A CHARACTERIZATION OF SUPERLINEAR CONVERGENCE AND ITS APPLICATION TO QUASI-NEWTON METHODS[†]

J.E. Dennis, Jr.

Jorge J. Moré

TR 73-157

January 1973

Department of Computer Science
Cornell University
Ithaca, New York 14850

[†]Research Supported by NSF GJ-27528

A Characterization of Superlinear Convergence and its Application to Quasi-Newton Methods

by

J. E. Dennis, Jr.

Jorge J. Moré

Abstract

Let F be a mapping from real n-dimensional Euclidean space into itself. Most practical algorithms for finding a zero of F are of the form

$$x_{k+1} = x_k - B_k^{-1}Fx_k$$

where $\{B_k\}$ is a sequence of non-singular matrices. The main result of this paper is a characterization theorem for the superlinear convergence to a zero of F of sequences of the above form. This result is then used to give a unified treatment of the results on the superlinear convergence of the Davidon-Fletcher-Powell method obtained by Powell for the case in which exact line searches are used, and by Broyden, Dennis, and Moré for the case without line searches. As a by-product, several results on the asymptotic behavior of the sequence $\{B_k\}$ are obtained.

An interesting aspect of these results is that superlinear convergence is obtained without any consistency conditions; i.e. without requiring that the sequence $\{B_k\}$ converge to the Jacobian

1. Introduction.

This paper will be concerned with iterations of the form

$$(1.1) \qquad x_{k+1}=x_k-B_k^{-1}Fx_k, \quad k=0,1,\ldots,$$

for the solution of $Fx=0$, where $F:R^n \to R^n$ is a mapping of Euclidean n-space R^n into itself. Each nonsingular matrix B_k is directly or indirectly (as in various quasi-Newton methods) intended to approximate $F'(x_k)$, the Jacobian matrix of F at x_k. Virtually all practical methods for generating a sequence $\{x_k\}$ of approximate zeros of F have the basic form (1.1) although they are often implemented in their damped version,

$$(1.2) \qquad x_{k+1}=x_k-\lambda_k B_k^{-1}Fx_k, \quad k=0,1,\ldots,$$

where λ_k is a scalar chosen to prevent divergence of the sequence $\{x_k\}$. Even in this case, the hope is always that $\lambda_k=1$ will suffice near x^*, a root of F.

No matter which of (1.1) or (1.2) is used to generate $\{x_k\}$, the utility of the method is in large part dependent on the speed with which $\{x_k\}$ converges to x^*. In order to be practical a method should be at least Q-linear (for material on rates of convergence see [6 , Chapter 9]) when F satisfies reasonable hypotheses and $\{x_k\}$ converges to x^*, i.e. for some $r\in(0,1)$, $||x_{k+1}-x^*|| \leq r||x_k-x^*||$, $k=0,1,\cdots$. Furthermore, if the method is to be competitive then it should in fact be

Q-superlinear, i.e. $\{||x_{k+1}-x^*||/||x_k-x^*||\}$ converges to zero with k. Higher orders of convergence are desirable when obtainable without unreasonable additional computations although for general problems, orders higher than two are of minimal importance.

The standard technique for proving that a method is Q-superlinear has always been to show that as $\{x_k\}$ converges to x^*, $\{B_k\}$ converges to $F'(x^*)$. It is well known that this consistency condition is sufficient but not necessary. (See Voigt [10] for a compendium of interesting theorems and examples.) Since until recently all the practical methods for which there were published proofs of Q-superlinearity satisfied this property, it seemed to be effectively necessary.

Recently, M.J.D. Powell [7] proved that the Davidon-Fletcher-Powell method is Q-superlinear when implemented in the form (1.2) with λ_k chosen by the standard line minimization criterion. On the other hand, Broyden, Dennis and Moré [2] showed that many of the most important quasi-Newton methods are Q-superlinear when implemented in the form (1.1). Powell [8] has also furnished an example which shows that the celebrated Davidon-Fletcher-Powell method is indeed a practical Q-superlinear method which does not have the property that $\{B_k\}$ converges to $F'(x^*)$.

Although the Powell and Broyden, Dennis, Moré convergence proofs are quite distinct, the parts of both papers which deal with Q-superlinearity turns out to be based on the same principle. The main purpose of this paper is to enunciate that principle and some of its consequences.

In the next section we prove a characterization theorem for Q-superlinear convergence to a zero of F which applies to sequences generated by either (1.1) or (1.2). We also obtain a simple, but apparently new, justification for using $||x_{k+1}-x_k||$ as an estimate for $||x_k-x^*||$.

In Section 3 we apply our results to the Davidon-Fletcher-Powell method and show how the superlinear results of Powell and those of Broyden, Dennis and Moré can be obtained, while in Section 4 we indicate how these results can be applied to other quasi-Newton methods and we give Powell's example. We conclude with some general remarks. In particular, we compare our results to those obtained by McCormick and Ritter [5], and Ritter [9].

2. Characterization of Superlinear Convergence.

Let $||\cdot||$ denote an arbitrary vector norm in R^n or the operator norm it induces in $L(R^n)$ - the space of real matrices of order n. Now recall [6 , Chapter 9] that if $\{x_k\} \subset R^n$ converges to x^*, then $\{x_k\}$ converges Q-superlinearly to x^* if and only if either $x_k=x^*$ for all sufficiently large k or $x_k \neq x^*$ for $k \geq k_0$ and

$$\lim_{k\to+\infty} \frac{||x_{k+1}-x^*||}{||x_k-x^*||} = 0.$$

In this paper, when we assume that $\{x_k\}$ converges to x^*, we will also make the unstated assumption that $x_k \neq x^*$ for all sufficiently large k. In the context of (1.1) this simply means $x_{k+1} \neq x_k$ for any $k \geq 0$ and is thus quite reasonable.

With these preliminaries out of the way it is possible to proceed with the stated purpose of this section, the development of a characterization theorem for methods of the form (1.1) which are Q-superlinearly convergent. We begin with a result that gives a necessary condition for an arbitrary sequence.

Lemma 2.1 Let $\{x_k\} \subset R^n$ converge Q-superlinearly to x^*. Then

$$(2.1) \qquad \lim_{k\to+\infty} \frac{||x_{k+1}-x_k||}{||x_k-x^*||} = 1.$$

<u>Proof</u>: Just note that

$$\left| \frac{||x_{k+1} - x_k||}{||x_k - x^*||} - \frac{||x_k - x^*||}{||x_k - x^*||} \right| \le \frac{||x_{k+1} - x^*||}{||x_k - x^*||} .$$

The converse does not hold. For example, if $x_{2k-1} = (k!)^{-1}$ and $x_{2k} = 2x_{2k-1}$ for $k \geq 1$, then this sequence satisfies (2.1) with $x^*=0$, but does not converge Q-superlinearly to zero.

Lemma 2.1 justifies the very commonly used computational technique of estimating $||x_k - x^*||$ with $||x_{k+1} - x_k||$, if the underlying method is Q-superlinear.

To obtain a characterization of Q-superlinear convergence for (1.1) we will assume that F is defined and (Gateaux) differentiable in an open, convex set D; that is, its Jacobian matrix $F'(x)$ satisfies

$$\lim_{t\to 0} \frac{F(x+th) - F(x)}{t} = F'(x)h$$

for each x in D and h in R^n. See, for example, [6, Chapter 3]. This facilitates the proofs but the existence of $F'(x)$ is all that is really needed.

<u>Theorem 2.2</u> Let $F:R^n \to R^n$ be differentiable in the open, convex set D in R^n, and assume that for some x^* in D, F' is continuous at x^* and $F'(x^*)$ is nonsingular.

Let $\{B_k\}$ in $L(R^n)$ be a sequence of nonsingular matrices and suppose that for some x_0 in D the sequence $\{x_k\}$ where

$$x_{k+1} = x_k - B_k^{-1} Fx_k \tag{2.2}$$

remains in D and converges to x^*. Then $\{x_k\}$ converges Q-superlinearly to x^* and $F(x^*)=0$ if and only if

$$\lim_{k\to+\infty} \frac{||[B_k - F'(x^*)](x_{k+1} - x_k)||}{||x_{k+1} - x_k||} = 0. \tag{2.3}$$

<u>Proof</u>: Assume first that (2.3) holds. Since

$$\begin{aligned} [B_k - F'(x^*)](x_{k+1} - x_k) &= -Fx_k - F'(x^*)(x_{k+1} - x_k) \\ &= Fx_{k+1} - Fx_k - F'(x^*)(x_{k+1} - x_k) - Fx_{k+1}, \end{aligned} \tag{2.4}$$

the continuity of F' and x^* and (2.3) implies that

$$\lim_{k\to+\infty} \frac{||Fx_{k+1}||}{||x_{k+1} - x_k||} = 0. \tag{2.5}$$

Hence, $F(x^*)=0$, and since $F'(x^*)$ is nonsingular, there is a $\beta > 0$ such that

$$||Fx_{k+1}|| = ||Fx_{k+1} - Fx^*|| \geq \beta||x_{k+1} - x^*|| ,$$

and therefore,

$$\frac{||Fx_{k+1}||}{||x_{k+1} - x_k||} \geq \frac{\beta||x_{k+1} - x^*||}{||x_{k+1} - x^*|| + ||x_k - x^*||} = \beta \frac{\rho_k}{1 + \rho_k}$$

where

$$\rho_k = \frac{||x_{k+1} - x^*||}{||x_k - x^*||}$$

Thus (2.5) implies that $\rho_k/(1+\rho_k)$ converges to zero and hence $\{\rho_k\}$ also converges to zero as desired.

Conversely, assume that $\{x_k\}$ converges Q-superlinearly to x^* and $F(x^*)=0$. Since

$$\frac{||Fx_{k+1}||}{||x_{k+1} - x_k||} = \frac{||Fx_{k+1} - Fx^*||}{||x_k - x^*||} \cdot \frac{||x_k - x^*||}{||x_{k+1} - x_k||},$$

Lemma 2.1 and the hypotheses on F' imply that (2.5) holds. It then follows from (2.4) that condition (2.3) is satisfied.

In some cases it is very easy to verify that (2.3) holds. For example, in a Newton or discretized Newton method, $\{B_k\}$ converges to $F'(x^*)$ and hence, (2.3) holds. We will later show that (2.3) also holds for many quasi-Newton methods even though in this case $\{B_k\}$ does not necessarily converge to $F'(x^*)$ (See Section 4).

The Newton, discretized Newton, and quasi-Newton methods are sometimes implemented in the form (1.2) where the sequence $\{\lambda_k\}$ is chosen so as to enlarge the domain of convergence. The next result shows that this sequence will still be Q-superlinear if and only if $\{\lambda_k\}$ converges to unity.

<u>Corollary 2.3</u> Let $F:R^n \to R^n$ satisfy the hypotheses of Theorem 2.2 on the set D. Let $\{B_k\}$ be a sequence of nonsingular matrices, and suppose that for some x_0 in D the sequence $\{x_k\}$ generated by (1.2) remains in D and converges to x^*. If (2.3) holds then $\{x_k\}$ converges Q-superlinearly to x^* where $Fx^*=0$ if and only if $\{\lambda_k\}$ converges to unity.

<u>Proof</u>: Assume that $\{x_k\}$ converges Q-superlinearly to x^* and $F(x^*)=0$. By Theorem 2.2 we must have

$$(2.6) \qquad \lim_{k\to+\infty} \frac{\|[\lambda_k^{-1} B_k - F'(x^*)](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0,$$

and therefore (2.3) implies that

$$\lim_{k\to+\infty} \frac{\|(\lambda_k^{-1} - 1)\, B_k(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0.$$

But $B_k(x_{k+1} - x_k) = -\lambda_k\, Fx_k$, so that

$$\lim_{k\to+\infty} \frac{\|(\lambda_k - 1)\, Fx_k\|}{\|x_{k+1} - x_k\|} = 0.$$

Now $F'(x^*)$ is nonsingular, and thus there is a $\beta > 0$ such that $||Fx_k|| \geq \beta||x_k - x^*||$. Therefore, Lemma 2.1 implies that $\{\lambda_k\}$ must converge to unity.

Conversely, if $\{\lambda_k\}$converges to unity, it is clear from (2.3) that (2.6) holds. Hence, Theorem 2.2 guarantees that the sequence given by (1.2) converges Q-superlinearly to x^* where $Fx^*=0$.

If (2.3) holds, Corollary 2.3 explains why techniques for finding $\{\lambda_k\}$ must eventually produce values close to unity if there is to be Q-superlinear convergence to a zero of F. For example, in the often-quoted technique of Goldstein and Price [4] for discretized Newton methods, $\lambda_k = 1$ for all large k. Corollary 2.3 then explains why they were able to obtain Q-superlinear convergence. Later on, we will prove that a number of quasi-Newton methods also satisfy (2.3) so that the above remarks also apply to these methods.

Finally note that the results of this section can be used to prove that certain methods of the form (2.2) are not Q-superlinearly convergent to a zero of F. For example, in the "reset" methods one sets $B_{k_i}=B$ for some fixed nonsingular matrix, and some subsequence $\{k_i\}$. If $\{x_k\}$ converges Q-superlinearly to a zero of F, then Theorem 2.2 implies that (2.3) holds, and thus $Bs=F'(x^*)s$ for some vector s with $||s||=1$. In particular, 1 is an eigenvalue of $B^{-1}F'(x^*)$. This indicates that unless B is chosen with great care, resetting will prevent Q-superlinear convergence to a zero of F.

3. Superlinear Convergence of the Davidon - Fletcher - Powell Method.

For a mapping $F:R^n \to R^n$ the Davidon - Fletcher - Powell method is defined by

$$(3.1) \qquad x_{k+1} = x_k - \lambda_k H_k F x_k$$

where

$$(3.2) \qquad H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} ,$$

and

$$(3.3) \qquad \begin{aligned} s_k &= x_{k+1} - x_k, \\ y_k &= Fx_{k+1} - Fx_k. \end{aligned}$$

There are two results which guarantee that the Davidon - Fletcher - Powell method is Q-superlinearly convergent. The first result is due to Powell [7]; he assumes that (a) F is the gradient of a uniformly convex functional $f:R^n \to R^1$, (b) that λ_k is chosen so that

$$f(x_k - \lambda_k p_k) = \min\{f(x_k - \lambda p_k): \lambda \geq 0\}$$

where $p_k = H_k \nabla f(x_k)$, and (c) that F is a continously differentiable mapping on R^n which satisfies the one-sided

Lipschitz condition

$$||F'(x) - F'(x^*)|| \le L||x - x^*||$$

in the level set $\{x: f(x) \le f(x_0)\}$. Here x^* is the unique solution of $Fx = 0$. Under these hypotheses, Powell proved that for any x_0 in R^n and any symmetric, positive definite H_0, the Davidon - Fletcher - Powell method is Q-superlinearly convergent to x^*.

On the other hand, Broyden, Dennis, and Moré [2] assumed that (a) $F: R^n \rightarrow R^n$ is differentiable in an open convex neighborhood D of a point x^* for which $Fx^* = 0$ and $F'(x^*)$ is symmetric and positive definite, and that (b) F' satisfies the one-sided Lipschitz condition

$$||F'(x) - F'(x^*)|| \le K||x - x^*||^p \tag{3.4}$$

for some $p > 0$ and all x in D. With these hypotheses they proved that if (x_0, H_0) is sufficiently close to $(x^*, F'(x^*)^{-1})$, then the Davidon - Fletcher - Powell method with $\lambda_k = 1$ for all k is Q-superlinearly convergent to x^*.

These two convergence results are really more complementary than comparable and their proofs are quite distinct. We will show now that the results of Section 2 not only give a unified treatment of both rate of convergence proofs, but lead to the conclusion that Powell's minimizing

$\{\lambda_k\}$ converges to 1.

It is well known that if H_{k+1} is defined from H_k, s_k and y_k by (3.2) and if $H_k^{-1} = B_k$ exists, then as long as $y_k^T s_k \neq 0$, H_{k+1}^{-1} is defined by

$$(3.5) \qquad B_{k+1} = B_k + \frac{(y_k - B_k s_k)^T y_k + y_k(y_k - B_k s_k)^T}{y_k^T s_k} - \frac{s_k^T(y_k - B_k s_k) y_k y_k^T}{y_k^T s_k}$$

Note that $\{B_k\}$is defined provided the sequences $\{y_k\}$ and $\{s_k\}$ satisfy $y_k^T s_k \neq 0$. We assume that $\{y_k\}$ and $\{s_k\}$ are generated by (3.3), but for the moment we will leave $\{x_k\}$ unspecified.

We now plan to investigate the behavior of the sequence $\{B_k\}$ generated by (3.5). For this first recall that for $A \in L(R^n)$ the Frobenius norm is defined by

$$||A||_F^2 = \sum_{i,j=1}^{n} |a_{ij}|^2$$

We also need the following special case of a result of Broyden, Dennis, and Moré [2].

<u>Lemma 3.1</u> Let $M \in L(R^n)$ be a nonsingular symmetric matrix such that

(3.6) $$||My - M^{-1}s|| \leq \beta||M^{-1}s||$$

for some $\beta \in [0, 1/3]$ and vectors y and s in R^n with $s \neq 0$. Then $y^T s > 0$ and thus $\bar{B} \in L(R^n)$ can be defined by

$$\bar{B} = B + \frac{(y-Bs)^T y + y(y-Bs)^T}{y^T s} - \frac{s^T(y - Bs)yy^T}{(y^T s)^2}$$

where $B \in L(R^n)$ is symmetric. Moreover if $||\cdot||_M$ is the matrix norm defined by

(3.7) $$||Q||_M = ||MQM||_F,$$

then there are positive constants α, α_1 and α_2 (depending only on M and n) such that for any symmetric $A \in L(R^n)$,

$$||\bar{B} - A||_M \leq \left[\sqrt{1 - \alpha\theta^2} + \alpha_1 \frac{||My - M^{-1}s||}{||M^{-1}s||}\right] ||B - A||_M$$

$$+ \alpha_2 \frac{||y - As||}{||M^{-1}s||}$$

where $\alpha \in [0,1]$, and

$$\theta = \frac{||M[B - A]s||}{||B - A||_M \; || M^{-1}s||}, \quad \text{for } B \neq A,$$

with $\theta = 0$ otherwise.

In [2], a more general result was given, but only for the case when $||\cdot||$ was the ℓ_2 norm and $B \neq A$. Those restrictions are unnecessary in the present case. Note that this lemma needs only a way of choosing M in order to be applicable to $\{B_k\}$. In the following lemma we show that if $F'(x^*)$ is symmetric and positive definite then M can be taken to be the symmetric, positive definite, square root of $F'(x^*)^{-1}$, i.e., $F'(x^*)^{-1/2}$. It is interesting to note that we do not assume $\{x_k\}$ is generated by (3.1) or that $\{H_k\}$ exists or even that $F(x^*)=0$.

<u>Lemma 3.2</u> Let $F\colon R^n \to R^n$ be differentiable in an open, convex neighborhood D of a point x^* for which $F'(x^*)$ is symmetric and positive definite, and suppose that F satisfies the Lipschitz condition (3.4) in a neighborhood of x^*. If for some sequence $\{x_k\} \subset D$ which converges to x^*, we define $\{s_k\}$ and $\{y_k\}$ by (3.3), then $y_k^T s_k > 0$ for $k \geq k_0$. Moreover for any symmetric $B_{k_0} \in L(R^n)$ the sequence $\{B_k\}$ is well-defined by (3.5) for $k \geq k_0$, and there are positive constants α, α_3, and α_4 such that

$$(3.8)\quad ||B_{k+1}-F'(x^*)||_M \leq [\sqrt{1-\alpha\theta_k^2} + \alpha_3\ \sigma_k]\,||B_k-F'(x^*)||_M + \alpha_4\,\sigma_k ,$$

where

$$\sigma_k = \max\{||x_{k+1}-x^*||^p,\ ||x_k-x^*||^p\}.$$

Here $\alpha \in (0,1]$, $||\cdot||_M$ is defined by (3.7) with $M = F'(x^*)^{-1/2}$, and

$$\theta_k = \frac{||M[B_k - F'(x^*)]\, s_k||}{||B_k - F'(x^*)||_M ||M^{-1}s_k||} \quad , \text{ for } B_k \neq F'(x^*),$$

and $\theta_k = 0$ otherwise.

<u>Proof</u>: For $M = F'(x^*)^{-1/2}$ we have $||My_k - M^{-1}s_k|| \leq ||M||\ ||y_k - F'(x^*)s_k||$, and it is not difficult (e.g. Lemma 3.1 in [2]) to verify that if (3.4) holds then

$$||y_k - F'(x^*)s_k|| \leq K \max\{||x_{k+1} - x^*||^p, ||x_k - x^*||^p\}||s_k||.$$

Hence, (3.6) is satisfied if k_0 is taken so that $K||M||\ ||x_k - x^*||^p \leq 1/3$ for $k \geq k_0$. The result now follows from Lemma 3.1.

Our next task is to deduce properties of $\{B_k\}$ from (3.8). The following result is applicable if we assume that

(3.9) $$\sum_{k=1}^{\infty} ||x_k - x^*||^p < +\infty \quad .$$

<u>Lemma 3.3</u> Let $\{\phi_k\}$ and $\{\delta_k\}$ be sequences of nonnegative numbers such that

(3.10) $$\phi_{k+1} \leq (1 + \delta_k)\phi_k + \delta_k$$

and

$$(3.11) \qquad \sum_{k=1}^{\infty} \delta_k < +\infty.$$

Then $\{\phi_k\}$ converges.

Proof: We first show that $\{\phi_k\}$ is bounded (above). Let

$$\mu_k = \prod_{j=1}^{k-1} (1 + \delta_j).$$

Then $\mu_k \geq 1$, and (3.11) implies that $\mu_k \leq \mu$ for some constant μ. But (3.10) yields

$$\frac{\phi_{k+1}}{\mu_{k+1}} \leq \frac{\phi_k}{\mu_k} + \frac{\delta_k}{\mu_{k+1}} \leq \frac{\phi_k}{\mu_k} + \delta_k ,$$

and therefore,

$$\frac{\phi_{m+1}}{\mu_{m+1}} \leq \frac{\phi_1}{\mu_1} + \sum_{k=1}^{m} \delta_k ,$$

which in view of (3.11) and the boundedness of $\{\mu_k\}$ implies that $\{\phi_k\}$ is bounded.

Since $\{\phi_k\}$ is bounded it certainly has at least one limit point. Suppose there are subsequences $\{\phi_{k_n}\}$, $\{\phi_{k_m}\}$ which converge respectively to limit points ϕ' and ϕ''. We will show that $\phi' \leq \phi''$ and by symmetry $\phi'' \leq \phi'$. Thus $\phi'' = \phi'$ and consequently $\{\phi_k\}$ converges.

Let ϕ bound $\{\phi_k\}$. It then follows from (3.10) that for $k_n \geq k_m$

$$\phi_{k_n} - \phi_{k_m} \leq (1 + \phi) \sum_{j=k_m}^{\infty} \delta_j .$$

Thus, by the choice of $\{k_n\}$,

$$\phi' - \phi_{k_m} \leq (1 + \phi) \sum_{j=k_m}^{\infty} \delta_j$$

and by the choice of $\{k_m\}$,

$$\phi' - \phi'' \leq 0.$$

This complete the proof.

Theorem 3.4 Assume that $F: R^n \to R^n$ satisfies the hypotheses of Lemma 3.2 on the set D and let $\{x_k\}$ be a sequence in D which satisfies (3.9). Then, there is a positive integer k_0 such that for any symmetric $B_{k_0} \in L(R^n)$, the sequences $\{s_k\}$, $\{y_k\}$ and $\{B_k\}$ are well defined for $k \geq k_0$ by (3.3) and (3.5). Furthermore,

(i) $\lim_{k\to\infty} ||F'(x^*)^{-1/2} B_k F'(x^*)^{-1/2} - I||_F$ exists,

and

(ii) $$\lim_{k\to\infty} \frac{||[B_k - F'(x^*)]s_k||}{||s_k||} = 0.$$

<u>Proof</u>: Statement (i) follows directly from Lemmas 3.2 and 3.3. To prove statement (ii) note that since $\sqrt{1 - \alpha\theta_k^2} \le 1 - \left(\frac{\alpha}{2}\right)\theta_k^2$, equation (3.8) can be written as

$$(\alpha\theta_k^2/2)||B_k - F'(x^*)||_M \le ||B_k - F'(x^*)||_M - ||B_{k+1} - F'(x^*)||_M$$
$$+ \sigma_k[\alpha_3||B_k - F'(x^*)||_M + \alpha_4],$$

and therefore, summing both sides,

$$(\alpha/2) \sum_{k=1}^{\infty} \theta_k^2 ||B_k - F'(x^*)||_M < +\infty.$$

If some subsequence of $\{||B_k - F'(x^*)||_M\}$ converges to zero, then (i) implies that the whole sequence converges to zero and thus (ii) is trivially satisfied. Otherwise, $\{||B_k - F'(x^*)||_M\}$ is bounded away from zero, and the last equation implies that $\{\theta_k\}$ converges to zero. Statement (ii) follows.

Note that Theorem 3.4 holds regardless of how the sequence $\{x_k\}$ is generated provided (3.9) is satisfied. For example, suppose the sequence is generated by

$$x_{k+1} = x_k - \lambda_k B_k^{-1} Fx_k$$

where $\{\lambda_k\}$ is any sequence with $|\lambda_k - 1| \le \lambda < 1$ for $k \ge 0$ and $\{B_k\}$ is generated by (3.5). Then using the techniques of [2] and inequality (3.8), it is not difficult to show that if $Fx^*=0$ then there is a constant $r \in (0,1)$ with

$$||x_{k+1} - x^*|| \leq r||x_k - x^*||$$

provided (x_0, B_0) is close enough to $(x^*, F'(x^*))$. Hence, (3.11) is satisfied and if $\{\lambda_k\}$ converges to unity, (ii) above implies that

$$\lim_{k\to+\infty} \frac{||[\lambda_k^{-1} B_k - F'(x^*)]s_k||}{||s_k||} = 0.$$

Thus, Lemma 2.2 implies that $\{x_k\}$ is Q-superlinearly convergent. For $\lambda_k \equiv 1$, this reduces to a result of Broyden, Dennis, and Moré [2].

We now would like to show that Theorem 3.4 also yields the superlinear convergence of the version of the Davidon - Fletcher - Powell algorithm considered by Powell [7]. As mentioned at the beginning of this section, Powell chooses λ_k via exact line minimization, so that

$$\nabla f(x_{k+1})^T s_k = 0.$$

<u>Theorem 3.5</u> Let $F: R^n \to R^n$ be differentiable in the open, convex set D in R^n, and assume that for some x^* in D, $F'(x^*)$ is nonsingular, and F' is continuous at x^*. Let the sequence $\{x_k\}$ be defined by

$$x_{k+1} = x_k - \lambda_k B_k^{-1} Fx_k$$

where the nonsingular matrix B_k and scalar λ_k are such that

$$Fx_{k+1}{}^T s_k = 0,$$

and suppose that $\{x_k\}$ lies in D and converges to x*. If

$$\lim_{k\to +\infty} \frac{||[B_k - F'(x^*)]s_k||}{||s_k||} = 0,$$

then $\{\lambda_k\}$ converges to unity, Fx*=0 and $\{x_k\}$ converges Q-superlinearly to x*.

Proof: Define, for v, w $\in E^n$,

$$<v,w> = [F'(x^*)^{-1}v]^T w$$

so that by the mean value theorem,

$$0 = <Fx_{k+1},s_k> = <Fx_k,s_k> + <F'(x_k + t_k s_k)s_k, s_k>$$

where $|t_k| < 1$. Hence,

$$<F'(x_k + t_k s_k)s_k,s_k> = -<Fx_k,s_k> = \lambda_k^{-1} <B_k s_k, s_k> .$$

Therefore,

$$(\lambda_k - 1) < F'(x^*)s_k, s_k> = \lambda_k < [F'(x^*) - F'(x_k + t_k s_k)]s_k, s_k>$$

$$+ <[B_k - F'(x^*)]s_k, s_k> .$$

Using the definition of $<\cdot,\cdot>$, the above equation can be written as

$$(3.12) \qquad \lambda_k - 1 = \lambda_k \alpha_k + \beta_k$$

where

$$|\alpha_k| \leq ||F'(x^*) - F'(x_k + t_k s_k)|| \; ||F'(x^*)^{-1}||,$$

and

$$|\beta_k| \leq \frac{||[B_k - F'(x^*)]s_k||}{||s_k||} \cdot ||F'(x^*)^{-1}||.$$

It follows that $\{\alpha_k\}$ and $\{\beta_k\}$ converge to zero, and thus (3.12) implies that $\{\lambda_k\}$ converges to unity as desired. The Q-superlinear convergence of $\{x_k\}$ to x^* where $Fx^*=0$ follows from Corollary 2.3.

Theorems 3.4 and 3.5 guarantee the Q-superlinear convergence of the Davidon - Fletcher - Powell method to a local minimum if λ_k is chosen by an exact line search, and the sequence $\{x_k\}$ converges to x^* in such a way that (3.9) is satisfied. For example, under the conditions on F mentioned at the beginning of this section, Powell [7 , Theorem 2] proved that there is a $\mu < 1$ such that

$$||x_k - x^*|| \leq \eta\mu^k$$

for some constant $\eta > 0$ and all k. Hence, (3.9) holds, and we obtain an independent proof of Powell's result of Q-superlinear convergence.

Note, however, that Q-superlinear convergence to a zero of F is also assured if $\{\lambda_k\}$ is any sequence wich converges to unity and makes (3.9) hold. Therefore, a good choice of $\{\lambda_k\}$ would give us (3.9) while minimizing function evaluations; exact line minimization is an expensive way to do this.

4. Extensions

The results of the previous section do not hold only for the Davidon - Fletcher - Powell method. In particular, the all-important (ii) of Theorem 3.4 holds for the following updates:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k}$$

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) y_k^T}{y_k^T s_k}$$

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^T + s_k (y_k - B_k s_k)^T}{s_k^T s_k} - \frac{s_k^T (y_k - B_k s_k) s_k s_k^T}{(s_k^T s_k)^2}$$

Only in the third update is it necessary to take B_0 symmetric. To verify that these updates satisfy (ii) of Theorem 3.4 we only need to show that they satisfy a relationship like (3.8), and this is done in Lemmas 4.2 and 5.2 of [2]. The assumptions are as in Theorem 3.4 but for the first and third updates M is the identity matrix, so that $F'(x^*)$ need only be symmetric in the third update, while existence is all that is required in the first update. Conclusion (i) of Theorem 3.3 also holds but for the first and third updates it should be modified to read

$$\lim_{k\to+\infty} ||B_k - F'(x^*)||_F \text{ exists.}$$

For more information on these methods and their properties see [2].

As we mentioned in the introduction, results on Q-superlinearity are usually proved by assuming $\{B_k\}$ converges to $F'(x^*)$. Nothing, so far, in our results says that this approach can't be applied to the quasi-Newton methods, but consider the following class of examples. Define $F: R^n \to R^n$ such that $f_2,\ldots,f_n$ are independent of x_1 and $f_1(x)=x_1$. Let x_0 have a zero in its first coordinate and B_0 zeros in its first row and column except for possibly the (1,1) element. If $\{x_k\}$ and $\{B_k\}$ are generated by the Davidon-Fletcher-Powell or any of the above three methods, then the first row and column of B_k will remain unchanged while the rest of B_k will be the matrix generated by the corresponding method when applied to $f_2,\ldots,f_n$ as a function of the n-1 variables $(x_2,\ldots,x_n)$. In particular, the sequence $\{B_k\}$ does not converge to $F'(x)$ for any x in R^n. Of course, it is always possible for the iteration to terminate or to break down because of a division by zero, but by choosing $f_2,\ldots,f_n$ appropriately, convergence will be assured. Also note that the above example applies to the "damped" method

$$x_{k+1} = x_k - \lambda_k B_k^{-1} F x_k, \quad \lambda_k \neq 0.$$

The first person to produce an example of this type was Powell [8]. He pointed out that if the Davidon - Fletcher - Powell method with exact line minimization is used to minimize the function

$$f(x_1,\ldots,x_n) = x_1^2 + g(x_2,\ldots,x_n)$$

then the above behavior occurs from initial points of the type specified above.

The previous results also apply to the so-called complementary or Fletcher-dual methods [3]. For example, the complementary Davidon-Fletcher-Powell method is given by

$$x_{k+1} = x_k - H_k F x_k$$

where

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)^T s_k + s_k (s_k - H_k y_k)^T}{y_k^T s_k} - \frac{y_k^T (s_k - H_k y_k)\ s_k s_k^T}{(y_k^T s_k)^2} .$$

The Fletcher-dual is therefore obtained by replacing B's by H's and interchanging y_k and s_k. For this method, (i) and (ii) of Theorem 3.4 now read

(i) $\lim_{k\to+\infty} ||F'(x^*)^{1/2} H_k F'(x^*)^{1/2} - I||_F$ exists,

$$\text{(ii)} \qquad \lim_{k\to+\infty} \frac{||[H_k - F(x^*)^{-1}]y_k||}{||y_k||} = 0,$$

where we are now assuming that $y_k \neq 0$ for $k \geq 0$. Similar observations apply to the dual of the three methods defined in this section.

Q-superlinear convergence to a zero of F also follows from (ii) above if we assume that $\{||H_k^{-1}||\}$ is uniformly bounded. In this case,

$$[H_k - F'(x^*)^{-1}]y_k = H_k F x_{k+1} - F'(x^*)^{-1}[y_k - F'(x^*)s_k],$$

and if F satisfies the assumption of Theorem 2.2 then there is a $\gamma > 0$ such that

$$||y_k|| = ||Fx_{k+1} - Fx_k|| \geq \gamma \, ||x_{k+1} - x_k||.$$

All together,

$$\lim_{k\to+\infty} \frac{||Fx_{k+1}||}{||x_{k+1} - x_k||} = 0.$$

This is just (2.5), so from (2.4) we see that $\{x_k\}$ must converge Q-superlinearly to x^* and $F(x^*)=0$.

For the complementary Davidon-Fletcher-Powell method the matrices $B_k = H_k^{-1}$ are generated by

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} , \tag{4.1}$$

In this case it is possible to prove, with the technique that Powell used in Theorem 3 of [7], that under the assumptions of Theorem 3.4 there is a positive integer k_0 such that for any symmetric $B_{k_0} \in L(R^n)$ the sequences $\{s_k\}$, $\{y_k\}$ and $\{B_k\}$ are well defined for $k \geq k_0$ by (3.3) and (4.1), and moreover $\{||B_k||\}$ is uniformly bounded. Hence, the results of Section 3 apply to the complementary Davidon-Fletcher-Powell method without explicitly assuming that $\{||H_k^{-1}||\}$ is uniformly bounded.

5. Concluding Remarks

It is interesting to compare our results with those of McCormick and Ritter [5], and Ritter [9]. These authors prove, under more restrictive hypotheses than those of Theorem 2.2, that $\{x_k\}$ is Q-superlinearly convergent if

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

They fail to note that this condition is also sufficient (and hence, equivalent to (2.3)), and moreover, they do not prove that any of the single - or double-rank methods satisfy their hypotheses. They do make the interesting point that if

$$\frac{||[B_k - F'(x^*)](x_{k+1}-x_k)||}{||Fx_k||} = \mathcal{O}(||Fx_{k-\nu}||) \tag{5.1}$$

for some ν, $0 \leq \nu \leq n$, then

$$\frac{||x_{k+1}-x^*||}{||x_k-x^*||} = \mathcal{O}(||x_{k-\nu}-x^*||).$$

In this case, it follows that the R-order of (2.2) is at least r where r is the unique positive root of $t^{\nu+1}-t^{\nu}-1=0$. See, for example, [6,page 291]. It would be very interesting to prove that some quasi-Newton method satisfies (5.1) or some similar relationship that would guarantee an order greater than one.

We also note that the techniques of this paper can also be used to establish Q-superlinear convergence of algorithms that use the quasi-Newton philosophy. For example, K. Brown and J.E. Dennis, in unpublished work, have proven Q-superlinearity of the nonlinear least squares method given in [1].

Finally, readers familiar with the majorant approach to the convergence theory of nonlinear iterative methods will perhaps agree with us that (2.3) is a sort of directional-norm analogue of consistency, i.e., norm convergence of $\{B_k\}$ to $F'(x^*)$. It would be interesting to find a reasonable sufficient condition for local convergence of (1.1) based on such a directional norm approach.

Acknowledgments

We would like to thank M.J.D. Powell for allowing us to publish the example mentioned in Section 4, P.T. Boggs who informed us of its existence, and C.G. Broyden, who taught us a lot of what we know about quasi-Newton methods.

References

[1] Brown, K.M. and Dennis, J.E., Jr. (1971), "A new algorithm for nonlinear least-squares curve fitting" in Mathematical Software, John R. Rice, ed., Academic Press, New York, pp. 391-396.

[2] Broyden, C.G., Dennis, J.E., and Moré, J.J. (1972), "On the local and superlinear convergence of quasi-Newton methods", Cornell Comptuer Science Technical Report 72-137.

[3] Fletcher, R. (1970), "A new approach to variable metric algorithms", Comput. J. 13, pp. 317-322.

[4] Goldstein, A. and Price, J. (1967),"An effective algorithm for minimization", Numer. Math.10, pp. 184-189.

[5] McCormick G. and Ritter, K. (1972), "Methods of conjugate directions versus quasi-Newton methods", Math. Prog. 3, pp. 101-116.

[6] Ortega, J.M. and Rheinboldt, W.C. (1970), Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York.

[7] Powell, M.J.D. (1971), "On the convergence of the variable metric algorithm", J.I.M.A. 7, pp. 21-36.

[8] Powell, M.J.D. (1972), Private communication.

[9] Ritter, K. (1972), "Superlinearly convergent methods for unconstrained minimization problems", Proc. ACM '72, Boston, pp. 1137-1145.

[10] Voigt,R.(1969)"Rates of convergence for iterative methods for nonlinear systems of equations", Ph.D. Thesis, University of Maryland.