

On the Global Convergence of Broyden's Method*

By J. J. Moré and J. A. Trangenstein

Abstract. We consider Broyden's 1965 method for solving nonlinear equations. If the mapping is linear, then a simple modification of this method guarantees global and Q -superlinear convergence. For nonlinear mappings it is shown that the hybrid strategy for nonlinear equations due to Powell leads to R -superlinear convergence provided the search directions form a uniformly linearly independent sequence. We then explore this last concept and its connection with Broyden's method. Finally, we point out how the above results extend to Powell's symmetric version of Broyden's method.

1. Introduction. Let $F: R^n \rightarrow R^n$ be a mapping with domain and range in real n -dimensional Euclidean space R^n , and consider the problem of finding a solution to the system of equations $F(x) = 0$ by Broyden's [1] method.

In this paper we show that a simple modification of Broyden's method leads to global and Q -superlinear convergence if F is an affine function with nonsingular coefficient matrix. This improves on a result of Broyden [2] which gives local and R -superlinear convergence to the unmodified method. For future reference, recall (for more information see [7, Chapter 9]) that if a sequence $\{x_k\}$ converges to x^* , then $\{x_k\}$ converges R -superlinearly to x^* if

$$\lim_{k \rightarrow +\infty} \|x_k - x^*\|^{1/k} = 0,$$

and that $\{x_k\}$ converges Q -superlinearly to x^* if there is a sequence $\{\alpha_k\}$ converging to zero such that

$$\|x_{k+1} - x^*\| \leq \alpha_k \|x_k - x^*\|, \quad k \geq 0.$$

Clearly, Q -superlinear convergence implies R -superlinear convergence but the converse does not hold.

If F is not affine, the above modification of Broyden's method fails to be globally convergent, although an improvement of a result of Broyden, Dennis and Moré [3], shows that it is locally and superlinearly convergent under very reasonable conditions.

In order to ensure the global convergence of Broyden's method we follow Powell's [8] hybrid method. For this algorithm Powell proved a global convergence result, but did not analyze the rate of convergence. In this paper we show that if the sequence $\{x_k\}$ generated by the hybrid method converges to a point x^* then $F'(x^*)^T F(x^*) = 0$ where $F'(x^*)$ denotes the Jacobian matrix of F at x^* . Thus if $F'(x^*)$ is nonsingular then $F(x^*) = 0$, and under this condition, we show that in general $\{x_k\}$ converges R -superlinearly to x^* .

Received January 14, 1975; revised July 10, 1975.

AMS (MOS) subject classifications (1970). Primary 65H10, 65K05.

*This research was done while the authors were at Cornell University and was supported in part by the National Science Foundation under Grants GJ-40903 and GZ-03527.

Copyright © 1976, American Mathematical Society

The hybrid method requires "special iterations" which guarantee that the Jacobian approximations in Broyden's method do not differ radically from the true Jacobians. Powell's [8] special iterations guarantee this by making sure that the directions generated by the algorithm are uniformly linearly independent. In Section 5 we examine this concept and show that the various definitions in the literature are equivalent. This leads to particularly easy proofs of the results of Powell [8] on the behavior of the matrices generated by Broyden's update.

Finally, in Section 6 we discuss the extension of the previous results to Powell's [10] symmetric form of Broyden's update.

As far as notation is concerned, we assume that R^n is equipped with the usual inner product $\langle x, y \rangle = x^T y$ and $\|\cdot\|$ denotes the l_2 vector norm or the corresponding operator norm in $L(R^n)$ —the linear space of all real matrices of order n . We shall also use the Frobenius norm

$$(1.1) \quad \|A\|_F = [\text{trace}(A^T A)]^{1/2}$$

and the fact that for any A and B in $L(R^n)$,

$$(1.2) \quad \|AB\|_F \leq \|A\| \|B\|_F.$$

2. Broyden's Method. Let $F: R^n \rightarrow R^n$ be given. In its simplest form Broyden's method is of the form

$$(2.1) \quad x_{k+1} = x_k - B_k^{-1} F(x_k),$$

where, given an approximation B_0 to $F'(x_0)$, the matrices $\{B_k\}$ are generated by

$$(2.2) \quad B_{k+1} = B_k + (y_k - B_k s_k) s_k^T / \|s_k\|^2$$

and

$$(2.3) \quad y_k = F(x_{k+1}) - F(x_k), \quad s_k = x_{k+1} - x_k.$$

The motivation for Broyden's method is that the matrices generated by (2.2) are good approximations to the Jacobian matrices and thus (2.1) resembles Newton's method, but with the difference that (2.2) only requires one function evaluation and $O(n^2)$ arithmetic operations while the Jacobian matrix requires the evaluation of n^2 partial derivatives. Moreover, (2.1) can be carried out in $O(n^2)$ operations while Newton's method requires $O(n^3)$.

There are two ways to compute x_{k+1} in $O(n^2)$ arithmetic operations. In the first method the inverse of $\{B_k\}$ can be computed by the Sherman-Morrison formula as

$$(2.4) \quad H_{k+1} = H_k + (s_k - H_k y_k) s_k^T H_k / \langle s_k, H_k y_k \rangle,$$

while in the second method a QR factorization of B_k is carried along; for example, see the technique of Gill and Murray [6]. Either method can be done in $O(n^2)$ arithmetic operations per iteration but the latter method is recommended for stability reasons.

Further information and motivation for Broyden's method can be found in the

survey paper [5]; in particular, that paper contains a discussion of the following result of Broyden, Dennis and Moré [3].

THEOREM 2.1. *Let $F: R^n \rightarrow R^n$ be continuously differentiable in an open convex set D , and assume that $F(x^*) = 0$ and $F'(x^*)$ is nonsingular for some $x^* \in D$. In addition, suppose that F' is Lipschitz continuous at x^* , and consider Broyden's method as defined by (2.1), (2.2) and (2.3). Then Broyden's method is locally and Q -super-linearly convergent at x^* .*

To be more precise, the conclusion of this theorem means that there is an $\epsilon > 0$ and a $\delta > 0$ such that if $\|x_0 - x^*\| < \epsilon$ and $\|B_0 - F'(x^*)\| < \delta$, then Broyden's method is well defined, and if $\{x_k\}$ is the sequence generated, then either $x_k = x^*$ for some k at which place the iteration stops, or $\{x_k\}$ converges Q -superlinearly to x^* .

Broyden's method is sometimes modified by defining $\{x_k\}$ by (2.1) and (2.2) but instead of (2.3),

$$(2.5) \quad y_k = F(x_k + s_k) - F(x_k)$$

for some nonzero vector s_k . The proof of Theorem 2.1 shows that this version of Broyden's method is locally and linearly convergent if s_k satisfies a relationship of the form

$$\|s_k\| \leq \eta \max\{\|x_{k+1} - x^*\|, \|x_k - x^*\|\}$$

provided x_k, x_{k+1} belong to D . However, superlinear convergence will be lost unless the direction of s_k is chosen with some care. For example, the choice $s_k = \|F(x_{k+1})\|u$ for some fixed vector u leads to local and linear convergence, but rarely to superlinear convergence. In this connection note that under the assumptions of Theorem 2.1, Dennis and Moré [4] proved that if the sequence x_k generated by (2.1) converges to x^* then $\{x_k\}$ converges Q -superlinearly to x^* if and only if

$$(2.6) \quad \lim_{k \rightarrow +\infty} \frac{\|[B_k - F'(x^*)](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0.$$

This explains why the choice $s_k = \|F(x_{k+1})\|u$ rarely leads to superlinear convergence. However, in Section 5 we will show that if the direction of s_k is chosen so that $\{s_k/\|s_k\|\}$ is uniformly linearly independent, then the matrices $\{B_k\}$ generated by (2.2) converge to $F'(x^*)$ and thus (2.6) holds. Hence, in this case we also have Q -super-linear convergence, but note that if $s_k \neq x_{k+1} - x_k$, then the computation of (2.2) would require an additional function evaluation unless $F(x_k + s_k)$ is used at some other stage of the calculation.

3. Broyden's Method for Linear Equations. We would like to improve Theorem 2.1 if $F: R^n \rightarrow R^n$ is affine with nonsingular coefficient matrix; that is,

$$(3.1) \quad F(x) = Ax - b, \quad A \in L(R^n) \text{ nonsingular.}$$

To investigate this problem, first note that the matrices generated by (2.2) may be singular. In fact, it is easy to verify that if B_k is nonsingular then B_{k+1} is nonsingular if and only if $\langle s_k, B_k^{-1}y_k \rangle \neq 0$. This also follows from the following result, whose simple proof can be found, for example, in [5, Lemma 4.4].

LEMMA 3.1. *Let $u, v \in R^n$. Then $\det(I + uv^T) = 1 + \langle u, v \rangle$.*

This result also shows how to avoid singularity in B_{k+1} . Powell [9] sets

$$(3.2) \quad B_{k+1} = B_k + \theta_k(y_k - B_k s_k) s_k^T / \|s_k\|^2,$$

where θ_k is chosen so that B_{k+1} is nonsingular. To be more precise, given $\sigma \in (0, 1)$ we choose θ_k so that

$$(3.3) \quad |\det B_{k+1}| \geq \sigma |\det B_k|, \quad |\theta_k - 1| \leq \sigma.$$

To see that this is possible, note that Lemma 3.1 implies that

$$|\det B_{k+1}| = |\det B_k| |(1 - \theta_k) + \theta_k \langle B_k^{-1} y_k, s_k \rangle / \|s_k\|^2|.$$

Thus, if γ_k is defined by $\langle B_k^{-1} y_k, s_k \rangle \equiv \gamma_k \|s_k\|^2$ then we can choose

$$\theta_k = \begin{cases} 1, & |\gamma_k| \geq \sigma, \\ \frac{1 - \text{sign}(\gamma_k)\sigma}{1 - \gamma_k}, & |\gamma_k| < \sigma, \end{cases}$$

where $\text{sign}(0) = 1$. It is not too difficult to show that this choice of θ_k provides a number closest to unity so that (3.3) is satisfied. In the rest of the paper we will only assume that θ_k is chosen to satisfy

$$(3.4) \quad B_{k+1} \text{ nonsingular}, \quad |\theta_k - 1| \leq \hat{\theta} < 1.$$

THEOREM 3.2. *Let $F: R^n \rightarrow R^n$ be given by (3.1) and consider Broyden's method as defined by (2.1), (2.3), (3.2) and (3.4). Then Broyden's method is globally and Q -superlinearly convergent.*

Proof. The result follows from a careful estimation of the difference between $\|E_{k+1}\|_F^2$ and $\|E_k\|_F^2$ where $E_k = B_k - A$ and $\|\cdot\|_F$ is the Frobenius norm. For this note that

$$E_{k+1} = E_k(I - \theta_k s_k s_k^T / \|s_k\|^2),$$

and therefore, direct calculation with $\|E\|_F^2 = \text{trace}(E^T E)$ yields

$$\|E_{k+1}\|_F^2 = \|E_k\|_F^2 - \theta_k(2 - \theta_k)(\|E_k s_k\| / \|s_k\|)^2.$$

This implies that

$$(1 - \hat{\theta})^2 \sum_{k=0}^{\infty} \left(\frac{\|E_k s_k\|}{\|s_k\|} \right)^2 \leq \|E_0\|_F^2,$$

and in particular,

$$(3.5) \quad \lim_{k \rightarrow \infty} \frac{\|E_k s_k\|}{\|s_k\|} = 0.$$

Now note that $(B_k - A)s_k = -F(x_{k+1}) = -A(x_{k+1} - x^*)$ where $x^* = A^{-1}b$.

Therefore, if ϵ_k is defined by

$$\|A^{-1}\| \|E_k s_k\| = \epsilon_k \|s_k\|,$$

then

$$\|x_{k+1} - x^*\| \leq \epsilon_k \|s_k\| \leq \epsilon_k [\|x_{k+1} - x^*\| + \|x_k - x^*\|],$$

and (3.5) clearly shows that $\{\epsilon_k\}$ converges to zero. The above inequality then implies that

$$\|x_{k+1} - x^*\| \leq \epsilon_k \|x_k - x^*\| / (1 - \epsilon_k)$$

for k sufficiently large, and this proves that $\{x_k\}$ converges Q -superlinearly to x^* .

Theorem 3.2 is interesting because to our knowledge it is the only iterative method which is globally and superlinearly convergent for arbitrary nonsingular linear systems.

While the above-mentioned modification of Broyden's method leads to global and superlinear convergence in the linear case, this does not hold for general nonlinear functions. In one dimension Broyden's method essentially reduces to the secant method, and this method can cycle.

Example 3.3. Let $f: R \rightarrow R$ be any continuously differentiable function such that

$$f(\pm 1) = \pm 1, \quad f(\pm(\sqrt{5} - 2)) = \pm \left(\frac{\sqrt{5} - 1}{2} \right).$$

For example, $f(x) = \alpha \arctan(\beta x)$ with $\alpha = 0.733 \dots$ and $\beta = 4.75 \dots$. It can then be verified that Broyden's method as defined by (2.1), (2.2) and (2.3) cycles if $x_0 = 1$ and $B_0 = (3 - \sqrt{5})^{-1}$. To be more specific, it turns out that $x_{2k+1} = (-1)^k(\sqrt{5} - 2)$ and that $x_{2k} = (-1)^k x_0$. Also note that if B_k is defined by (3.2) and (3.3), instead of by (2.2), then $\theta_k = 1$ satisfies (3.3) if $\sigma \leq 0.37$.

Example 3.3 shows that Broyden's method or its modification, may cycle and diverge. On the other hand, Theorem 2.1 actually shows that the modification is still locally and superlinearly convergent. This follows because the proof of Theorem 2.1 shows that $\epsilon > 0$ and $\delta > 0$ can be chosen so that if $\|x_0 - x^*\| < \epsilon$ and $\|B_0 - F'(x^*)\| < \delta$, then $\|B_k - F'(x^*)\| < 2\delta$ for all $k \geq 0$. Certainly δ can be further restricted so that $\theta_k = 1$ satisfies either (3.3) or (3.4). However, even if $\theta_k = 1$ is not chosen, we still have local and superlinear convergence.

THEOREM 3.4. *Let $F: R^n \rightarrow R^n$ satisfy the assumptions of Theorem 2.1, and consider Broyden's method as defined by (2.1), (2.3), (3.2) and (3.4). Then Broyden's method is locally and superlinearly convergent at x^* .*

This result follows from a modification to the proof of Theorem 2.1 as given by Broyden, Dennis and Moré [3], so we will omit its proof. Note that since Theorem 3.4 lets us choose any θ_k which satisfies (3.4), this gives a certain amount of stability to Broyden's method.

4. Powell's Hybrid Method. In view of Example 3.3, Broyden's method must be modified in order to achieve global convergence. In this section we outline a modification due to Powell [8] which achieves this aim; for a more thorough presentation see the original papers [8], [9].

Powell's hybrid method was designed to find solutions to $F(x) = 0$ where $F: R^n \rightarrow R^n$ is continuously differentiable in some open set D , but we are not able or willing to calculate the Jacobian matrix. Basically, the method attempts to minimize the functional $\psi: R^n \rightarrow R^n$ defined by

$$(4.1) \quad \psi(x) = (1/2)\|F(x)\|^2,$$

while making full use of the form of ψ .

At the beginning of the k th iteration, we have the iterate x_k , an approximation J_k to $F'(x_k)$ and a step-bound Δ_k such that the quadratic

$$(4.2) \quad \Phi_k(p) = (1/2)\|F(x_k) + J_k p\|^2$$

is a good approximation to $\psi(x_k + p)$ for $\|p\| \leq \Delta_k$. Below we specify how a correction p_k is determined with $\|p_k\| \leq \Delta_k$. Once this is done then the next iterate is given by

$$(4.3) \quad \begin{aligned} x_{k+1} &= x_k + p_k && \text{if } \psi(x_k + p_k) < \psi(x_k), \\ &= x_k && \text{otherwise.} \end{aligned}$$

Thus to complete the description of the k th iteration we need to define p_k , Δ_{k+1} , and J_{k+1} .

The correction p_k can either be chosen by an ordinary iteration or by a special iteration. The idea behind the choice of p_k in an ordinary iteration is that the Newton direction of F

$$(4.4) \quad p_k^N = -J_k^{-1}F(x_k)$$

is suitable if $\|p_k^N\| \leq \Delta_k$. Otherwise p_k should be chosen as a convex combination of p_k^N and some multiple of the steepest (or gradient) direction of Φ_k ,

$$(4.5) \quad p_k^G = -J_k^T F(x_k),$$

which reduces Φ_k in some sense. Hence, if $\|p_k^N\| \leq \Delta_k$ then $p_k = p_k^N$, but if $\|p_k^N\| > \Delta_k$ we examine Φ_k on the ray along p_k^G . It is not difficult to verify that on this ray Φ_k stops decreasing at

$$(4.6) \quad \hat{p}_k = (\|p_k^G\|/\|J_k p_k^G\|)^2 p_k^G.$$

Thus, if $\|\hat{p}_k\| \geq \Delta_k$ then it is reasonable to choose

$$p_k = \Delta_k p_k^G / \|p_k^G\|.$$

If $\|\hat{p}_k\| < \Delta_k$ we can decrease Φ_k further by proceeding toward p_k^N ; in this case we choose p_k as the convex combination of \hat{p}_k and p_k^N which minimizes Φ_k subject to $\|p_k\| \leq \Delta_k$. Hence, if $\|p_k^N\| > \Delta_k$ and $\|\hat{p}_k\| < \Delta_k$, then p_k is determined by finding $\alpha \in (0, 1)$ such that

$$\|(1 - \alpha)\hat{p}_k + \alpha p_k^N\| = \Delta_k$$

and setting $p_k = (1 - \alpha)\hat{p}_k + \alpha p_k^N$. To change the step-bound Δ_k in an ordinary iteration we test whether or not

$$(4.7) \quad \psi(x_k) - \psi(x_k + p_k) \geq \rho[\Phi_k(0) - \Phi_k(p_k)],$$

where $\rho \in (0, 1)$ is a given constant. If (4.7) holds, then the iteration is *successful* and $\Delta_{k+1} \in [\Delta_k, \mu\Delta_k]$ for some $\mu \geq 1$. Otherwise the iteration is *unsuccessful* and $\Delta_{k+1} \in [\rho_1\Delta_k, \rho_2\Delta_k]$, where $\rho_1 \leq \rho_2 < 1$. In the program given by Powell [9] the values $\rho = 0.1$, $\mu = 2$, $\rho_1 = \rho_2 = 0.5$ are used.

This completes the description of how the correction vector p_k is calculated and how the step-bound Δ_{k+1} is changed in an ordinary iteration. Special iterations are needed because sometimes it is convenient not to define p_k by the procedure outlined above. This is particularly true if J_{k+1} is determined from J_k by Broyden's method; see the discussion after Eq. (4.8). At this point the particular method for determining p_k in a special iteration is not important but we assume that $\|p_k\| \leq \Delta_k$ and that at most n consecutive special iterations are necessary. Finally, in a special iteration $\Delta_{k+1} = \Delta_k$.

The matrix J_{k+1} is determined from x_k, p_k and J_k in such a way that for some fixed $\gamma > 0$,

$$(4.8) \quad \begin{aligned} & \text{(a) } J_k \text{ is nonsingular and } \|J_k\| \leq \gamma \text{ for } k \geq 0. \\ & \text{(b) If } \{x_k\} \text{ converges to } x \text{ in } D \text{ and } \{p_k\} \\ & \quad \text{converges to zero, then } \{J_k\} \text{ converges to } F'(x). \end{aligned}$$

There are several ways to define $\{J_k\}$ so that (4.8) is satisfied. We will be particularly interested if it is determined by Broyden's method:

$$(4.9) \quad J_{k+1} = J_k + \theta_k \frac{[F(x_k + p_k) - F(x_k) - J_k p_k] p_k^T}{\|p_k\|^2}$$

and θ_k is chosen so that J_{k+1} is nonsingular and $|\theta_k - 1| \leq \theta$. In this case, however, (4.8) does not hold unless careful use is made of the special iterations. For example, if the sequence $\{p_k\}$ does not span R^n , then there is a $v \neq 0$ with $\langle v, p_k \rangle = 0$ for $k \geq 0$ and then (4.9) implies that $J_k v = J_0 v$. Hence, (4.8)(b) will not hold unless the choice of J_0 was somewhat fortunate. On the other hand, in the next section we prove that if the special iterations are used to guarantee that $\{p_k\}$ satisfies a uniform linear independence condition, then (4.8) holds.

One way to guarantee that $\{p_k\}$ satisfies a uniform linear independence condition is to choose, at periodic intervals, p_k to be a suitable multiple of a unit basis vector so that for some integer $m \geq n$,

$$\{e_1, \dots, e_n\} \subset \left\{ \pm \frac{p_{k+1}}{\|p_{k+1}\|}, \dots, \pm \frac{p_{k+m}}{\|p_{k+m}\|} \right\}.$$

If this strategy is used in connection with (4.9) then this amounts to replacing, at periodic intervals, a column of J_k by a divided difference. To see this, note that if $\theta_k = 1$ and $p_k = \eta e_j$ is used in (4.9), then the j th column of J_k is replaced by

$$[F(x_k + \eta e_j) - F(x_k)] / \eta$$

and the other columns of J_k are unchanged. An equivalent method would just replace, at periodic intervals, J_k by a divided difference approximation to $F'(x_k)$.

Of course, it will not be possible to define J_{k+1} so that (4.8) is satisfied unless F' is bounded on a set which contains the iterates. With this in mind, note that $x_{k+1} \in L$ where

$$L = \{x \in D: \|F(x)\| \leq \|F(x_0)\|\},$$

but that $x_k + p_k$ may not lie in L . Therefore, the algorithm requires a $\Delta > 0$ such that if

$$L_\Delta = \{y \in R^n: \|y - x\| \leq \Delta \text{ for some } x \in L\},$$

then $L_\Delta \subset D$, and in all cases Δ_{k+1} is not allowed to exceed Δ . Note that $L_\Delta \subset D$ is automatically satisfied if $D = R^n$ while if D is open but otherwise arbitrary and L is compact, then there is always a $\Delta > 0$ such that $L_\Delta \subset D$.

In what follows, Powell's hybrid method refers to the algorithm outlined above, where in particular, the sequence $\{J_k\}$ satisfies (4.8). The main convergence theorem for this algorithm is due to Powell [8, Theorem 5].

THEOREM 4.1. *Let D be an open set such that $L_\Delta \subset D$ and assume that $F: R^n \rightarrow R^n$ is continuously differentiable on D and F' is bounded on L_Δ . Then for each $\epsilon > 0$ Powell's hybrid method produces a k such that $\|J_k^T F(x_k)\| < \epsilon$.*

Since the main purpose of Powell's theorem is to prove that the algorithm terminates, he leaves some questions unanswered. For example, if $\{x_k\}$ converges to some x^* in D , does it follow that $F'(x^*)^T F(x^*) = 0$? Also, if $F'(x^*)$ is nonsingular (and hence $F(x^*) = 0$), at what rate does $\{x_k\}$ converge to x^* ? In the remainder of this section we answer these two questions.

THEOREM 4.2. *Let $F: R^n \rightarrow R^n$ satisfy the assumptions of Theorem 4.1 on the open set D . If the sequence $\{x_k\}$ generated by Powell's hybrid method converges to some x^* in D , then $F'(x^*)^T F(x^*) = 0$.*

Proof. We first assume that there is an infinite number of successful Newton iterations. In this case, since (4.7) implies that

$$\|F(x_{k+1})\| \leq (1 - \rho)\|F(x_k)\|$$

whenever the k th iteration is a successful Newton iteration, and since for all iterations $\|F(x_{k+1})\| \leq \|F(x_k)\|$, it follows that if there are an infinite number of successful Newton iterations then $\{\|F(x_k)\|\}$ converges to zero and hence, $F(x^*) = 0$.

Suppose now that there is a $k_0 > 0$ such that if $k \geq k_0$ then the k th iteration is not a successful Newton iteration. In this case, if k corresponds to an ordinary iteration, then $\Delta_{k+1} \leq \rho_2 \Delta_k$ if the iteration is unsuccessful, or $\Delta_{k+1} \leq \mu \Delta_k$ if the iteration is successful. Moreover, in the latter instance $\Delta_k = \|x_{k+1} - x_k\|$ since $p_k \neq p_k^N$. Hence, in an ordinary iteration k with $k \geq k_0$,

$$\Delta_{k+1} \leq \max\{\rho_2 \Delta_k, \mu \|x_{k+1} - x_k\|\}.$$

A special iteration sets $\Delta_{k+1} = \Delta_k$, and there are at most n consecutive special iterations. Thus, since $\{\|x_{k+1} - x_k\|\}$ converges to zero, it follows that $\{\Delta_k\}$ and hence, $\{p_k\}$ converges to zero. Now (4.8) guarantees that $\{J_k\}$ converges to $F'(x^*)$, and then Theorem 4.1 shows that $F'(x^*)^T F(x^*) = 0$ as desired.

If we assume in Theorem 4.2 that $F'(x^*)$ is nonsingular, then $F(x^*) = 0$. The following result shows that in this case the sequence $\{x_k\}$ will usually converge R -superlinearly to x^* .

THEOREM 4.3. *Let $F: R^n \rightarrow R^n$ satisfy the assumptions of Theorem 4.1 on the open set D , and assume that the sequence $\{x_k\}$ generated by Powell's hybrid method converges to a point x^* in D and that $\{p_k\}$ converges to zero. If $F'(x^*)$ is nonsingular, then $F(x^*) = 0$ and $\{x_k\}$ converges R -superlinearly to x^* .*

Proof. That $F(x^*) = 0$ follows from Theorem 4.2. Moreover, since we have assumed that $\{p_k\}$ converges to zero, (4.8) guarantees that $\{J_k\}$ converges to $F'(x^*)$, and since $F'(x^*)$ is nonsingular, there is a $\sigma > 0$ such that $\|J_k^{-1}\| \leq \sigma$. It also follows that if

$$\eta_k = \sup\{\|F'(x_k + tp_k) - J_k\|: 0 \leq t \leq 1\},$$

then $\{\eta_k\}$ converges to zero.

For the most part, the proof consists of showing that eventually all the ordinary iterations are successful. This means that there is a $k_1 > 0$ such that if k is an ordinary iteration and $k \geq k_1$, then

$$(4.10) \quad x_{k+1} = x_k + p_k, \quad \Delta_{k+1} \geq \Delta_k.$$

But $\Delta_{k+1} = \Delta_k$ for special iterations so that $\Delta_{k+1} \geq \Delta_k$ for all $k \geq k_1$. In particular, since $\|J_k^{-1}\| \leq \sigma$ it follows that

$$\|p_k^N\| \leq \sigma \|F(x_k)\| \leq \Delta_k,$$

and thus all ordinary iterations eventually choose $p_k = p_k^N$. The first equation in (4.10) now shows that

$$\|F(x_{k+1})\| = \|F(x_{k+1}) - F(x_k) - J_k p_k\| \leq \eta_k \|p_k\| \leq \sigma \eta_k \|F(x_k)\|,$$

where k corresponds to an ordinary iteration. Since $\{\eta_k\}$ converges to zero, given ϵ in $(0, 1)$ there is a $k_2 > 0$ such that $\sigma \eta_k \leq \epsilon$ for $k \geq k_2$. Now recall that $\|F(x_{k+1})\| \leq \|F(x_k)\|$ in all cases and that we have assumed that there is at least one ordinary iteration in each set of $n + 1$ consecutive iterations. Hence, if $l \leq (k - k_2)/m$ where $m = n + 1$ then

$$\|F(x_k)\| \leq \epsilon \|F(x_{k-m})\| \leq \dots \leq \epsilon^l \|F(x_{k-lm})\|,$$

so that if $l = l(k)$ is the largest integer that does not exceed $(k - k_2)/m$, then

$$\limsup_{k \rightarrow +\infty} \|F(x_k)\|^{1/k} \leq \limsup_{k \rightarrow +\infty} \epsilon^{l(k)} \|F(x_{k_2})\|^{1/k} \leq \epsilon^{1/m}.$$

Since $\epsilon > 0$ was arbitrary, it follows that

$$(4.11) \quad \lim_{k \rightarrow +\infty} \|F(x_k)\|^{1/k} = 0,$$

and since $\{x_k\}$ converges to x^* and $F'(x^*)$ is nonsingular, (4.11) implies that $\{x_k\}$ converges R -superlinearly to x^* .

To complete the proof it is only necessary to show that eventually all the ordinary iterations are successful. For this we first prove that if k corresponds to an ordinary iteration then

$$(4.12) \quad \psi(x_k) - \Phi_k(p_k) \geq (1/2) \|p_k^G\| \min \left\{ \Delta_k, \frac{\|p_k^G\|^3}{\|J_k p_k^G\|^2} \right\},$$

$$(4.13) \quad \psi(x_k + p_k) - \Phi_k(p_k) \leq \eta_k \|p_k\| \{(\eta_k/2) \|p_k\| + \|F(x_k) + J_k p_k\|\},$$

where ψ and Φ_k are defined by (4.1) and (4.2), respectively. To prove (4.12) note that

if $0 \leq \lambda \leq \Delta_k$ then

$$\Phi_k(p_k) \leq \Phi_k(\lambda p_k^G / \|p_k^G\|).$$

In particular, if $\lambda_k = \min\{\Delta_k, \|\hat{p}_k\|\}$ where \hat{p}_k is defined by (4.6) then

$$(4.14) \quad \psi(x_k) - \Phi_k(p_k) \geq \Phi_k(0) - \Phi_k(\lambda_k p_k^G / \|p_k^G\|).$$

To estimate the right side of this inequality note that for any $\lambda \geq 0$

$$\Phi_k(0) - \Phi_k\left(\lambda \frac{p_k^G}{\|p_k^G\|}\right) = \lambda \|p_k^G\| \left[1 - \left(\frac{\lambda}{2}\right) \frac{\|J_k p_k^G\|^2}{\|p_k^G\|^3} \right],$$

and since $\lambda_k \|J_k p_k^G\|^2 \leq \|p_k^G\|^3$,

$$\Phi_k(0) - \Phi_k(\lambda_k p_k^G / \|p_k^G\|) \geq (\lambda_k / 2) \|p_k^G\|.$$

It now follows from (4.14) and the definitions of λ_k and \hat{p}_k that (4.12) holds.

To prove (4.13) note that

$$\|F(x_k + p_k)\| - \|F(x_k) + J_k p_k\| \leq \eta_k \|p_k\|,$$

and since $\alpha^2 - \beta^2 \leq |\alpha - \beta| \{|\alpha - \beta| + 2|\beta|\}$, inequality (4.13) follows immediately from the definitions of ψ and Φ .

It is now easy to prove that eventually all the ordinary iterations are successful. Note that $\|p_k\| \leq \|p_k^N\|$ and therefore,

$$(4.15) \quad \|p_k\| \leq \sigma^2 \|p_k^G\|.$$

Hence (4.12) implies that

$$(4.16) \quad \psi(x_k) - \Phi_k(p_k) \geq (1/2) \|p_k^G\| \|p_k\| \min\{1, (\sigma\gamma)^{-2}\}.$$

Next note that $\|F(x_k)\| \leq \sigma \|p_k^G\|$ so that (4.13) and (4.15) imply that

$$(4.17) \quad \psi(x_k + p_k) - \Phi_k(p_k) \leq \eta_k \|p_k\| \|p_k^G\| \{((\eta_k/2) + \gamma)\sigma^2 + \sigma\}.$$

It is now clear from (4.16) and (4.17) that there is an index $k_1 > 0$ such that for $k \geq k_1$,

$$(1 - \rho) [\psi(x_k) - \Phi_k(p_k)] \geq \psi(x_k + p_k) - \Phi_k(p_k),$$

or equivalently,

$$\psi(x_k) - \psi(x_k + p_k) \geq \rho [\Phi_k(0) - \Phi_k(p_k)].$$

This shows that eventually all the ordinary iterations of the algorithm are successful and concludes the proof.

Theorem 4.3 assumes that the sequence $\{p_k\}$ converges to zero. At first sight it would seem that this follows from the fact that $\{x_k\}$ converges, but the following example shows that if the choice of $\{J_k\}$ is careless enough, then $\{p_k\}$ may not converge to zero.

Example 4.4. Let $f: R \rightarrow R$ be defined by $f(x) = x$, and consider Powell's hybrid method with $\rho = \rho_1 = \rho_2 = 0.5$ and $\mu = 2$. Assume that $x_k = (1/2)^k$ and $\Delta_k \geq 0.5$; we will show that it is possible to choose $\{J_k\}$ so that $x_{k+2} = (1/2)^{k+1}$ and $\Delta_{k+2} = \Delta_k$.

To see this, note that if $J_k = 2$ then $p_k = p_k^N$ is successful and hence, $x_{k+1} = (1/2)^{k+1}$. Since the iteration is successful, we are allowed to take $\Delta_{k+1} = 2\Delta_k$. Now choose $J_{k+1} = -2^{-k}$. Then $p_k = p_k^N$, but in this case the step is unsuccessful and moreover, $x_{k+2} = x_{k+1}$ and $\Delta_{k+2} = (1/2)\Delta_{k+1}$. Hence, $x_{k+2} = (1/2)^{k+1}$ and $\Delta_{k+2} = \Delta_k$ so that the same pattern can be repeated. Thus $\{x_k\}$ converges to $x^* = 0$ but the rate of convergence is not superlinear.

Example 4.4 shows that some subsequence of $\{\|p_k\|\}$ can be bounded away from zero and therefore, (4.8) does not force $\{J_k\}$ to converge to $F'(x^*)$. There are several ways to remedy the situation.

One solution consists of setting $J_{k+1} = J_k$ if $\|F(x_k + p_k)\| > \|F(x_k)\|$ in the ordinary iterations, and taking care, in the special iterations, that the correction vector p_k converges to zero if $\{x_k\}$ converges. For example, this can be done by taking $\|p_k\| = O(\|x_k - x_{k-1}\|)$. This modification would guarantee that $\{J_k\}$ is only calculated from a sequence $\{p_k\}$ which converges to zero if $\{x_k\}$ converges.

Another solution would be to modify the definition of the step-bound and require

$$\Delta_{k+1} \in [\rho_1 \|p_k\|, \rho_2 \|p_k\|]$$

if the k th step is unsuccessful, and $\Delta_{k+1} \in [\|p_k\|, \mu\|p_k\|]$ if the step is successful. In this case

$$\Delta_{k+1} \leq \max\{\rho_2 \Delta_k, \mu\|x_{k+1} - x_k\|\},$$

and thus if $\{x_k\}$ converges then Δ_k converges to zero and hence, $\{p_k\}$ must also converge to zero. This choice of step-bound is analogous to the one used in Powell's [10] hybrid method for unconstrained minimization.

Note that the second solution is actually a modification of the hybrid method and thus requires new proofs for the previous three results. This can be done, so from a theoretical point of view both modifications appear to be equally reasonable. Therefore, it would be interesting to compare numerically the above two approaches.

5. Uniform Linear Independence. The purpose of this section is to study the concept of uniform linear independence and to show that most of the definitions available in the literature are, in fact, equivalent. As our starting point, we take a definition that Ortega and Rheinboldt [7] used in the study of iterative methods for unconstrained minimization.

Definition 5.1. A sequence of unit vectors $\{u_j\}$ in R^n is uniformly linearly independent if there is a $\beta > 0$, a $k_0 \geq 0$ and an $m \geq n$, such that for $k \geq k_0$ and $\|x\| = 1$,

$$\max\{|\langle x, u_j \rangle|: j = k + 1, \dots, k + m\} \geq \beta.$$

This definition requires that each set of m consecutive vectors in the sequence $\{u_j\}$ spans R^n . However, it requires more. For example, if $u_{2k} = (1/\sqrt{k^2 + 1})(k, 1)$ and $u_j = (1, 0)$ for j odd, then each $m = 2$ consecutive vectors spans R^2 but this sequence is not uniformly linearly independent.

Also note that the term "uniformly linearly independent" is a misnomer since, of course, the sequence $\{u_j\}$ is not linearly independent. It would be better to say that

$\{u_j\}$ spans R^n uniformly if the sequence satisfies Definition 5.1.

LEMMA 5.2. *Let u_1, \dots, u_m be unit vectors in R^n and assume that $\theta_j \in (0, 2)$ for $1 \leq j \leq m$. Then $\{u_1, \dots, u_m\}$ spans R^n if and only if*

$$\left\| \prod_{j=1}^m [I - \theta_j u_j u_j^T] \right\| < 1.$$

Proof. Assume first that $\{u_1, \dots, u_m\}$ does not span R^n , and let

$$P = \prod_{j=1}^m [I - \theta_j u_j u_j^T].$$

Then there is an $x \neq 0$ such that $\langle x, u_j \rangle = 0$ for $1 \leq j \leq m$, and thus $Px = x$. In particular, $\|P\| \geq 1$.

Assume now that $\{u_1, \dots, u_m\}$ spans R^n . To show that $\|P\| < 1$ choose z_1 in R^n and define

$$(5.1) \quad z_{j+1} = (I - \theta_j u_j u_j^T) z_j, \quad j = 1, \dots, m,$$

so that $z_{m+1} = Pz_1$. We now show that $\|z_{m+1}\| < 1$ if $\|z_1\| = 1$. For this note that

$$\|z_{j+1}\|^2 = \|z_j\|^2 - \theta_j(2 - \theta_j)\langle u_j, z_j \rangle^2.$$

In particular, $\|z_{j+1}\| \leq \|z_j\|$, so that if $\|z_{m+1}\| = 1$ then $\|z_{j+1}\| = \|z_j\|$, and by the previous relationship $\langle u_j, z_j \rangle = 0$ for $1 \leq j \leq m$. Thus (5.1) implies $z_{j+1} = z_j$ for $1 \leq j \leq m$ and therefore,

$$\langle u_j, z_j \rangle = \langle u_j, z_1 \rangle = 0, \quad 1 \leq j \leq m.$$

Since $\{u_1, \dots, u_m\}$ spans R^n it follows that $z_1 = 0$. This contradicts the assumption $\|z_1\| = 1$ and therefore proves the result.

Lemma 5.2 is closely related to a result of Powell's [11, Theorem 6] in which he shows that the special iterations generated by his algorithm satisfy (5.2)(b) below. In the result that follows A^+ denotes the generalized inverse $A^T(AA^T)^{-1}$ of an n by m matrix of rank n .

THEOREM 5.3. *Let $\{u_k\}$ be a sequence of unit vectors in R^n . Then the following conditions are equivalent.*

(a) *The sequence $\{u_k\}$ is uniformly linearly independent.*

(b) *For any $\hat{\theta} \in [0, 1)$ there is a constant $\alpha \in (0, 1)$ such that if $|\theta_j - 1| \leq \hat{\theta}$, then*

$$(5.2) \quad \left\| \prod_{j=k+1}^{k+m} [I - \theta_j u_j u_j^T] \right\| \leq \alpha, \quad k \geq k_0.$$

(c) *There is a constant $\gamma > 0$ such that for each $\|x\| = 1$ and $k \geq k_0$ there are coefficients $\eta_j(x)$ such that for*

$$(5.3) \quad x = \sum_{j=k+1}^{k+m} \eta_j(x) u_j, \quad |\eta_j(x)| \leq \gamma.$$

(d) *If the n by m matrix $A_{k,m}$ is defined by*

$$(5.4) \quad A_{k,m} = [u_{k+1}, \dots, u_{k+m}],$$

then there is a constant $\mu > 0$ such that for $k \geq k_0$, $A_{k,m}$ has full rank and $\|A_{k,m}^+\| \leq \mu$.

Proof. Assume first that $\{u_k\}$ is uniformly linearly independent according to Definition 5.1, and let us show that (5.2) holds. If not, there is a subsequence $\{k_i\}$ such that

$$(5.5) \quad \lim_{i \rightarrow \infty} \left\| \prod_{j=k_i+1}^{k_i+m} [I - \theta_j u_j u_j^T] \right\| = 1.$$

A compactness argument now shows that there is a subsequence of $\{k_i\}$ (without loss of generality we assume that it is the full sequence) such that u_{k_i+j} and θ_{k_i+j} converge for $1 \leq j \leq m$. If u_j^* and θ_j^* are the values to which they converge, then (5.5) implies that

$$\left\| \prod_{j=1}^m [I - \theta_j^* u_j^* u_j^{*T}] \right\| = 1.$$

Since $\theta_j^* \in (0, 2)$, Lemma 5.2 implies that $\{u_1^*, \dots, u_m^*\}$ do not span R^n . However, Definition 5.1 implies that

$$\max\{|\langle x, u_j^* \rangle| : j = 1, \dots, m\} \geq \beta,$$

and this in turn implies that u_1^*, \dots, u_m^* span R^n . This contradiction shows that (a) implies (b).

Assume now that (b) holds, and let $\hat{\theta} = 0$. Then there is a constant $\alpha \in (0, 1)$ such that for $k \geq k_0$,

$$\|P_k\| \equiv \left\| \prod_{j=k+1}^{k+m} [I - u_j u_j^T] \right\| \leq \alpha.$$

We now proceed as in the proof of Lemma 5.1 and define

$$(5.6) \quad z_{j+1} = (I - u_j u_j^T) z_j, \quad k+1 \leq j \leq k+m.$$

Then $z_{k+m+1} = P_k z_{k+1}$ and

$$z_{k+1} - z_{k+m+1} = \sum_{j=k+1}^{k+m} \langle u_j, z_j \rangle u_j.$$

But z_{k+1} is arbitrary, so for any given $\|x\| = 1$ it can be chosen so that $(I - P_k)z_{k+1} = x$. Thus the above expression implies that $x = \sum_{j=k+1}^{k+m} \langle u_j, z_j \rangle u_j$. To bound the coefficients $\langle u_j, z_j \rangle$ note that (5.6) implies that $\|z_{j+1}\| \leq \|z_j\|$ and thus,

$$|\langle u_j, z_j \rangle| \leq \|z_{k+1}\| \leq \|(I - P_k)^{-1}\| \leq (1 - \alpha)^{-1}.$$

Hence, (5.3) holds with $\gamma = (1 - \alpha)^{-1}$.

Assume that (c) holds and let $\|x\| = 1$ be given. Then (5.3) implies that A_k is of full rank and that

$$A_k z = x, \quad z = (\eta_{k+1}, \dots, \eta_{k+m}).$$

Since $\|A_k^+ x\| \leq \|z\|$ if $A_k z = x$, we have that $\|A_k^+ x\| \leq m^{1/2} \gamma$ and therefore, $\|A_k^+\| \leq m^{1/2} \gamma$. Thus (d) holds with $\mu = m^{1/2} \gamma$.

If (d) holds and $\|x\| = 1$, then since A_k has full rank, $x = A_k (A_k^+ x)$. Hence,

$$1 = \|x\|^2 = \langle A_k^T x, A_k^+ x \rangle \leq \mu \|A_k^T x\|.$$

It follows that (a) holds with $\beta = 1/(m^{1/2} \mu)$.

As noted before, (a) is due to Ortega and Rheinboldt [7]. Conditions (b) and (c)

were used by Powell [8], [10] in a hybrid strategy for unconstrained minimization and nonlinear equations, respectively, but Powell did not investigate the relationship between these two conditions. Finally, (d) seems to be new in the case $m > n$, although when $m = n$ it appears quite frequently.

It should also be clear that there are other variations of Theorem 5.3. In particular, (d) is equivalent to the existence of $\mu_1 > 0$ such that

$$(e) \quad \|(A_{km}A_{k,m}^T)^{-1}\| \leq \mu_1 \quad \text{for } k \geq k_0.$$

This follows from the fact that if the n by m matrix A is of full row rank then $\|A^+\|^2 = \|(AA^T)^{-1}\|$.

THEOREM 5.4. *Let $\{u_k\}$ be a sequence of unit vectors in R^n . Then the following conditions are equivalent:*

- (a) *The sequence $\{u_k\}$ is uniformly linearly independent with $m = n$.*
- (b) *There is a $\sigma > 0$ such that for $k \geq k_0$, $|\det A_{k,n}| \geq \sigma$.*
- (c) *There is a $\mu > 0$ such that for $k \geq k_0$, $\|A_{k,n}^{-1}\| \leq \mu$.*

Proof. Theorem 5.3 implies that (a) and (c) are equivalent. Now if (c) holds and λ is an eigenvalue of $A_{k,n}$, then $|\lambda| \geq 1/\mu$. Thus, $|\det A| \geq 1/\mu^n$ so that (b) holds with $\sigma = 1/\mu^n$.

To show that (b) implies (c), let $A_{k,n} = QL$ where L is lower triangular and Q is orthogonal. Then all the columns of L are of unit norm and thus, $|l_{ij}| \leq 1$. Moreover, since $|\det L| \geq \sigma$, we also have that $|l_{ii}| \geq \sigma$. Now, given x in R^n with $\|Lx\| = 1$ it follows, by induction, that $|\xi_j| \leq 2^{j-1}/\sigma^j$ where $x = (\xi_j)$. Hence, $\|x\| \leq (2/\sigma)^n$ and therefore,

$$\|A_{k,n}^{-1}\| = \|L^{-1}Q^{-1}\| = \|L^{-1}\| \leq (2/\sigma)^n.$$

Thus, (c) holds with $\mu = (2/\sigma)^n$.

To illustrate the usefulness of Theorem 5.3 we present simple proofs of the asymptotic behavior of the matrices generated by Broyden's update. For this purpose consider

$$(5.7) \quad J_{k+1} = J_k + \theta_k \frac{[y_k - J_k p_k] p_k^T}{\|p_k\|^2}, \quad y_k = F(x_k + p_k) - F(x_k),$$

where $F: R^n \rightarrow R^n$ is continuously differentiable in an open set D , and assume that the following conditions are satisfied.

- (a) The sequence $\{x_k\}$ remains in some set $D_0 \subset D$ and $|\theta_k - 1| \leq \theta$ for some $\theta \in (0, 1)$.
- (b) The sequence of nonzero vectors $\{p_k\}$ is uniformly linearly independent, and the line segment from x_k to $x_k + p_k$ lies in D_0 .

As an initial step in analyzing (5.7), we will need the following simple result.

LEMMA 5.5. *Let $\{\phi_k\}$ and $\{\delta_k\}$ be sequences of nonnegative numbers such that $\phi_{k+m} \leq \alpha \phi_k + \delta_k$ for some fixed integer $m \geq 1$ and $\alpha \in (0, 1)$. If $\{\delta_k\}$ is bounded, then $\{\phi_k\}$ is also bounded, and if in addition, $\{\delta_k\}$ converges to zero, then $\{\phi_k\}$ converges to zero.*

Proof. Assume first that $m = 1$. It can then be verified that

$$\phi_k \leq \alpha^k \phi_0 + \sum_{j=0}^{k-1} \alpha^{k-j} \delta_j,$$

so that if δ is a bound for $\{\delta_k\}$, then $\phi_k \leq \alpha^k \phi_0 + \delta(1 - \alpha)^{-1}$. It follows that $\{\phi_k\}$ is bounded; a similar argument shows that if $\{\delta_k\}$ converges to zero then $\{\phi_k\}$ also converges to zero. If $m > 1$ let $\hat{\phi}_k = \phi_{km+i}$ for any integer $0 \leq i \leq m - 1$. Then $\hat{\phi}_{k+1} \leq \alpha \hat{\phi}_k + \delta_{km+i}$, and thus the above argument shows that if $\{\delta_k\}$ is bounded, then $\{\hat{\phi}_k\}$ is bounded for any $0 \leq i \leq m - 1$, and therefore, $\{\phi_k\}$ is bounded. Similarly, if $\{\delta_k\}$ converges to zero, then $\{\phi_k\}$ also converges to zero.

The following two results are due to Powell [8], but since he used version (c) of Theorem 5.3, his proofs are quite involved.

THEOREM 5.6. *Let $F: R^n \rightarrow R^n$ be continuously differentiable on the open set D and assume that F' is bounded on some $D_0 \subset D$. If the sequence $\{J_k\}$ is defined by (5.7) and assumptions (5.8) hold, then $\{J_k\}$ is bounded.*

Proof. Equation (5.7) shows that

$$J_{k+1} = J_k Q_k + \theta_k \frac{y_k p_k^T}{\|p_k\|^2}, \quad Q_k = I - \theta_k \frac{p_k p_k^T}{\|p_k\|^2}.$$

Now note that $\|Q_k\| \leq 1$ and that $|\theta_k| \leq 2$ so that an induction argument on m yields

$$\|J_{k+m+1}\| \leq \|J_{k+1} Q_{k+1} \cdots Q_{k+m}\| + 2 \sum_{j=k+1}^{k+m} \frac{\|y_j\|}{\|p_j\|}.$$

If $\|F'(x)\| \leq \mu$ for x in D_0 , then $\|y_j\| \leq \mu \|p_j\|$. Therefore, Theorem 5.3 shows that there is a $k_0 \geq 0$ and $\alpha \in (0, 1)$ such that $\|J_{k+m+1}\| \leq \alpha \|J_{k+1}\| + 2m\mu$. The result now follows from Lemma 5.5.

For the application of this result to the hybrid method, $D_0 = L_\Delta$. Also note that Theorem 5.6 has applications to least squares methods since this result is unchanged if F maps R^n into R^p for some $p \neq n$. These same remarks apply to the next result.

THEOREM 5.7. *Let $F: R^n \rightarrow R^n$ be continuously differentiable in the open set D and consider the sequence $\{J_k\}$ defined by (5.7). If assumptions (5.8) hold, and in addition $\{x_k\}$ converges to some x in D and $\{p_k\}$ converges to zero, then $\{J_k\}$ converges to $F'(x)$.*

Proof. The proof of this result is very similar to that of Theorem 5.6. In fact, (5.7) shows that

$$J_{k+1} - F'(x) = [J_k - F'(x)] Q_k + \theta_k \frac{[y_k - F'(x)p_k] p_k^T}{\|p_k\|^2}.$$

Also note that $\|y_k - F'(x)p_k\| \leq \epsilon_k \|p_k\|$ where

$$\epsilon_k = \max\{\|F'(x_k + tp_k) - F'(x)\|: 0 \leq t \leq 1\}.$$

Thus an induction argument on m and Theorem 5.3 show that there is a $k_0 \geq 0$ and $\alpha \in (0, 1)$ such that

$$\|J_{k+m+1} - F'(x)\| \leq \alpha \|J_{k+1} - F'(x)\| + 2 \sum_{j=k+1}^{k+m} \epsilon_j.$$

Since $\{\epsilon_k\}$ converges to zero, the result follows from Lemma 5.5.

The proofs of Theorems 5.6 and 5.7 are similar to those presented by Powell [10] for the symmetric form of Broyden's update. However, here the assumptions are weaker,

and our formulation clearly shows that these results do not depend on the particular algorithms which generate $\{x_k\}$ and $\{p_k\}$.

Finally we note that Powell [8] and Schwetlick [13] discuss algorithms for maintaining uniform linear independence.

6. Powell's Symmetric Version of Broyden's Method. Assume as before that $F: R^n \rightarrow R^n$ is continuously differentiable on some open set D , but in addition, suppose that $F'(x)$ is symmetric for all x in D . In this case it is advantageous to modify update (3.2) so as to take into account the symmetry of F' . One such modification is due to Powell [10]:

$$(6.1) \quad B_{k+1} = B_k + \theta_k \frac{v_k s_k^T + s_k v_k^T}{\|s_k\|^2} - \theta_k^2 \frac{\langle v_k, s_k \rangle}{\|s_k\|^4} s_k s_k^T,$$

$$v_k = y_k - B_k s_k,$$

where the parameter θ_k is chosen so that B_{k+1} is nonsingular. In this section we point out how the results of the previous section are changed if (3.2) is replaced by (6.1).

For the motivation and derivation of this update in the case $\theta_k = 1$ we refer to the survey article [5]. In this paper we follow Powell [11] and outline how θ_k can be chosen so that B_{k+1} is nonsingular and (3.4) holds for some $\hat{\theta}$.

It is not too difficult to show (see, for example, Lemma 7.6 in [5]) that as a consequence of Lemma 3.1,

$$\det(I + uv^T + pq^T) = (1 + \langle u, v \rangle)(1 + \langle p, q \rangle) - \langle u, q \rangle \langle v, p \rangle.$$

From this identity and after some manipulation it follows that if $B_k = H_k^{-1}$ is nonsingular, then $\det B_{k+1} = \phi_k(\theta_k) \det B_k$ where

$$\phi_k(\theta) = 1 - 2\theta \frac{\langle s_k, H_k v_k \rangle}{\|s_k\|^2} + \theta^2 \frac{\langle s_k, H_k v_k \rangle^2 - \langle s_k, H_k s_k \rangle \langle v_k, H_k v_k + s_k \rangle}{\|s_k\|^4}.$$

Given σ in $(0, 1)$, Powell [11] chooses $\theta_k = 1$ if $|\phi_k(1)| \geq \sigma$, and otherwise, θ_k is chosen to be a number closest to unity such that $\phi_k(\theta) = \sigma$. An important point about this choice of θ_k is that

$$|\theta_k - 1| \leq [2\sigma/(\sigma + 1)]^{1/2}.$$

However, we emphasize that in this paper θ_k need only satisfy (3.4).

It is now natural to consider the symmetric analogue of Broyden's method in which the sequence $\{x_k\}$ is defined by (2.1), (2.3) and $\{B_k\}$ is generated by (6.1) and (3.4) with B_0 symmetric and nonsingular. This method is known as the Powell-symmetric-Broyden algorithm.

THEOREM 6.1. *Let $F: R^n \rightarrow R^n$ satisfy the assumptions of Theorem 2.1 and in addition, suppose that $F'(x^*)$ is symmetric. Then the Powell-symmetric-Broyden algorithm is locally and superlinearly convergent at x^* .*

This result is due to Broyden, Dennis and Moré [3] if $\theta_k \equiv 1$; if θ_k is just restricted by (3.4), the proof is very similar, so it is omitted.

Theorem 6.1 is the analogue of Theorem 3.4; the following result parallels Theorem 3.2.

THEOREM 6.2. *Let $F: R^n \rightarrow R^n$ be defined by $F(x) = Ax - b$ where A in $L(R^n)$ is symmetric and nonsingular. Then the Powell-symmetric-Broyden algorithm is globally and superlinearly convergent.*

Proof. If $E_k = B_k - A$ then (6.1) implies that

$$E_{k+1} = Q_k E_k Q_k, \quad Q_k = I - \theta_k s_k s_k^T / \|s_k\|^2,$$

and since $\|Q_k\| \leq 1$, estimate (1.2) yields that

$$\|E_{k+1}\|_F^2 \leq \|E_k Q_k\|_F^2.$$

However, in the proof of Theorem 3.2 we proved

$$\|E_k Q_k\|_F^2 = \|E_k\|_F^2 - \theta_k (2 - \theta_k) (\|E_k s_k\| / \|s_k\|)^2,$$

so that the last three estimates show that

$$(1 - \hat{\theta})^2 (\|E_k s_k\| / \|s_k\|)^2 \leq \|E_k\|_F^2 - \|E_{k+1}\|_F^2.$$

This inequality implies (3.5) and thus the proof proceeds as in Theorem 3.2.

It should now be clear that Theorems 5.6 and 5.7 remain essentially unchanged if $\{J_k\}$ is generated by

$$(6.2) \quad J_{k+1} = J_k + \theta_k \frac{v_k p_k^T + p_k v_k^T}{\|p_k\|^2} - \theta_k^2 \frac{\langle v_k, p_k \rangle}{\|p_k\|^4} p_k p_k^T,$$

$$v_k = F(x_k + p_k) - F(x_k) - J_k p_k,$$

with J_0 symmetric; the only difference is that now $F'(x)$ is assumed to be symmetric for x in D .

7. Concluding Remarks. It is interesting to compare our results with those obtained for Powell's 1970 hybrid method [10] for the unconstrained minimization of a functional f .

Powell [12] shows that if update (6.1) is used without special iterations, then there is global and superlinear convergence. However, his results do not apply to the functional ψ defined by (4.1) because he assumes that the gradient of the functional can be calculated exactly. It is an open question whether Theorems 4.1 and 4.2 hold for the sequence $\{J_k\}$ defined by (5.7) if no special iterations are performed, but our numerical experiments show that in most cases special iterations are numerically desirable.

Thomas [14] in his Ph.D. thesis shows that if Powell's 1970 hybrid method is slightly modified then, with special iterations, the sequence $\{\nabla f(x_k)\}$ converges to zero, while Powell only shows that this holds for some subsequence. It would be interesting to show that a similar result holds for the hybrid method of this paper so that in Theorem 4.1 we actually obtain the convergence of $\{\|J_k^T F(x_k)\|\}$ to zero.

Acknowledgment. We would like to thank M. J. D. Powell for his helpful comments on an earlier draft of this paper.

Applied Mathematics Division
Argonne National Laboratory
Argonne, Illinois 60539

Department of Mathematics
University of California, San Diego
LaJolla, California 92093

1. C. G. BROYDEN, "A class of methods for solving nonlinear simultaneous equations," *Math. Comp.*, v. 19, 1965, pp. 577–593. MR 33 #6825.
2. C. G. BROYDEN, "The convergence of single-rank quasi-Newton methods," *Math. Comp.*, v. 24, 1970, pp. 365–382. MR 43 #5714.
3. C. G. BROYDEN, J. E. DENNIS & J. J. MORE, "On the local and superlinear convergence of quasi-Newton methods," *J. Inst. Math. Appl.*, v. 12, 1973, pp. 223–245. MR 49 #6599.
4. J. E. DENNIS & J. J. MORE, "A characterization of superlinear convergence and its application to quasi-Newton methods," *Math. Comp.*, v. 28, 1974, pp. 549–560. MR 49 #8322.
5. J. E. DENNIS & J. J. MORE, *Quasi-Newton Methods, Motivation and Theory*, Cornell University Computer Science Technical Report TR 74–217, 1974.
6. P. E. GILL & W. MURRAY, "Quasi-Newton methods for unconstrained minimization," *J. Inst. Math. Appl.*, v. 9, 1972, pp. 91–108. MR 45 #9456.
7. J. M. ORTEGA & W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970. MR 42 #8686.
8. M. J. D. POWELL, "A hybrid method for nonlinear equations," *Numerical Methods for Non-Linear Algebraic Equations*, Gordon and Breach, London, 1970, pp. 87–114. MR 49 #8330a.
9. M. J. D. POWELL, "A FORTRAN subroutine for solving systems of nonlinear algebraic equations," *Numerical Methods for Non-Linear Algebraic Equations*, Gordon and Breach, London, 1970, pp. 115–161. MR 49 #8330b.
10. M. J. D. POWELL, "A new algorithm for unconstrained optimization," *Nonlinear Programming*, Academic Press, New York, 1970, pp. 31–65. MR 42 #7043.
11. M. J. D. POWELL, *A FORTRAN Subroutine for Unconstrained Minimization, Requiring First Derivatives of the Objective Function*, Atomic Energy Research Establishment, Harwell, R.6469, 1970.
12. M. J. D. POWELL, "Convergence properties of a class of minimization algorithms," *Nonlinear Programming 2*, Academic Press, New York, 1974, pp. 1–27.
13. H. SCHWETLICK, "Über die Realisierung und Konvergenz von Mehrschrittverfahren zur iterativen Lösung nichtlinearer Gleichungen," *Z. Angew. Math. Mech.*, v. 54, 1974, pp. 479–493.
14. S. M. THOMAS, *Sequential Estimation Techniques for Quasi-Newton Algorithms*, Cornell University Ph.D. Thesis, 1974.