_{n+1}$ is given by

$$B_{n+1} = B_n - \frac{(B_n y_n - y_n) s_n}{s_n^T s_n}$$

(2.2)

In what follows we denote $s_n = (s_{i1}^{(n)})$, $y_n = (y_{i1}^{(n)})$, $B_n = (b_{ij}^{(n)})$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$.

LEMMA 2.1. $(I - \frac{s_n}{s_n^T})$ is a square matrix whose diagonal elements are positive and off-diagonal elements are non-positive provided $s_{i1}^{(n)} > 0$ and $s_{i2}^{(n)} > 0$ for at least one $i$. 

---
The proof is trivial.

**Lemma 2.2.** Let the following conditions be fulfilled:

(i) \( B_n = \{b_{ij}^{(n)}\}, b_{ij}^{(n)} < 0 \) for all \( i, j \), \( i = j \), \( b_{ii}^{(n)} > 0 \)

(ii) \( B_n \) is a strictly diagonally dominant matrix

(iii) \( \{x_n\} \) is a monotonic increasing sequence in the order sense.

(iv) \( \sum_{k} b_{ik}^{(n)} (s_{k}^{(n)} - s_{k}^{(n-1)}) = (y_{k}^{(n)} - y_{k}^{(n-1)}) s_{k}^{(n)} \)

(v) \( \sum_{i} b_{ij}^{(n)} \) is small compared to \( \left| b_{ij}^{(n)} \right| \)

Then \( B_{n+1} \) is an M-matrix.

**Proof.** It follows from conditions (i) and (ii) that \( B_n \) is an M-matrix [12].

Since \( B_n \) satisfies the Quasi-Newton equation (2.1)

\[
(B_n - y_n)S_n = \left[ (s_n - s_{n-1}) - (y_n - y_{n-1}) \right] S_n^T
\]

and the \( (i,j) \)th element of \( B_{n+1} \) is given by

\[
b_{ij}^{(n+1)} = \sum_{k} b_{ik}^{(n)} (s_{k}^{(n)} - s_{k}^{(n-1)}) - (y_{k}^{(n)} - y_{k}^{(n-1)}) s_{k}^{(n)}
\]

since \( x_n \) is a monotonic increasing sequence \( s_n = x_{n+1} - x_n > 0 \). In the neighborhood of the solution \( x^* \), \( s^{(n)} \) will be small and \( s^{(n)} - s^{(n-1)} \) will still be smaller in magnitude. \( s_{j}^{(n)} / \sum_{j} s_{j}^{(n)} \) will also have at most a finite magnitude.

If \( b_{ij}^{(n)} \neq 0, i \neq j \), the signs of \( b_{ij}^{(n)} \) will determine the signs of \( b_{ij}^{(n+1)} \).

Therefore the diagonal elements of \( b_{ij}^{(n+1)} \) will be strictly positive and the off-diagonal elements strictly negative. Since \( B_n \) is a strictly diagonally dominant matrix the strict diagonal dominance property of \( B_{n+1} \) will be maintained by virtue of condition (v).

Let us denote

\[
\sum_{k} b_{ik}^{(n)} (s_{k}^{(n)} - s_{k}^{(n-1)}) - (y_{k}^{(n)} - y_{k}^{(n-1)}) s_{k}^{(n)}
\]

by \( c_{ij}^{(n)} \).

Condition (v) implies that

\[
\sum_{j} \left| b_{ij}^{(n+1)} \right| < \left| b_{ii}^{(n+1)} \right|\]

and since \( \left| b_{ij}^{(n+1)} \right| \) is large compared to \( \left| c_{ij}^{(n)} \right| \), (2.4) implies that

\[
\sum_{i,j} \left| b_{ij}^{(n+1)} \right| < \left| b_{ii}^{(n+1)} \right|
\]
REMARK 2.1. If however \( B_{n+1} \) ceases to be an \( M \)-matrix, we can without violating the constraint (2.1), pre-multiply the first term in the expression for \( B_{n+1} \) by a strictly positive diagonal matrix \( P_n \) of the same order of \( B_n \), such that \( P_n B_n (I - \frac{s_n s_n^T}{s_n^T}) = \frac{s_n s_n^T}{s_n^T} \) becomes the predominant term in the expression for \( B_{n+1} \) as given below.

\[
\tilde{B}_{n+1} = P_n B_n (I - \frac{s_n s_n^T}{s_n^T}) + \frac{y_n y_n^T}{s_n^T}.
\]  

(2.5)

Thus, if \( B_n \) is already a strictly diagonally dominant matrix with \( b^{(n)}_{ij} < 0 \), \( i \neq j \) and \( b^{(n)}_{ii} > 0 \), \( B_{n+1} \) will also be a strictly diagonally dominant matrix with \( b^{(n+1)}_{ij} < 0 \), \( i \neq j \), and \( b^{(n+1)}_{ii} > 0 \). Therefore, if \( B_n \) is an \( M \)-matrix we can always construct \( B_{n+1}(= \tilde{B}_{n+1}) \) as an \( M \)-matrix without affecting the condition \( B_n s_n = y_n \).

Writing \( B_n^{-1} = H_n \) and \( B_{n+1}^{-1} = H_{n+1} \), we obtain from (2.5)

\[
H_{n+1} = [P_n B_n (I - \frac{s_n s_n^T}{s_n^T}) - \frac{s_n s_n^T}{s_n^T}]^{-1}
\]

\[
= [I - \frac{s_n s_n^T}{s_n^T} + \frac{b^{(n)}_{ii} y_n y_n^T}{s_n^T}]^{-1} - \frac{s_n s_n^T}{s_n^T}
\]

which is approximately equal to

\[
[I + (s_n - H_n^{-1} y_n - n)] H_n^{-1}.
\]

Writing \( s_n = \tilde{B}_{n+1}^{-1} y_n = H_{n+1} y_n \), which is approximately equal to \( H_n^{-1} y_n \), \( H_{n+1} \) can be taken as

\[
H_{n+1} = H_n^{-1} + (s_n - H_n^{-1} y_n) \frac{y_n y_n^T}{s_n^T} H_n^{-1} H_n^{-1} y_n.
\]

(2.6)

Therefore, \( H_{n+1} \) as expressed by (2.6) satisfies the Quasi-Newton equation

\[
H_{n+1} y_n = s_n.
\]

**LEMMA 2.3.** Let the diagonal elements of \( b^{(n)}_{ij} \) of the strictly diagonally dominant matrix \( B_n \) satisfy the condition \( |b^{(n)}_{ii}| < M \), for all \( i \), where \( M(>0) \) is a finite constant. Then

\[
\frac{1}{|\lambda^{(n)}_i|} > \frac{1}{2n}, \quad \lambda^{(n)}_i \quad \text{are the eigenvalues of } B_n.
\]

**PROOF.** By the Gerschgorin theorem,

\[
|\lambda^{(n)}_i - b^{(n)}_{ii}| < \sum_{j \neq i} |b^{(n)}_{ij}|
\]

By strict diagonal dominance of \( B_n \),

\[
|\lambda^{(n)}_i| < 2|b^{(n)}_{ii}| < 2M.
\]
Therefore,
\[ \frac{1}{|\lambda_1^{(n)}|} > \frac{1}{MN} > 0. \]

The above lemma ensures that the lower bound of the eigenvalues of \( H_n = B^{-1}_n \) are strictly positive for all \( n \) and hence \( H_n \) remains nonsingular at each iteration. The stability of the iteration process is therefore guaranteed.

In what follows \( ||.|| \) denotes an arbitrary norm in \( \mathbb{R}^m \) and the operator norm is induced by the particular vector norm.

3.1. CONVERGENCE.

**Theorem 3.1.** Let the following conditions be fulfilled:

(i) \( F: D \subset \mathbb{R}^m \rightarrow \mathbb{R}^m \) is Fréchet differentiable on a convex set \( D \subset D_0 \).

(ii) \( D_0 \) includes the null element and \( x_0, y_0 \in D \) such that if \( x < y \) then the order interval \( <x, y>D_0 \).

(iii) \( F(y) < 0 \), for all \( x \in D_0 \) and \( F \) is isotone i.e. if \( x < y \), \( F(x) < F(y) \) for all \( x, y \in D_0 \).

(iv) \( x_0 \in D_0 \) is an initial approximation to the solution and \( x_0 \geq 0 \) and \( F(x_0) < 0 \).

(v) The operator \( G = I - \gamma \cdot H \cdot F \) is such that \( GD \subset D_0 \).

(vi) The operator \( B = (b_{ij}) \) is a strictly diagonally dominant matrix with \( b_{ii}^{(o)} < 0 \), for all \( i, j, i \neq j \), and \( b_{ii}^{(o)} > 0 \).

(vii) For \( B_n = b_{ij}^{(n)} \)

\[
\sum_{k} \frac{b_{ij}^{(n)}(s_k^{(n)} - s_k^{(n-1)}) - (y_k^{(n)} - y_k^{(n-1)}) s_k^{(n)}}{s_k^{(n)} - s_1^{(n)}} - \sum_{i} \frac{b_{ij}^{(n)}(s_i^{(n)} - s_i^{(n-1)}) - (y_i^{(n)} - y_i^{(n-1)}) s_i^{(n)}}{s_i^{(n)} - s_1^{(n)}} \]

is small compared to \( |b_{ij}^{(n)}|, n = 0,1,2,... \)

(viii) \( \sum_{j \neq i} b_{ij}^{(n)} < b_{ii}^{(n)} + \frac{\sum_{k} b_{ik}^{(n)}(s_k^{(n)} - s_k^{(n-1)}) - (y_k^{(n)} - y_k^{(n-1)}) s_k^{(n)}}{s_k^{(n)} - s_1^{(n)}} - \frac{\sum_{i} b_{ij}^{(n)}(s_i^{(n)} - s_i^{(n-1)}) - (y_i^{(n)} - y_i^{(n-1)}) s_i^{(n)}}{s_i^{(n)} - s_1^{(n)}} \),

\( n = 0,1,2,... \)

(ix) \( |b_{ii}^{(n)}| < (K > 0), n = 0,1,2,... \)

(x) The scalars \( \gamma_n (> \alpha > 0) \) are to be chosen such that

(a) \( \gamma_{n-1} \cdot H_{n-1} \cdot F(x_0) < \gamma_n \cdot H_n \cdot F(x_0) < 0, n > 1 \)

(b) \( \sup \{ I - \gamma_n \cdot H \cdot F'(x) \} < \sup \{ I - \gamma_{n-1} \cdot H \cdot F'(x) \}, x \in [x_{n-1}, x] \)

(xi) \( F'(x) \) is Fréchet differentiable for all \( x \in D_0 \).
Then starting from $x_0 \neq 0$, an initial approximation to the solution of $F(x) = 0$ is the sequence $(x_n)$ of Broyden-like approximations defined by

$$x_{n+1} = x_n - \gamma_n F(x_n), \quad n = 0, 1, \ldots$$

(3.1)

which converges to a solution of the equation $F(x) = 0$.

If in particular, it is possible to find a positive convergent matrix $A$, i.e., $A^n \to 0$ as $n \to \infty$, then the error at the $n$th stage is given by

$$||x^n - x^*|| \leq ||A^{n-1}(x_1 - x_0)||$$

(3.2)

PROOF. Conditions (iii), (iv) and (xa) yield

$$x_1 = x_o - \gamma H F(x_o) > x_0 \quad (3.3)$$

By (v), $0 < x_1 \in D_o \subset D$. Hence $Gx_1 \in D_o$.

$B_0$ is an $M$-matrix [12] by (vi). Hence $H_0 = B_0^{-1} > 0$, $s_0 = x_1 - x_0 > 0$.

Because of conditions (vi), (vii) and (viii) and because $s_0 > 0$, it follows from (2.2) that $B_1$ is an $M$-matrix. Therefore, $H_1 = B_1^{-1} > 0$.

Isotonicity of $F$, together with condition (xa) yield,

$$0 > \gamma_1 H_1 F(x_1) \geq \gamma H F(x_0).$$

Hence $x_2 > x_1 > 0$. Thus $x_1 < x_2 = x_1 - \gamma H F(x_o) = Gx_1 \in D_o$.

Let us assume by way of induction that $0 < x_k \in D_o$, $k = 1, 2, \ldots, n$, and $B_k$ is an $M$-matrix, $k = 1, 2, \ldots, n$, with $b_{ij}^{(k)} < 0$, $i \neq j$.

Since $B_n$ is an $M$-matrix, $H_n > 0$.

By conditions (iii) and (xa) $\gamma_n H_n F(x_n) > \gamma H F(x_o)$

$$\gamma H F(x_o)$$

Therefore, $x_n < x_{n+1} < x_n - \gamma_n H F(x_o)$

$$< x_n - \gamma H F(x_o)$$

$$= Gx_n \in D_o$$

Using the fact that $B_n$ is an $M$-matrix with $b_{ij}^{(n)} < 0$, $i \neq j$, $x_n < x_{n+1}$, and the conditions (vii) and (viii), we can conclude from Lemma (2.2) that $B_{n+1}$ is an $M$-matrix with $b_{ij}^{(n+1)} < 0$, $i \neq j$.

Therefore the induction is completed.

Thus, $(x_k)$ is a monotonic increasing sequence in the order sense and $x_k \in D_o$, for all $k$.

Now, $0 < x_2 - x_1 = x_1 - x_0 - \gamma_1 H_1 [F(x_1) - F(x_o)]$

$$- [\gamma_1 H_1 F(x_o) - \gamma H F(x_o)]$$

$$< x_1 - x_0 - \gamma_1 H_1 [F(x_1) - F(x_o)].$$
D being a convex set and using the mean value theorem in $\mathbb{R}^m$ we obtain from (3.5)

$$\frac{x_2 - x_1}{\gamma_1} < \int \frac{1}{\gamma_1} \frac{\partial}{\partial t} \frac{1}{\gamma_1} \frac{F'(x_0 + (t_1 - t_1^*)x_2)}{\gamma_1} dt^* \quad 0 < t < 1 \tag{3.6}$$

Here $t^*$ stands for differentiation in the Frechet sense.

Again since $F'(x_0 + t(x_1 - x_0))$, $0 < t < 1$ is $C$ differentiable, $F'(x_0 + t(x_1 - x_0))$ is semi-continuous [12].

Therefore, the operator $G_1(t) = F'(x_0 + t(x_1 - x_0))$ is continuous in $t [0,1]$. We thus have the following from (3.6)

$$0 < x_2 - x_1 < \sup_t \frac{1}{\gamma_1} \frac{F'(x_0 + t(x_1 - x_0))}{\gamma_1} (x_1 - x_0)$$

assuming the supremum is attained at $t = t^*$, $0 < t < 1$.

By (xb), (3.7) further simplifies to

$$0 < x_2 - x_1 < \frac{1}{\gamma_1} \frac{F'(x_0)}{\gamma_1} (x_1 - x_0) \tag{3.8}$$

Arguing analogously as before,

$$x_{n+1} - x_n < \frac{1}{\gamma_1} \frac{\partial}{\partial t} \frac{1}{\gamma_1} \frac{F'(x_{n-1} + t(x_n - x_{n-1}))}{\gamma_1} (x_n - x_{n-1})$$

assuming that the supremum is attained at $t = \tilde{t}$.

Condition (xb) further reduces (3.9) to

$$x_{n+1} - x_n < \frac{1}{\gamma_1} \frac{F'(x_{n-1})}{\gamma_1} (x_n - x_{n-1}) \tag{3.10}$$

Hence $x_{n+p} - x_n < \sum_{k=n}^{n+p-1} \frac{1}{\gamma_1} \frac{F'(x_{n-1})}{\gamma_1} (x_1 - x_0)$

Since $[1 - \gamma_1 F'(x_0)]^n (x_1 - x_0) + 0$ as $n \to \infty$, it follows from (xii) that $\{x_n\}$ is a Cauchy sequence and the space is complete,

$$\lim_{n \to \infty} x_n = \mathbb{R}^m \tag{3.12}$$

Using convergence of $\{x_n\}$ it follows from (3.12) that

$$\lim_{n \to \infty} \gamma_1 H F(x_n) = 0. \tag{3.13}$$

Now, $F(x_n) = \frac{1}{\gamma_1} B_n (x_{n+1} - x_n)$.

Therefore using (ix), (x) and the strict diagonal dominance of $B_n$

$$||F(x_n)|| < 2k^\alpha \quad ||x_{n+1} - x_n|| \to 0 \quad \text{as} \quad n \to \infty. \tag{3.14}$$

Continuity of $F$ yields that

$$\star \quad F(x^*) = 0.$$ 

Hence $x^*$ is a solution of the equation $F(x) = 0$.

Further if $1 - \gamma_1 F'(x_0) < A$, (3.11) reduces to
\[ x_{n+p} - x_n = \sum_{k=n}^{n+p-1} \Delta^k (x_1 - x_o) \quad (3.15) \]

The error (3.2) follows from (3.15) by making \( p \to \infty \) and then taking \( \| \cdot \|_p \).

In the next theorem we provide a bound for the inverse Jacobian approximation.

**Theorem 3.2.** Let in addition to the conditions of theorem (3.1), the following conditions be satisfied:

(a) \( \| F'(x) - F'(y) \| < \gamma' \| x - y \| \), for all \( x, y \in D_o \)

(b) \( [F'(x_o)]^{-1} \) exists and \( \| [F'(x_o)]^{-1} \| < \beta' \)

(c) \( \| \sum_{n=0}^{N-1} (I - \gamma_{ho} F'(x_o)) \| (x_1 - x_o) \| < \beta' \gamma' \)

Then
\[ \| [F'(x_n)]^{-1} - \gamma_n H_n \| < \frac{\| (I - \gamma_{ho} F'(x_o)) \|}{1 - \beta' \gamma'} \frac{\| \sum_{n=0}^{N-1} (I - \gamma_{ho} F'(x_o)) \| (x_1 - x_o) \|}{1 - \beta' \gamma'} \]

**Proof.** Following Rheinboldt and Ortega [12] we show that \( [F(x)]^{-1} \) exists at all the iteration points.

Let \( D_1 \) denote the closed sphere with \( x_o \) as the center and \( \beta' \gamma' \) as the radius.

Let \( D_1 = \cap \{ x_o, (\beta' \gamma')^{-1} \} \cap D_o \)

Then, for \( x \in D_1 \) we have
\[ \| F'(x) - F'(x_o) \| < \gamma \| x - x_o \| < 1/\beta' \quad (3.17) \]

Now,
\[ \| I - [F'(x_o)]^{-1} F'(x_1) \| \]
\[ \| [F'(x_o)]^{-1} [F'(x_o) - F'(x_1)] \| < \beta' \frac{1}{\beta'} = 1 \quad (3.18) \]

Therefore \( [F'(x_o)]^{-1} F'(x_1) \) has an inverse and hence for \( x \in D_1 \), \( F'(x) \) has an inverse.

Moreover, for \( x \in D_1 \), using the Neumann Lemma we get
\[ [F'(x)]^{-1} = \sum_{n=0}^{N-1} [F'(x_o)]^{-1} [F'(x_o) - F'(x)]^{-1} \]

Therefore,
\[ \| \sum_{n=0}^{N-1} ([F'(x_o)]^{-1} [F'(x_o) - F'(x)]^{-1} \| \]
\[ \| \gamma_n \| (\beta' \gamma') \| x - x_o \| \| (x_1 - x_o) \| \]
\[ \leq \frac{1}{1 - \beta' \gamma'} \| x - x_o \| \]

since \( \beta' \gamma' \| x - x_o \| < 1 \)

Therefore for \( x \in D_1 \)
\[ \| F'(x) \| < \frac{\beta'}{1 - \beta' \gamma'} \| x - x_o \| \]

\( \{ x_n \} \) being a monotonic increasing sequence it follows from (3.10) and condition (c) of Theorem (3.2),
\[ \| x_n - x_o \| < \| \sum_{k=0}^{N-1} (I - \gamma_{ho} F(x_o)) \| (x_1 - x_o) \| < \beta' \gamma' \quad (3.22) \]

Hence \( x_n \in \cap \{ x_o, (\beta' \gamma') \} \cap D_o = D_1 \)

Utilizing condition (xb) and (3.21) we get
Using (3.10) and the monotonic increasing property of \( \{x_n\} \) we further conclude that

\[
\left| \left| \left( F'(x) \right)^{-1} - y_n H_n \right| \right| < \frac{b'}{1 - b' r} \left| \left| I - y H F'(x, o) \right| \right| \left| \left| F'(x) \right| \right|^{-1}
\]

\[
(3.23)
\]

**THEOREM 3.3.** If condition (xh) of Theorem 3.1 is replaced by the following condition:

\[
(I - y H F'(x, n-1)) \succ a' \sup_{x \in D} (I - y H F'(x, n))
\]

and all other conditions of theorem 3.1 are fulfilled then the superlinear convergence of the sequence \( \{x_n\} \) is ensured.

**PROOF.**

\[
\lim_{n \to \infty} (I - y H F'(x, n)) \preceq \sup_{x \in D} (I - y H F'(x, n))
\]

\[
< \frac{1}{a' n} (I - y H F'(x, o)), (a' > 1)
\]

\[
\Rightarrow 0 \text{ as } n \to \infty.
\]

\[
(3.25)
\]

If \( \lim_{n \to \infty} x_n = x^* \) then

\[
0 \prec x^* - x_{n-1} = x^* - x_n - y H [F(x^*) - F(x_n)]
\]

\[
= \int_0^1 [I - y H F'(x, n + t(x^* - x_n))] (x^* - x_n) dt.
\]

By arguments analogous to theorem 3.1

\[
\left| \left| x^* - x_{n+1} \right| \right| \preceq \sup_{t \in [0,1]} \left| \left| I - y H F'(x, n + t(x^* - x_n)) \right| \right| \left| \left| x - x_n \right| \right|.
\]

Therefore,

\[
\left| \left| x^* - x_{n+1} \right| \right| \preceq \sup_{t \in [0,1]} \left| \left| I - y H F'(x, n + t(x^* - x_n)) \right| \right|,.
\]

The continuity of \( F'(x) \) for \( x \in D \) and relation (3.25) yield,

\[
\lim_{n \to \infty} \left| \left| x^* - x_{n+1} \right| \right| < \lim_{n \to \infty} \left| \left| x^* - x_n \right| \right| = 0
\]

\[
(3.26)
\]

This proves superlinear convergence of \( \{x_n\} \).

**REMARK 3.1.** It may be noted that conditions (vii) and (viii) of theorem 3.1 are required only to prove that \( B_{n+1} \) is an M-matrix provided \( B_n \) is so. The question may be raised as to how one can know these conditions in advance. From computational experience one can say that such conditions are usually satisfied. If that is not so, we have indicated in remark 2.1 how \( B_{n+1} \) can be made an M-matrix when \( B_n \) is so.

4. **ALGORITHM.**

**Step 1.** Find \( D_o \) such that \( x < x^* \) in which \( F(x) \) is isotone.
Step 2. Choose $x_0 \in D$ and $x_0 > 0$ s.t. $F(x_0) < 0$

Step 3. Choose $H_0 > 0$, $\gamma_0 = 1$ and compute $x_1$, $k = 0$.

Step 4. Compute $A_0 = I - \gamma_0 H_0 F'(x_0)$.

If $(I - \gamma_0 H_0 F'(x_0))^n (x_1 - x_n) + 0$ for some $n > n_0$ go to step 5, otherwise go to step 3.

Step 5. Compute $s_k = \gamma H_0 F'(x_k)$, $x_{k+1} = x_k + s_k$ and $F(x_{k+1})$

Step 6. If $F(x_{k+1}) = 0(10^{-4})$ stop, otherwise go to step 7.

Step 7. Compute $y_k = F(x_{k+1}) - F(x_k)$, $s_k^T H_k y_k$

If $s_k^T H_k y_k = 0$, $H_{k+1} = H_k$ go to step 8.

Otherwise $H_{k+1} = H_k + (s_k - H_k y_k)[\frac{H_k s_k y_k}{s_k^T H_k y_k}]^T$

Step 8. Compute $y_{k+1}$ s.t. $y_{k+1}^T H_{k+1} F(x_k) = y_k^T H_k F(x_k)$

Step 9. If $I - \gamma H_k F'(x_k) \sup_{x \in \langle x_k, x_{k+1} \rangle} [I - \gamma_{k+1} H_{k+1} F'(x)]$ go to step 10, otherwise go to step 8.

Step 10. If $k = k+1$, go to step 5.

REMARK 4.1. (i) The implementation of the conditions (xa) and (xb) for the determination of the scalar $\gamma_n$ can be done by a computer. (ii) The matrix $P_n$ in note (2.1) is arbitrary except that the elements of $P_n$ are greater than $p_n$ ($>0$). This must lead to some arbitrariness in $H_{n+1}$. In that case the choice of $\gamma_{n+1}$ covertly depends on $P_n$ in addition to the conditions of (xa) and (xb). $H_{n+1}$ so generated is however a solution of the Quasi-Newton equation. (iii) The total number of multiplications and divisions in finding $x_{k+1}$, $H_{k+1}$ and $y_{k+1}$ respectively is $(m^2 + 2m)$, $2m^2 + m$, $2m^2 + 3m$, i.e. $5m^2 + 5m$. (iv) The number of function evaluations is $m + 2m$.

5.1 NUMERICAL EXAMPLE.

We take an example [14] in which every equation is linear except for the last equation which is highly non-linear. We choose

$$F(x) = [f_i(x)]^T, \quad i = 1, 2, \ldots N-1$$

where

$$f_i(x) = -(N+1) + 2x_i + \sum_{j=1}^{N} x_j^i, \quad i = 1, 2 \ldots N-1$$

and

$$f_N(x) = -1 + \Pi_{j=1}^{N} x_j^i.$$

The problem was run for $N = 5, 10$ and $30$. We take $x_0^i = 0.5$ so that $F(x_0) < 0$. Incidentally $F(x)$ is isotone.

Define $D_0$, the rectangular parallelepiped given by $0.5 \leq x_i \leq 1.0, \quad i = 1, 2, \ldots N$, and choose $\gamma_0 = 1$. $H_0 = (h_{ij}^0)$ where $h_{ij}^0 = 0.1, \quad i \neq N, \quad h_{NN}^0 = 0.5, \quad h_{ij}^0 = .01$.

$A = [I - \gamma H_0 F'(x_0)]$ is a convergent matrix.

We summarize our computational experience in table I. The computations were performed on a Burroughs computer at the R.C.C. Calcutta using a FORTRAN IV language. We solved the problem and in each case the exact solution was obtained. In
Table 2 provides a comparison of our results with those of Newton's method and Brown's method. For computational results of Newton's method and Brown's method see [14].

<table>
<thead>
<tr>
<th>Dimension(N)</th>
<th>No. of Iterations(n)</th>
<th>x</th>
<th>F(x)</th>
<th>CPU Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>(1,1,1)^T</td>
<td>0</td>
<td>3.637 sec</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>(1,1,...,1)^T</td>
<td>0</td>
<td>4.462 sec</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
<td>(1,1,...,1)^T</td>
<td>0</td>
<td>8.548 sec</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Dimension(N)</th>
<th>Newton's method</th>
<th>Brown's Method</th>
<th>Broyden-like method</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Converged in 18 its</td>
<td>Converged in 6 its</td>
<td>Converged in 5 its</td>
</tr>
<tr>
<td>10</td>
<td>Diverged</td>
<td></td>
<td>|x</td>
</tr>
<tr>
<td>30</td>
<td>Diverged ||x</td>
<td></td>
<td>_2^2 \sim 10^6</td>
</tr>
</tbody>
</table>

Although Brown's method has taken a larger number of iterations, it may take less CPU time than that of the Broyden-like method because Brown's method is quadratically convergent.

6. DISCUSSIONS.

(i) In the case of Broyden's method the initial estimate B_o is found by taking a finite difference analogue of F'(x_o). We are considering the case where the complete computation of F'(x) is infeasible. In our Broyden-like method we start with B_o an M-matrix so that H_o > 0. Since F is an isotone mapping it could be that all the entries of F'(x_o) are non-negative. Nevertheless we can choose B_o such that \|B_o - F'(x_o)\| is sufficiently small. Such a B_o satisfies the Quasi-Newton equation. Moreover we have used Broyden updates (good method) and as such we have called our method the Broyden-like method. (ii) In the case of DFP's method, H_o is always taken as a symmetric positive-definite matrix. Moreover, a symmetric M-matrix is positive definite [12]. Hence our \H_n can sometimes turn out to be a positive-definite matrix as in DFP's method. (iii) Theorem 3.1 can only provide us with non-negative solutions of nonlinear equations. With a suitable translation we can transform the given equation into another equation having non-negative solutions only. (iv) Our convergence proof seems to be much more elegant than the cases where "majorization principle" has been utilized or where the Euclidean norm of \E_i (E_i = A^{-1}B_i - I, for linear system of equations A\_x = b) has been utilized. (v) The convergence theorem 3.1 is applicable where F is isotone. This restricts the sphere of applicability of the theorem. However, in a large number of problems, the
nonlinear operators are Monotonically Decomposable (MDO) [13] and convergence theorems along the line of theorem 3.1 can be developed for such operators. The result would be communicated in a separate paper.

ACKNOWLEDGEMENTS.

The authors are extremely grateful to Dr. L. Dixon of Hatfield Polytechnic U.K. for some constructive comments.

The second author, who is a N.B.H.M (India) research awardee, is grateful to the National Board for Higher Mathematics, India for financial support to carry out the research.

REFERENCES