Internat. J. Math. & Math. Sci.
VOL. 14 NO. 2 (1991) 349-362

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 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 R^n, M-matrix.
1980 AMS SUBJECT CLASSIFICATION CODE. 65K.

1. INTRODUCTION.

Let $\varphi : D \subset R^n \longrightarrow R^1$ be an F-differentiable functional having an optimum $x^* \in \text{int}(D)$.
Then the vector x^* at which the optimum of φ is realized, satisfies the equation

$$\nabla\varphi(x) = P(x) = 0$$

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

$$F(x) = 0; \; F: R^m \longrightarrow R^m \tag{1.1}$$

in the case when the complete computation of F' is infeasible.

In the case, F = P, solving (1.1) means in effect finding the minimizer of φ.

The algorithm under consideration takes the form

$$x_{n+1} = x_n - \gamma_n H_n F(x_n) \tag{1.2}$$

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

$$H_{n+1} (F(x_{n+1}) - F(x_n)) = x_{n+1} - x_n \tag{1.3}$$

is satisfied at each step.

The step length γ_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 $\gamma_n = \gamma_n^*$ 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_o R^{mxm} and also x_o R^m.

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

γ_n being chosen such that

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

$$x_{n+1} = x_n + s_n \tag{1.5}$$

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

If $y_n = 0$ then $\quad H_{n+1} = H_n$ (1.7)

and choose v_n R^m such that $v_n^T y_n = 1$ and $v_n^T H_n s_n \neq 0$.

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

Broyden's [3] method (sometimes called his first or good method) results from choosing $v_n = H_n^T s_n / s_n^T H_n y_n$ 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 $y_n^T H_n^{-1} s_n \neq 0$.

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

[10]. 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_o to be an M-matrix [12] unlike Broyden [3]. We have used componentwise partial ordering in 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 decomposible (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 R^m is defined as follows:

For $x, y \in R^m$, $x = (x_1, x_2, \ldots x_m)^T$

$y = (y_1, y_2, \ldots y_m)^T$, $x > y$ if and only if

$x_i > y_i$, $i = 1, 2, \ldots m$,

$x > y \iff x > y$ and $x \neq y$.

DEFINITION 2.2. We define for any $x, y \in R^m$, such that $x < y$, the order interval [11]

$$\langle x, y\rangle = \{u \quad R^m / x < u < y\}.$$

DEFINITION 2.3. A mapping $F: D \quad R^m \longrightarrow R^m$ is isotone (antitone) [12] on $D_o \quad D$ if $Fx < Fy$ $(Fx > Fy)$ whenever $x < y$, $x, y \in D_o$.

DEFINITION 2.4. We denote by $L(R^m)$ the space of mxm matrices.

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

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

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

$$B_n s_{n-1} = y_{n-1}, \quad n = 1, 2, \ldots \tag{2.1}$$

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

$$B_{n+1} = B_n - \frac{(B_n s_n - y_n) s_n^T}{s_n^T s_n} \tag{2.2}$$

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

LEMMA 2.1. $(I - \frac{s_n s_n^T}{s_n^T s}) $ is a square matrix whose diagonal elements are positive and off-diagonal elements are non-positive provided $s_i^{(n)} > 0$ and $s^{(n)} > 0$ 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) $$\left| \frac{\sum_k b_{ik}^{(n)}(s_k^{(n)} - s_k^{(n-1)}) - (y_k^{(n)} - y_k^{(n-1)})s_j^{(n)}}{\sum_i (s_i^{(n)})^2} \right|$$

is small compared to $|b_j^{(n)}|$

(v) $$\sum_{i\neq j} |b_{ij}^{(n)}| < b_{ii}^{(n)} + \frac{\sum_j \sum_k [b_{ik}^{(n)}(s_k^{(n)} - s_k^{(n-1)}) - (y_k^{(n)} - y_k^{(n-1)})]s_j^{(n)}}{\sum_i (s_i^{(n)})^2}$$

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)

$$(B_n \dot{s}_n - y_n)\dot{s}_n^T = \frac{[B_n(s_n - s_{n-1}) - (y_n - y_{n-1})]\, s_n^T}{s_n s_n^T}$$

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

$$b_{ij}^{(n)} - \frac{\sum_k [b_{ik}^{(n)}(s_k^{(n)} - s_k^{(n-1)}) - (y_k^{(n)} - y_k^{(n-1)})]\, s_j^{(n)}}{\sum_j s_j^{(n)}}$$

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. $s_j^{(n)}/\sum_j (s_j^{(n)})^2$ 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

$$\frac{\sum_k [b_{ik}^{(n)}(s_k^{(n)} - s_k^{(n-1)}) - (y_k^{(n)} - y_k^{(n-1)})]\, s_j^{(n)}}{\sum_j (s_j^{(n)})^2}$$

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

Condition (v) implies that

$$\sum_{\substack{j \\ i \neq j}} -(b_{ij}^{(n)} + c_{ij}^{(n)}) < b_{ii}^{(n)} + c_{ii}^{(n)} \qquad (2.4)$$

and since $|b_{ij}^{(n)}|$ is large compared to $|c_{ij}^{(n)}|$,(2.4) implies that

$$\sum_{i\neq j} |b_{ij}^{(n+1)}| < |b_{ii}^{(n+1)}|$$

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 $P_nB_n(I-\frac{s_ns_n^T}{s_n^Ts_n})$ becomes the predominant term in the expression for $\tilde{B}_{n+1}$ as given below.

$$\tilde{B}_{n+1} = P_nB_n\,(I - \frac{s_ns_n^T}{s_n^Ts_n}) + \frac{y_ns_n^T}{s_n^Ts_n}\,. \tag{2.5}$$

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

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

$$H_{n+1} = [P_nB_n(I - \frac{s_ns_n^T}{s_n^Ts} + \frac{B_n^{-1}P^{-1}y_ns_n^T}{s_n^Ts_n})]^{-1}$$

$$= [I - \frac{s_ns_n^T}{s_n^Ts_n} + \frac{B_n^{-1}P_n^{-1}y_ns_n^T}{s_n^Ts_n}]^{-1}\ B_n^{-1}P_n^{-1}$$

which is approximately equal to

$$[I + (s_n - H_nP_n^{-1}y_n)\frac{s_n^T}{s_n^Ts_n}]\ H_nP_n^{-1}.$$

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

$$H_{n+1} = H_nP_n^{-1} + (s_n - H_nP_n^{-1}y_n)\frac{s_n^TH_nP_n^{-1}}{s_n^TH_nP_n^{-1}y_n} \tag{2.6}$$

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 $b_{ij}^{(n)}$ of the strictly diagonally dominant matrix B_n satisfy the condition $|b_{ii}^{(n)}| < M$, for all i, where $M(>0)$ is a finite constant. Then

$$\frac{1}{|\lambda_i^{(n)}|} > \frac{1}{2\eta},\ \lambda_i^{(n)} \text{ are the eigenvalues of } B_n.$$

PROOF. By the Greschgorin theorem,

$$|\lambda_i^{(n)} - b_{ii}^{(n)}| < \sum_{i\neq j}|b_{ij}^{(n)}|\ .$$

By strict diagonal dominance of B_n,

$$|\lambda_i^{(n)}| < 2|b_{ii}^{(n)}| < 2M.$$

Therefore,

$$\frac{1}{|\lambda_i^{(n)}|} > \frac{1}{2M} > 0.$$

The above lemma ensures that the lower bound of the eigenvalues of $H_n = B_n^{-1}$ 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 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) $F:D\subset R^m \longrightarrow R^m$ is Frechet differentiable on a convex set $D_o \subset D$.

(ii) D_o includes the null element and x_o, $y_o \in D_o$ such that if $x \leq y$ then the order interval $\langle x,y\rangle \subset D_o$.

(iii) $F(x) \leq 0$, for all x D_o and F is isotone i.e. if $x \leq y$, $F(x) \leq F(y)$ for all $x,y \in D_o$.

(iv) $x_o \in D_o$ is an initial approximation to the solution and $x_o \geq 0$ and $F(x_o) < 0$.

(v) The operator $G = I - \gamma_o H_o F$ is such that $GD_o \subset D_o$.

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

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

$$\frac{\left|\sum_k [b_{ik}^{(n)}(s_k^{(n)} - s_k^{(n-1)}) - (y_k^{(n)} - y_k^{(n-1)})]\, s_j^{(n)}\right|}{\sum_i (s_i^{(n)})^2}$$

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

(viii) $$\sum_{\substack{j\\ i\neq j}} b_{ij}^{(n)} < b_{ii}^{(n)} + \frac{\sum_j \sum_k b_{ik}^{(n)}(s_k^{(n)} - s_k^{(n-1)}) - (y_k^{(n)} - y_k^{(n-1)})\, s_j^{(n)}}{\sum_i (s_i^{(n)})^2},$$

$n = 0,1,2,\ldots$

(ix) $|b_{ii}^{(n)}| \leq (K > 0)$, $n = 0,1,2,\ldots$

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

(a) $\gamma_{n-1} H_{n-1} F(x^o) \leq \gamma_n H_n F(x_o) \leq 0$, $n \geq 1$

(b) $\sup_{(x)} [I - \gamma_n H_n F'(x)] \leq [I - \gamma_{n-1} H_{n-1} F'(x_{n-1})]$, $x \in \langle x_{n-1}, x_n\rangle$

(xi) $F'(x)$ is Frechet differentiable for all $x \in D_o$.

(xii) $\lim_{n\to\infty} [I - \gamma_o H_o F'(x_o)]^n [x_1 - x_o] = 0$ in the order sense.

Then starting from $x_o > 0$, an initial approximation to the solution of $F(x) = 0$ is the sequence $\{x_n\}$ of Broyden-like approximations defined by

$$x_{n+1} = x_n - \gamma_n H_n F(x_n), \quad n = 0,1,\ldots \tag{3.1}$$

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 nth stage is given by

$$||x^* - x_n|| \leq ||A^n(I-A)^{-1} (x_1 - x_o)|| \tag{3.2}$$

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

$$x_1 = x_o - \gamma_o H_o F(x_o) \geq x_o \geq 0 \tag{3.3}$$

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

B_o is an M-matrix [12] by (vi). Hence $H_o = B_o^{-1} \geq 0$, $s_o = x_1 - x_o \geq 0$.

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

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

$$0 \geq \gamma_1 H_1 F(x_1) \geq \gamma_1 H_1 F(x_o) \geq \gamma_o H_o F(x_o). \tag{3.4}$$

Hence $x_2 \geq x_1 \geq 0$. Thus $x_1 \leq x_2 = x_1 - \gamma_o H_o F(x_o) = Gx_1 \in D_o$.

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

Since B_n is an M-matrix, $H_n \geq 0$,

By conditions (iii) and (xa)

$$\begin{aligned} \gamma_n H_n F(x_n) &\geq \gamma_n H_n F(x_o) \\ &\geq \ldots\ \ldots \\ &\geq \gamma_o H_o F(x_o) \end{aligned}$$

Therefore,

$$\begin{aligned} x_n \leq x_{n+1} &\leq x_n - \gamma_n H_n F(x_o) \\ &\leq x_n - \gamma_o H_o F(x_o) \\ &= Gx_n \in D_o \end{aligned}$$

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

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.

$$\begin{aligned} \text{Now, } 0 \leq x_2 - x_1 &= x_1 - x_o - \gamma_1 H_1 [F(x_1) - F(x_o)] \\ &\quad - [\gamma_1 H_1 F(x_o) - \gamma_o H_o F(x_o)] \\ &\leq x_1 - x_o - \gamma_1 H_1 [F(x_1) - F(x_o)]. \end{aligned}$$

D_o being a convex set and using the mean value theorem in R^m we obtain from (3.5)

$$x_2 - x_1 \leq \int_o^1 [I - \gamma_1 H_1 F'(x_o + t(x_1 - x_o))] (x_1 - x_o)dt \quad 0 < t < 1 \tag{3.6}$$

Here ' stands for differentiation in the Frechet sense.

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

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

$$0 \leq x_2 - x_1 \leq \sup_t [I - \gamma_1 H_1 F'(x_o + t(x_1 - x_o))](x_1 - x_o)$$
$$= [I - \gamma_1 H_1 F'(x_o + t(x_1 - x_o))] (x_1 - x_o) \tag{3.7}$$

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

By (xb), (3.7) further simplifies to

$$0 \leq x_2 - x_1 \leq [I - \gamma_o H_o F'(x_o)] (x_1 - x_2). \tag{3.8}$$

Arguing analogously as before,

$$x_{n+1} - x_n \leq \int_o^1 [(I - \gamma_n H_n F'(x_{n-1} + t(x_n - x_{n-1}))] (x_n - x_{n-1})dt$$
$$\leq \sup_t [I - \gamma_n H_n F'(x_{n-1} + t(x_n - x_{n-1}))] (x_n - x_{n-1})$$
$$= [I - \gamma_n F_n F'(x_{n-1} + \tilde{t}(x_n - x_{n-1}))] (x_n - x_{n-1}) \tag{3.9}$$

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

Condition (xb) further reduces (3.9) to

$$x_{n+1} - x_n \leq [I - \gamma_{n-1} H_{n-1} F'(x_{n-1})] (x_n - x_{n-1})$$
$$\leq [I - \gamma_o H_o F'(x_o)] (x_n - x_{n-1})$$
$$\leq [I - \gamma_o H_o F'(x_o)]^n (x_1 - x_o) \tag{3.10}$$

Hence $$x_{n+p} - x_n \leq \sum_{k=n}^{n+p-1} [I - \gamma_o H_o F'(x_o)]^k (x_1 - x_o) \tag{3.11}$$

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

$$x^* = \lim_{n\to\infty} x_n \quad R^m \tag{3.12}$$

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

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

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

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

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

Continuity of F yields that

$$F(x^*) = 0.$$

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

Further if $I - \gamma_o H_o F'(x_o) \leq A$, (3.11) reduces to

$$x_{n+p} - x_n \leq \sum_{k=n}^{n+p-1} A^k(x_1 - x_0) \tag{3.15}$$

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

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) $||F'(x) - F'(y)|| \leq \gamma'||x-y||$, for all $x,y \in D_0$

(b) $[F'(x_0)]^{-1}$ exists and $||[F'(x_0)]^{-1}|| \leq \beta'$

(c) $||\sum_{n=0}^{\infty} (I - \gamma_0 H_0 F'(x_0))^n(x_1 - x_0)|| \leq \beta'\gamma'$

Then

$$||[F'(x_n)]^{-1} - \gamma_n H_n|| \leq \frac{||I - \gamma_0 H_0 F'(x_0)||}{1-\beta'\gamma'||\sum_{k=0}^{n-1} (I - \gamma_0 H_0 F'(x_0))^k(x_1 - x_0)||} \tag{3.16}$$

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

Let $\underline{\Omega}(x_0,(\beta'\gamma')^{-1})$ denote the closed sphere with x_0 as the center and $(\beta'\gamma')^{-1}$ as the radius.

Let $D_1 = \underline{\Omega}(x_0,(\beta'\gamma')^{-1}) \cap D_0$

Then, for $x \in D_1$ we have

$$||F'(x) - F'(x_0)|| \leq \gamma'||x - x_0|| < 1/\beta'. \tag{3.17}$$

Now, $\qquad ||I - [F'(x_0)]^{-1} F'(x_1)||$

$$||[(F'(x_0)]^{-1} [F'(x_0) - F'(x_1)]|| < \beta' \frac{1}{\beta'} = 1 \tag{3.18}$$

Therefore $[F'(x_0)]^{-1} F'(x_1)$ 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

$$[F'(x)]^{-1} = \sum_{n=0}^{\infty}([F'(x_0)]^{-1} [F'(x_0) - F'(x)])^n[F'(x_0)]^{-1} \tag{3.19}$$

Therefore,

$$||\sum_{n=0}^{\infty}([(F'(x_0)]^{-1} [F'(x_0) - F'(x)])^n||$$

$$\leq \sum_{n=0}^{\infty} (\beta'\gamma'||x - x_0||)^n = \frac{1}{1-\beta'\gamma'||x-x_0||} \tag{3.20}$$

since $\beta'\gamma'||x - x_0|| < 1$

Therefore for $x \in D_1$

$$||[F'(x)]^{-1}|| \leq \frac{\beta'}{1-\beta'\gamma'||x - x_0||} \tag{3.21}$$

$\{x_n\}$ being a monotonic increasing sequence it follows from (3.10) and condition (c) of Theorem (3.2),

$$||x_n - x_0|| \leq ||\sum_{k=0}^{n-1} (I - \gamma_0 H_0 F(x_0))^k (x_1 - x_0)|| < \beta'\gamma' \tag{3.22}$$

Hence $\qquad x_n \in \underline{\Omega}(x_0,\beta'\gamma') \cap D_0 = D_1$

Utilizing condition (xb) and (3.21) we get

$$\|[F'(x_n)]^{-1} - \gamma_n H_n\| = \|[I - \gamma_n H_n F'(x_n)]\, [F'(x_n)]^{-1}\|$$
$$\leq \|I - \gamma_n H_n F'(x_n)\|\ \|[F'(x_n)]^{-1}\|$$
$$\leq \frac{\beta' \|I - \gamma_o H_o F'(x_o)\|}{1-\beta'\gamma'\|x_n - x_o\|} \tag{3.23}$$

Using (3.10) and the monotonic increasing property of $\{x_n\}$ we further conclude that

$$\|[F'(x)]^{-1} - \gamma_n H_n\| \leq \frac{\beta' \|I - \gamma_o H_o F'(x_o)\|}{1-\beta'\gamma'\|\sum_{k=0}^{n-1} [I - \gamma_o H_o F'(x_o)]^{-k} (x_1 - x_o)\|}$$

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

$$(I - \gamma_{n-1} H_{n-1} F'(x_{n-1})) \geq \alpha' \sup_{x \in D_o} (I - \gamma_n H_n F'(x_n)) \tag{3.24}$$
$$\alpha' > 1, \ {}^o n > 1$$

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

PROOF. $$\lim_{n\to\infty} (I - \gamma_n H_n F'(x_n)) \leq \sup_{x\ D_o} (I - \gamma_n H_n F'(x_n))$$
$$\leq \frac{1}{\alpha^n} (I - \gamma_o H_o F'(x_o)), \ (\alpha^1 > 1)$$
$$\to 0 \text{ as } n \to \infty. \tag{3.25}$$

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

$$0 \leq x^* - x_{n-1} = x^* - x_n - \gamma_n H_n [F(x^*) - F(x_n)]$$
$$= \int_0^1 [I - \gamma_n H_n F'(x_n + t(x^* - x_n))]\, (x^* - x_n) dt\ .$$

By arguments analogous to theorem 3.1

$$\|x^* - x_{n+1}\| \leq \sup_{t \in [0,1]} \|I - \gamma_n H_n F'(x_n + t(x^* - x_n))\|\ \|x - x_n\|.$$

Therefore,

$$\frac{\|x^* - x_{n+1}\|}{\|x^* - x_n\|} < \sup_{t \in [0,1]} \|I - \gamma_n H_n F'(x_n + t(x^* - x_n))\|, \ .$$

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

$$\lim_{n\to\infty} \frac{\|x^* - x_{n+1}\|}{\|x^* - x_n\|} \leq \lim_{n\to\infty} (I - \gamma_n H_n F'(x)) = 0 \tag{3.26}$$

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: $\underset{\sim}{x} \leq x \leq \tilde{x}$ in which $F(x)$ is isotone.

Step 2. Choose $x_o \in D_o$ and $x_o > 0$ s.t. $F(x_o) < 0$

Step 3. Choose $H_o > 0$, $\gamma_o = 1$ and compute x_1, $k = 0$.

Step 4. Compute $A_o = I - \gamma_o H_o F'(x_o)$.

If $(I - \gamma_o H_o F'(x_o))^n (x_1 - x_n) \to 0$ for some $n > n_o$ go to step 5, otherwise go to step 3.

Step 5. Compute $s_k = -\gamma_k H_k F(x_k)$, $x_{k+1} = x_k + s_k$ and $F(x_{k+1})$

Step 6. If $F(x_{k+1}) \simeq 0(10^{-4})$ 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.

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

Step 8. Compute γ_{k+1} s.t. $\gamma_{k+1} H_{k+1} F(x_o) > \gamma_k H_k F(x_o)$

Step 9. If $I - \gamma_k H_k F'(x_k) > \sup\limits_{x \in \langle x_k, x_{k+1} \rangle} [I - \gamma_{k+1} H_{k+1} F'(x)]$ 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 γ_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 γ_{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 γ_{k+1} respectively is $(m+m^2)$, $2m^2+m$, $2m^2 + 3m$, i.e. $5m^2 + 5m$. (iv) The number of function evaluations is $m^2 + 2m$.

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

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

where
$$f_i(x) = -(N+1) + 2x^i + \sum_{\substack{j=1 \\ i \neq j}}^{n} x^j, \; i = 1,2 \ldots N-1$$

and
$$f_N(x) = -1 + \prod_{j=1}^{N} x^j .$$

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

Define D_o, the rectangular parallelepiped given by $0.5 \leqslant x^i \leqslant 1.0$, $i = 1,2,\ldots N$, and choose $\gamma_o = 1$. $H_o = (h_{ij}^o)$ where $h_{ij}^o = 0.1$, $i \neq N$, $h_{NN}^o = 0.5$. $h_{ij} = .01$. $A = [I - \gamma_o H_o F'(x_o)]$ is a convergent matrix. $i \neq j$

We summarize our computational experience in table 1. 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 we provide 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 1

Dimension(N)	No. of iterations(n)	x	F(x)	CPU Time
5	5	$(1,1,1)^T$	0	3.637 sec
10	5	$(1,1,\dots,1)^T$	0	4.462 sec
30	6	$(1,1,\dots,1)^T$	0	8.548 sec

Table 2

Dimension(N)	Newton's method	Brown's Method	Broyden-like method
5	Converged in 18 its	Converged in 6 its	Converged in 5 its
10	Diverged $\|\|x\|\|_{\ell_2}^{1} \sim 10^3$	Converged in 9 its	Converged in 5 its
30	Diverged $\|\|x\|\|_{\ell_2}^{1} \sim 10^6$	Converged in 9 its	Converged in 6 its

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 positve-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 Decomposible (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

1. DAVIDON, W.C., Variable metric method for minimization. Argonne Nat. Labs. Report ANL-5990, Rev. Argonne II (1959).
2. FLETCHER, R. and POWELL, M.J.D., A rapidly convergent descent method for minimizations. Comput J. 6 (1963), 163-168.
3. BROYDEN, C.G., A class of methods for solving non-linear simultaneous equations, Maths. Comp. 19, (1965), 577-593.
4. DIXON, L.C.W., Numerical methods for non-linear optimization (1972) ed. F.A. Looftsma. Academic Press.
5. GAY, D.M., Some convergence properties of Broyden's method. SIAM. J. Num Aual 16(4), (1979), 623-630.
6. BROYDEN, C.G., Numerical methods for non-linear algebraic equations, ed. P. Rabinowitz (1970).
7. BROYDEN, C.G., DENNIS, J.E. and MORE, J.J., On the local and superlinear convergence of Quasi-newton methods, J. Inst. Maths. Appln.12, (1973), 223-245.
8. MORE, J.J. and TRANGSTEIN, J.A., On the global convergence of Broyden's method, Math. Comput. 30, (1976), 523-540.
9. DENNIS, J.R. and WALKER, H.F., Convergence theorems for least change secant update methods. SIAM J. Numerical Anal. 18, (1981), 949-987.
10. DENNIS, J.E. and MORE, J.J., A characterization of superlinear convergence and its applications to Quasi-Newton methods. Maths. Comp. 28, (1974), 549-560.
11. DECKER, D.W. and KELLEY, C.T., Broyden's method for a class of problems having singular Jacobian at the root, SIAM J. Numer. Anal. 22(1), (1985), 566-573.
12. ORTEGA, J.M. and RHEINBOLD, W.C., Iterative solution of nonliear equations in several variables, Academic Press (1970).
13. COLLATZ, L., Functional Analysis and Numerical Analysis Ed. L.B. Rall, Academic Press, N.Y. (1966).
14. BYRNE, G.D. and HALL, C.A., Numerical solution of system of non-linear algebraic equations , Academic Press N.Y. (1973).