A BRIEF SURVEY OF CONVERGENCE RESULTS
FOR QUASI-NEWTON METHODS

J.E. Dennis, Jr.+

TR 75-238

May 1975

Department of Computer Science
Cornell University
Ithaca, New York 14853

+This research was supported in part by the National Science Foundation under Grant GJ-40903.

A BRIEF SURVEY OF CONVERGENCE RESULTS FOR QUASI-NEWTON METHODS

J.E. Dennis, Jr.[1]

ABSTRACT

This paper highlights the important theoretical developments in the study of quasi-Newton or update methods and suggests avenues for future research. An attempt is made to present this material in a way reasonably compatible with history but organized for the novice.

1. INTENT

The purpose of this paper is to present an informal survey of that thread of research on so-called "quasi-Newton" methods which is most likely to be of interest to a mathematician considering work on practical optimization. The reader interested in a more complete or formal introduction to this active and fruitful area should consult, for example, Dennis and Moré [19], Lootsma [36] or Murray [40].

2. THE PROBLEM AND SOME METHODS

We will view these methods as successful multivariate extensions of the useful secant method for nonlinear scalar equations. The basic problem to be attacked here is:

$$\begin{aligned}&\text{Given } F: R^n \to R^n, \text{ find } x^* \in R^n\\ &\text{for which}\\ &F(x^*) = (f_1(x^*), f_2(x^*), \ldots, f_n(x^*))^T = 0.\end{aligned}$$

The classical Newton or Newton-Raphson method for generating a sequence of approximation to such an x^* is:

(2.1) Given $x_k \in R^n$, solve the linear system

$$F'(x_k)s_k = -F(x_k)$$

and set

$$x_{k+1} = x_k + s_k$$

Test x_{k+1} and if it is acceptable stop, otherwise replace k by k+1 and repeat the process. Generally the test for convergence involves some consideration of relative changes in x_{k+1} and $F(x_{k+1})$ as the iteration proceeds as well as $||F(x_{k+1})||$.

The advantages of Newton's method are given by the following theorem whose proof hardly needs repeating here.

AMS(MOS) subject classifications (1970). Primary 63H10, 65K05, Secondary 49D05, 49D10, 49D15.

[1]Research supported by the National Science Foundation under grant GJ-40903.

THEOREM 2.2 If F' is continuous in a neighborhood of some x^* for which $F(x^*) = 0$ and $F'(x^*)^{-1}$ exists, then there exists a positive constant Δ such that if $||x_0-x^*|| < \Delta$ then the sequence $\{x_k\}$ exists and converges to x^*. Moreover, if there is some K for which $||x-x^*|| < \Delta$ implies $||F'(x)-F'(x^*)|| \leq K \; ||x-x^*||$, then

$$\lim_{k \to \infty} \frac{||x_{k+1}-x^*||}{||x_k-x^*||^2} < \infty$$

The above behavior of the error sequence $\{x_k-x^*\}$ is called <u>Q-order 2</u>. It is so fast that any higher order, ie, an exponent higher than 2 on the denominator is probably not of much interest, particularly since much more effort is generally required by such higher order methods. Of course, this theorem is as important as it is because for a large class of problems, Newton's method not only converges but quickly begins to exhibit second order convergence. There is a very elegant type of analysis due to L.V. Kantorovich which relates Δ to the Lipschitz constant K, whose magnitude can be viewed as a measure of the nonlinearity of F. For an exposition see Dennis [13].

If the problem is sufficiently nonlinear, a very good x_0 might be required. Another important difficulty arises when $F'(x_k)$ is singular. We refer the reader to the more complete surveys mentioned above for these not so well understood difficulties. We will concentrate on the two drawbacks of (2.1) which quasi-Newton methods attempt to allay:

(2.3)
- (i) The computation of $\{F'(x_k)\}$.
- (ii) The solution of an nxn linear system for each s_k.

The difficulties in solving a linear system of equations are well set out in Forsythe and Moler [26] and so we will just be concerned with the expense of the $0(n^3)$ algebraic operations required. The quasi-Newton methods can reduce this "algebraic complexity" to $0(n^2)$.

The disadvantages in (2.3.i) are really much more interesting and important. On the surface one sees the important expense of $0(n^2)$ nonlinear functional evaluations. In fact, it is often possible to alleviate this somewhat by clever programming to make use of common subexpressions in F and F'. A very serious objection is the taking and programming of the partial derivatives since this is a time consuming and error prone task. Symbolic differentiation programs can be of some help but often F is the output of some program and there are no expressions to differientiate. Thus, a simpler and more general remedy is to use finite difference approximations to the partial derivatives. While this is usually effective in practice, it doesn't always work and one still has the "$0(n^2)$ function complexity". See Brown and Dennis [2], Dennis [14] and Goldstein and Price [29] for more detail.

In the case of a single equation in a single unknown the secant method is a very efficient algorithm.

$$\text{Given } t_k, \text{ set } t_{k+1} = t_k - \frac{t_k - t_{k-1}}{F(t_k) - F(t_{k-1})} F(t_k).$$

The secant method has the same basic local behavior as Newton's method but notice that no F-information is required except the value at the current iterate, which we would need anyway in the convergence test. Furthermore, if we charge the division to the approximation of $F'(t_k)$, then no linear system solution is required and the $0(n^2)$ functional complexity for $F'(t_k)$, required by either version of Newton's method, has been replaced by $0(n^2)$ algebraic complexity. Under appropriate conditions the Q-order is $(1+\sqrt{5})/2$ which is usually not appreciably worse than 2 in practice. A lot of effort has gone into multidimensional generalizations which attempt to preserve this order property but don't work very well on any but special problems. See Ortega and Rheinboldt [42] for a comprehensive survey. We will see that all the quasi-Newton methods use the minimal F-information of the secant method, most are fast, and all reduce to the secant method for n=1.

3. RANK ONE METHODS

The first quasi-Newton method was introduced by W.C. Davidon in 1959 ([12]) and reinterpreted and popularized by R. Fletcher and M.J.D. Powell in 1963 ([23]). Before discussing this work, we skip ahead to Broyden [4] who was the first in this area to deal with the general nonlinear equation problem of section 2. Broyden suggested the following class of algorithms:

(3.1) Given x_k, H_k, let $s_k = -H_k F(x_k)$.

Set $x_{k+1} = x_k + s_k$.

Compute $F(x_{k+1})$ and test for convergence. Either accept x_{k+1} as the final approximation to x^* or choose $d_k \in R^n \sim \{\Delta F_k\}$ where $\Delta F_k = F(x_{k+1}) - F(x_k)$. Set

$$H_{k+1} = H_k - \frac{(H_k \Delta F_k - \Delta x_k) d_k^T}{d_k^T \Delta F_k}$$

where $\Delta x_k = x_{k+1} - x_k$ and repeat the process with k+1 in place of k.

Notice that this Broyden class of single-rank methods all reduce to the secant method for n=1. Furthermore, (3.1) only requires minimal F-information. The approximation to $F'(x_k)$ is of algebraic rather than function complexity $0(n^2)$. Broyden singled out two specific choices of d_k. One method is $d_k = \Delta F_k$ and it is a rather obvious choice in view of the requirement that $d_k^T \Delta F_k \neq 0$. This method does not work well although it has much theoretical support as <u>Broyden's method</u>, $d_k = H_k^T H_k F(x_k)$ which does. The latter choice seems as natural as the former if we use the Sherman-Morrison-Woodbury formula (see [53]) to write

$^{-1}$ ... $[\Delta F_k - B_k \Delta x_k]\, c_k^T$ ... ^{T}

Now we see that $d_k = H_k^T H_k F(x_k)$ corresponds to $c_k = \Delta x_k$. Dennis and Moré [19] give other motivation.

There is as yet no satisfactory explanation for the radical difference in the behavior of these two methods.

We mention the following theorem (Dennis [15]) because the proof used techniques which later proved fruitful in obtaining a better result and because the question raised by this theorem also led to interesting research.

THEOREM 3.3 Let all the hypotheses of Theorem 2.2 hold. Then there exist Δ_1, Δ_2 such that for $||x_o-x^*||<\Delta_1$, $||B_o-F'(x^*)||<\Delta_2$, the sequence $\{x_k\}$ defined by Broyden's method exists and converges to x^* with $||x_{k+1}-x^*||<\alpha||x_k-x^*||$ for some $\alpha<1$ and every k.

The requirement that B_o be close to $F'(x^*)$ is reasonable. The disappointing part is that the rate of convergence proved was only "Q-linear". In practice the iteration seemed much better but in order to prove Q-superlinearity it had always seemed necessary to have a consistent method. By "consistent" we mean that if $\{x_k\}$ converges to x^* then $\{B_k\}$ converges to $F'(x^*)$. This did not seem to hold even in the simple case of an affine F.

The analysis was based on proving that

$$||B_{k+1}-F'(x_{k+1})||_2 \le ||B_k-F'(x_k)||_2 + K\cdot||\Delta x_k||_2 \qquad (3.4)$$

We interpret this as guaranteeing that the ℓ_2 norm of the error in the derivative approximation stays below a line parameterized by the length of the iteration path and with slope measured by the nonlinearity of the function. In Dennis [13] the interested reader will find a Kantorovich analysis for general iterations whose derivative approximations satisfy this "bounded deterioration" condition.

Moré and Trangenstein [39] have a global convergence result for Broyden's method on an affine F.

Two other important single-rank update formulae were given in Pearson [43] There, Pearson suggested $c_k = \Delta F_k$ and McCormick suggested $c_k = B_k^T \Delta x_k$ or, as it is more naturally viewed $d_k=\Delta x_k$. We will discuss these later. Another important choice is $c_k=\Delta F_k - B_k\Delta x_k$, called for obvious reasons the symmetric single rank, or SSR method.

4. ANOTHER PROBLEM AND SOME METHODS

The unconstrained minimization problem:

Given $f:R^n \to R^1$ find $x^* \in R^n$ such that $f(x^*) \le f(x)$
for every x in some neighborhood of x^*;

is a very important special case of the nonlinear programming problem, (see Mangasarian [37] or Fiacco and McCormick [22]). Our point of view has always been that it is a special case of the problem of section 2. In that case $F=\nabla f$ ie, $f_i = \partial_i f$, the ith partial derivative of f and under the standard assumptions on F' we have interesting special structure. The Jacobian matrix, F' is the hessian matrix of f, its matrix of second partial derivatives, and is hence symmetric everywhere and positive definite at

The Newton method is again the model method but clearly Theorem 2.2 insures local convergence to any isolated zero of ∇f. This means (2.1) might find a local maximum or saddle point instead of the desired local minimum. In order to guard against this and to widen the choice of successful starting points, s_k is chosen as before but $x_{k+1}=x_k+t_k s_k$. Here t_k is chosen to at least make $f(x_{k+1})\leq f(x_k)$.

Clearly the finite difference form of Newton's method is applicable here also.

A complete set of theorems for Newton's method with various strategies for choosing $\{t_k\}$ can be found in Goldstein [28]. A through discussion of desirable features in a choice rule for $\{t_k\}$ can be found in Daniel [11] or Ortega and Rheinboldt [42]. A clear exposition of computer routines often used can be found in Jacoby, Kowalik and Pizzo [33].

Generally speaking, however s_k is chosen, these "line searches" seek to find t_k so that ideally $f(x_{k+1})$ is at least a local minimum of $\phi_k(t)=f(x_k+ts_k)$. Of course, such a value is not guaranteed to exist unless f has some property such as, bounded level sets, or strict convexity and it is under such assumptions that so-called global theorems are proved.

Following the terminology of Dixon [20] we say an iteration is <u>perfect</u> if t_k was chosen to be the global minimum of ϕ_k. If $t_k=1$ the step is said to have been made by <u>direct prediction</u>.

5. RANK-TWO QUASI-NEWTON METHODS

It is certainly possible to apply the methods of section 3 to compute $\{s_k\}$ and to use them with a line search routine, to solve the unconstrained minimization problem. On the other hand, we know the hessian is symmetric and so it seems reasonable to use a symmetric approximation to it. In addition, if the approximate hessian is positive definite then s_k would be a descent direction.

We could obtain a symmetric hessian approximation by averaging any of the single rank update formulae for hessian or inverse hessian approximations from section 3 with its transpose. We could then use the same update formula on the result in order to preserve the generalized divided difference property but this would destroy the symmetry. If we repeat this procedure it has a limit and leads to the following class of rank two update formulae for generating hessian approximations. The single rank formula specified by c_k or d_k related by $c_k=B_k^T d_k$ yields

$$B_{k+1}=B_k+\frac{[\Delta F_k-B_k\Delta x_k]c_k^T+c_k[\Delta F_k-B_k\Delta x_k]^T}{c_k^T\Delta x_k}-\frac{\Delta x_k^T[\Delta F_k-B_k\Delta x_k]c_k c_k^T}{(c_k^T\Delta x_k)^2}$$

and

$$H_{k+1}=H_k+\frac{[\Delta x_k-H_k\Delta F_k]d_k^T+d_k[\Delta x_k-H_k\Delta F_k]^T}{d_k^T\Delta F_k}-\frac{\Delta F_k{}^T[\Delta x_k-H_k\Delta F_k]d_k d_k^T}{(d_k^T\Delta F_k)^2}.$$

For more details on this derivation, see Powell [45], Dennis [16] or Dennis and More [19]. Note that $H_k B_k=I$ does not generally imply $H_{k+1}B_{k+1}=I$ except for the symmetric single rank method. The following chart uses this symmetrization device to orgainze most of the named methods.

Single Rank	Double Rank B_k	Double Rank H_k
Broyden's method [4] $c_k=\Delta x_k$ $d_k=H_k^T H_k F(x_k)$	Powell symmetric Broyden method, the PSB [45]	unnamed
Pearson's method [43] $c_k=\Delta F_k$ $d_k=H_k^T \Delta F_k$	Davidon [12] Fletcher & Powell, the DFP [23]	Greenstadt [30]
The symmetric single rank method (SSR) $c_k = \Delta F_k - B_k \Delta x_k$, $d_k = H_k \Delta F_k - \Delta x_k$		
McCormick's method, see Pearson [43] $c_k=B_k^T \Delta x_k$ $d_k=\Delta x_k$	unnamed	Broyden [6], [8] Fletcher [24] Goldfarb [27] Shanno [52], the BFGS.
Broyden's "other method" [4] $c_k=B_k^T \Delta F_k$ $d_k=\Delta F_k$	unnamed	Greenstadt's "other method" [30]

The first theorems on these methods do not seem particularly relevant as theoretical justification in view of subsequent developments. They were very important to the way the subject developed and so we give one.

THEOREM 5.1 Let $f=(1/2)x^TAx+b^Tx+c$ where A is a symmetric positive definite real $n\times n$ matrix, $b\in R^n$, $c\in R^1$. If H_0 is symmetric and positive definite and $x_0\in R^n$, then the perfect DFP method converges to x^* in $m\leq n$ iterations. The corresponding $H_1, \ldots, H_m$ are symmetric and positive definite and if $m=n$ then $H_mA=I$.

Broyden [5] gave a characterization of all rank-2 update formulae for which this theorem holds. The BFGS is one such. A similar theorem but without requiring perfect line searches, holds for the SSR method when it manages to be defined. But it may not be even in simple cases (See Dennis and Moré [19]). Huang [31] proved that all the Broyden class generated precisely the same sequence $\{x_k\}$ under the hypotheses of the theorem. Huang and Levy [32] gave computational evidence that with perfect line searches they all generated the same iterates even for general nonlinear, ie, non-quadratic, functionals.

The first important convergence results for this problem were given by Powell [46]. We give one of them here.

THEOREM 5.2 Let f be a C^2 nonlinear functional and let $\nabla^2 f$ satisfy a Lipschitz condition on all of R^n. Moreover, let there exist some positive lower bound, independent of x, on the spectrum of $\nabla^2 f(x)$. Then the perfect DFP method converges Q-superlinearly to x^*, the global minimum of f, from any x_o and positive definite H_o, ie,

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

Dixon [20] gave a proof of Huang and Levy's observation so that we have the following corollary which essentially ends research on perfect quasi-Newton methods.

COROLLARY 5.2 (Dixon-Powell). Theorem 5.1 holds for every member of Broyden's class and in particular for the BFGS method.

Burmeister [10], Lenard [34], Powell [49], Schuller [50], Schuller and Stoer [51] and Stoer [54] have all recently begun research on the rank-2 methods with imperfect line searches. This promises to be a fruitful area but rather than survey this work and be almost immediately out of date, we point out that Dixon's result was completely unexpected. It had a great impact on this area because for years people had been publishing computational results favoring one method or the other. These experiments can now be interpreted as relating a particular line search subroutine to the methods tested. Dixon realized this and an important study is Dixon [21]. It now appears that the BFGS is the method to be preferred for line search implementations.

6. DIRECT PREDICTION RESULTS

We introduced quasi-Newton methods as multidimensional analogs of the secant method and it seems appropriate to discuss the convergence theorems which have justified this viewpoint.

In section 3 we gave the Dennis [15] theorem for Broyden's method and we mentioned the idea of bounded deterioration on which that proof was based. This principal holds in the ℓ_2 operator norm for that method (see (3.4)) but in fact one can also show locally that under the hypotheses of Theorem 3.2

$$||B_{k+1}-F'(x^*)||_F \le ||B_k-F'(x^*)||_F + K\sigma(x_{k+1},x_k),$$

where $||A||_F=(\Sigma(a_{ij})^2)^{1/2}$ is the Frobenius norm and $\sigma(x,y)=\max\{||x-x^*||_2, ||y-x^*||_2\}$.

The key to a local analysis of the direct prediction DFP and BFGS methods was to show that this result holds for the DFP under the hypotheses of Theorem 5.2 if the norm is the Frobenius norm of $\nabla^2 f(x^*)^{-1/2}(A)\nabla^2 f(x^*)^{-1/2}$ where $\nabla^2 f(x^*)^{-1/2}$ is the symmetric positive definite square root of $\nabla^2 f(x^*)$. Although this norm might seem artificial it is exactly what one would wish since it has a very natural interpetation as the relative error in approximation of $\nabla^2 f(x^*)$ by the DFP (and Pearson's update) approximate hessian. On the other hand, the BFGS (and McCormick's update) can be shown to be of bounded deterioration in the relative error norm as an approximation to

We can now see that the proper interpetation of the error relation for Broyden's method was in terms of the absolute as opposed to relative error. The same absolute error result can be shown if $F'(x^*)$ is symmetric, for the PSB update. An interesting fact is that Broyden's other method (and Greenstadt's other method) can both be shown to obey this principal as absolute approximations to $F'(x^*)^{-1}$ (if $F'(x^*)$ is symmetric). Since these latter methods don't seem to work well in practice, it is not clear that we should be pleased to have them yield to our analysis.

We state the following composite theorem from Broyden, Dennis and Moré [9]. It will not surprise the reader that the relationship between single and double rank updates given in section 5 was very important in discovering the proof, which follows from the bounded deterioration principal.

THEOREM 6.1 Let $x^* \in R^n$ be a zero of F, a continuously differentiable function from a neighborhood of x^* into R^n. Assume that $F'(x^*)$ is nonsingular and that for some $K \geq 0$ and $\gamma > 0$ and some neighborhood Ω of x^*, $x \in \Omega$ implies $||F'(x)-F'(x^*)|| \leq K||x-x^*||^{\gamma}$. Under these hypotheses there exist $\Delta_1 > 0$, $\Delta_2 > 0$ such that if $||B_o - F'(x^*)|| < \Delta_1$ and $||x_o - x^*|| \leq \Delta_2$ the following direct prediction methods are defined and converge Q-superlinearly to x^*:

(i) Broyden's two methods;

(ii) The PSB and Greenstadt's other method; if in addition $F = \nabla f$ for some nonlinear functional f, ie, $F'(x)$ is symmetric for $x \in \Omega$.

(iii) The DFP, BFGS, Pearson's and McCormick's methods; if in addition to the assumption of (ii), $\nabla^2 f(x^*)$ is positive definite.

We will discuss the method of proof of superlinearity in the next section.

7. SUPERLINEARITY OF QUASI-NEWTON METHODS

Work on the local convergence of quasi-Newton methods has made at least two contributions to the study of general iterative methods. The first was the idea of bounded deterioration and the second is a characterization theorem for Q-superlinear methods. For the simple proof see Dennis and Moré [18].

THEOREM 7.1 Let $F: R^n \to R^n$ be differentiable in the open convex set $D \subset R^n$, and assume that for some $x^* \in D$, F' is continuous at x^* and $F'(x^*)$ is nonsingular. Let $\{A_k\}$ be a sequence of nonsingular matrices and let $x_o \in D$ with the property that the iteration sequence $\{x_k\}$ defined by

$$x_{k+1} = x_k - A_k^{-1} F(x_k) \qquad (7.2)$$

remains in D and converges to x^*. Then, $\{x_k\}$ converges Q-superlinearly to x^*, ie, $\lim_{k\to\infty} \frac{||x_{k+1}-x^*||}{||x_k - x^*||} = 0$, and $F(x^*)=0$ if and only if

$$\lim_{k\to\infty} \frac{||[A_k - F'(x^*)](x_{k+1}-x_k)||}{||x_{k+1}-x_k||} \qquad (7.3)$$

Professor Moré and the author recognized that this result gave the Q-superlinear part of the Broyden-Dennis-Moré as well as the Powell-Dixon theorems. We did not publish it right away because we were not certain that these results would not follow from the standard proof technique. Always before, concrete methods which were shown to be superlinear were proved so by showing that they were consistent:

$$\lim_{k\to\infty} x_k = x^* \text{ implies } \lim_{k\to\infty} A_k = F'(x^*)$$

This obviously implies (7.3) and it was well known that trumped up examples could be given so that consistency was not a necessary condition for superlinearity.

Many people, misled by consistency for strictly convex quadratics proved in Theorem 5.1, assumed that quasi-Newton methods were consistent. M.J.D. Powell in a private communication suggested a way outlined in Dennis and More [18] to construct counterexamples. This ensured the importance of Theorem 7.1.

Since no specific example has ever been published and since it is quite instructive we give one below.

Take $f(x)=\xi_1^2+\xi_2^2+\cos\xi_2$, $x=(\xi_1,\xi_2)^T$. Thus $\nabla f(x)=(2\xi_1, 2\xi_2-\sin\xi_2)^T$ and $x^*=(0,0)^T$.

Furthermore

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2-\cos\xi_2 \end{pmatrix}$$

and so the hypotheses of all the theorems are satisfied $(K=1, \gamma=1)$.

Now to start arbitrarily close, we take $x_0=(0,\varepsilon)$, $B_0=\begin{pmatrix} 2+\varepsilon & 0 \\ 0 & 1+\varepsilon \end{pmatrix}$.

The direct prediction DFP method is just the secant method applied to the second coordinate of $\nabla f(x)=0$. The first coordinate of x_k and the first row and column of B_k remain at their starting values. Since the (1,1) element of $\nabla^2 f(0)$ is 2, the method is not consistent.

It is hard to hold the failure to provide consistent derivative approximations against the method since it seem to recognize that it had as good as it needed. In practice, one often observes similiar behavior but because of inexact arithmetic this could not convince the skeptics.

This is of real practical importance since for many problems on which quasi-Newton methods are used, $\nabla^2 f(x^*)^{-1}$ is an estimate of the so-called "variance-covariance matrix". Any user wishing to use the final quasi-Newton H_k should be aware of the example above.

8. STILL ANOTHER PROBLEM

A very important and interesting unconstrained minimization problem is the nonlinear least squares problem:

Given $F:R^n \to R^m (m \geq n)$, find $x^* \in R^n$ for which $\phi(x) = 1/2\|F(x)\|_2^2$ has a local minimum.

Clearly the problem of section 2 can be attacked in this way but it is generally regarded as a bad idea. This problem is usually found in the context of curve fitting where $f_i(x)$ is the residual at the ith data point of the curve parameterized by x. The author's fascination with the problem comes from the special structure in $\nabla\phi(x)=F'(x)^T F(x)$ and

$$\nabla^2\phi(x)=F'(x)^T F'(x)+\sum_{i=1}^{m} f_i(x)\nabla^2 f_i(x).$$

It would be possible to use a quasi-Newton method directly on ϕ but this is directly opposed to the philosophy of these methods since once $\nabla\phi(x_k)$ is computed, the $F'(x_k)^T F'(x_k)$ term of the hessian is known. If we expect $\Sigma f_i(x^*)\nabla^2 f_i(x^*)$ to be small relative to this first term we could neglect it (see Brown and Dennis [2]). The resulting Newton-like method is called the Gauss-Newton method. If, rather than neglect this term, we replace it by a diagonal matrix, we obtain the Levenberg-Marquardt method (Levenberg [35], Marquardt [38]). These considerations pretty well take care of the small residual problem, ie, $\phi(x^*) \doteq 0$.

Our interest is in the case when the second term of $\nabla^2\phi$ cannot be neglected. We have suggested several methods over the last few years to deal with this but it remains an important unsolved problem. These methods are all based on making a quasi-Newton approximation to $\Sigma f_i(x_{k+1})\nabla^2 f_i(x_{k+1})$.

Brown and Dennis [3] suggested making update approximations to each of the component hessians. This can be done since $\nabla f_i(x_k)$ and $\nabla f_i(x_{k+1})$ are already needed as rows of F'. This method has all the right theoretical properties but the storage requirements for the m nxn approximate component hessians probably make this approach of limited value.

Broyden and Dennis in Dennis [17] suggested approximating the whole weighted sum of component hessians, $\Sigma f_i(x_{k+1})\nabla^2 f_i(x_{k+1})$ by one quasi-Newton approximant B_{k+1} determined from any of the previous formulas with ΔF_k replaced by $\nabla\phi(x_{k+1})-\nabla\phi(x_k)-F'(x_{k+1})^T F'(x_{k+1})\Delta x_k$. The analysis of this method is straightforward and Nazareth [41] is developing a promising algorithm of this type.

Another, previously unpublished idea is again to use a single B_{k+1} to replace the whole second term of the hessian but this time ΔF_k is replaced in the update formula by $F'(x_{k+1})^T F(x_{k+1})-F'(x_k)^T F(x_{k+1})$. We derive this from

$$\left[\sum_{i=1}^{m} f_i(x_{k+1})\nabla^2 f_i(x_{k+1})\right](\Delta x_k)=\sum_{i=1}^{m} f_i(x_{k+1})\left[\nabla^2 f_i(x_{k+1})\Delta x_k\right]$$

$$\doteq\sum_{i=1}^{m} f_i(x_{k+1})\left[\nabla f_i(x_{k+1})-\nabla f_i(x_k)\right]=F'(x_{k+1})^T F(x_{k+1})-F'(x_k)^T F(x_{k+1})$$

This idea is currently being tested.

The long range goal of this work is to begin to develop techniques which would apply to general unconstrained minimization problems where the hessian can be written $\nabla^2 f(x)=A(x)+B(x)$ and $A(x)$ can be computed while $B(x)$ cannot.

REFERENCES

1. L. Armijo, Minimization of functions having Lipschitz continuous first partial derivatives, Pacific J. Math. 16 (1966), 1-3.

2. K. M. Brown and J. E. Dennis, Jr., Derivative-free analogues of the Levenberg-Marquardt and Gauss algorithms for nonlinear least squares approximation, Numer. Math. 18 (1970), 289-297.

3. K. M. Brown and J. E. Dennis, Jr., A new algorithm for nonlinear least-squares curve fitting, in Mathematical Software, ed. by J. Rice, Academic Press, New York, (1971), 391-396.

4. C. G. Broyden, A class of methods for solving nonlinear simultaneous equations, Math. Comp. 19 (1965), 577-593.

5. C. G. Broyden, Quasi-Newton methods and their application to function minimization, Math. Comp. 21 (1967), 368-381.

6. C. G. Broyden, A new double-rank minimization algorithm, AMS Notices 16 (1969), 670.

7. C. G. Broyden, The convergence of single-rank quasi-Newton methods, Math. Comp. 24 (1970), 365-382.

8. C. G. Broyden, The convergence of a class double-rank minimization algorithms, J. Inst. Math. Appl. 7 (1971), 76-90, 222-236.

9. C. G. Broyden, J. E. Dennis and J. J. Moré, On the local and superlinear convergence of quasi-Newton methods, J. Inst. Math. Appl. 12 (1973), 223-246.

10. W. Burmeister, Die Konvergenzordnung des Fletcher-Powell Algorithmus, ZAMM 53 (1973), 693-699.

11. J. W. Daniel, The Approximate Minimization of Functionals, Prentice-Hall, Englewood Cliffs, N.J. (1971).

12. W. C. Davidon, Variable metric method for minimization, Argonne Nat. Labs. (1959) ANL-5990 Rev.

13. J. E. Dennis, Toward a unified convergence theory for Newton-like methods, in Nonlinear Functional Analysis and Applications, ed. by Louis Rall, Academic Press, New York (1971), 425-472.

14. J. E. Dennis, Algorithms for nonlinear problems which use discrete approximations to derivatives, Proc. ACM 1971 Nat. Conf. Chicago.

15. J. E. Dennis, On the convergence of Broyden's method of nonlinear systems of equations, Math. Comp. 25 (1971), 559-567.

16. J. E. Dennis, On some methods based on Broyden's secant approximation of the Hessian, in Numerical Methods for Nonlinear Optimization, ed. by F. A. Lootsma, Academic Press, London (1972).

17. J. E. Dennis, Some computational techniques for the nonlinear least squares problem, in Numerical Solution of Systems of Nonlinear Equations, ed. by G. Byrne and C. Hall, Academic, New York (1973).

18. J. E. Dennis and J. J. Moré, A characterization of superlinear convergence and its application to quasi-Newton methods, Math. Comp. 28 (1974), 549-560.

19. J. E. Dennis and J. J. Moré, Quasi-Newton methods, Motivation and Theory, Cornell University, Computer Science Technical Report (1975). 74-217.

20. L. C. W. Dixon, All the quasi-Newton family generate identical points, J.O.T.A. 10 (1972), 34-40.

21. L. C. W. Dixon, The choice of step length, a crucial factor in the performance of variable metric algorithms, Numerical Methods for Non-Linear Optimization, Academic Press, London (1972).

22. A. V. Fiacco and G. P. McCormick, Nonlinear Programming .. Sequential Unconstrained Minimization Techniques, John Wiley, New York (1968).

23. R. Fletcher and M. J. D. Powell, A rapidly convergent descent method for minimization, Comput. J. 6 (1963), 163-168.

24. R. Fletcher, A new approach to variable metric algorithms, Comput. J. 13 (1970), 317-322.

25. R. Fletcher, A survey of algorithms for unconstrained optimization, in Numerical Methods for Unconstrained Optimization, ed. by W. Murray, Academic Press, London (1972).

26. G. Forsythe and C. Moler, Computer Solution of Linear Algebraic Systems, Prentice-Hall, Englewood Cliffs, N.I. (1967).

27. D. Goldfarb, A family of variable-metric methods derived by variational means, Math. Comp. 24 (1970), 23-26.

28. A. A. Goldstein, Constructive Real Analysis, Harper & Row, New York (1967).

29. A. A. Goldstein and J. F. Price, An effective algorithm for minimization, Numer. Math. 10 (1967), 184.

30. J. Greenstadt, Variations on variable-metric methods, Math. Comp. 24 (1970), 1-18.

31. H. Y. Huang, Unified approach to quadratically convergent algorithms for function minimization, J.O.T.A. 5 (1970), 405-423.

32. H. Y. Huang and A. V. Levy, Numerical experiments on quadratically convergent algorithms for function minimization, Rice University, Aero-Astronautices Report (1969).

33. S. L. S. Jacoby, J. S. Kowalik and J. T. Pizzo, Iterative Methods for Nonlinear Optimization Problems, Prentice-Hall, Englewood Cliffs, N.J. (1972).

34. M. L. Lenard, Practical Convergence Conditions for the Davidon-Fletcher-Powell Method, University of Wisconsin-Madison (1974). MRC #1356.

35. K. Levenberg, A method for the solution of certain nonlinear problems in least squares, Quart. Appl. Math 2 (1944), 164-168.

36. F. A. Lootsma, Numerical Methods for Nonlinear Optimization, Academic, London (1972).

37. O. L. Mangasarian, Nonlinear Programming, McGraw-Hill, New York (1969).

38. D. W. Marquardt, An algorithm for least squares estimation of nonlinear parameters, SIAM J. Appl. Math. 11 (1963), 431-441.

39. J. J. Moré and J. A. Trangenstein, On the global convergence of Broyden's method, Cornell University, Computer Science Technical Report, 74-216.

40. W. Murray, Numerical Methods for Unconstrained Optimization, Academic Press, London (1972).

41. L. Nazerath, A hybird least squares method, presented at Computer Science Conf. at Washington, D.C., Feb. 1975.

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

43. J. D. Pearson, Variable metric methods of minimization, Comput. J. 12 (1969), 171-178.

44. M. J. D. Powell, A FORTRAN subroutine for unconstrained minimization, requiring first derivatives of the objective function, A.E.R.E., Harwell (1970), R.6469.

45. M. J. D. Powell, A new algorithm for unconstrained optimization, in Nonlinear Programming, ed. by Rosen, Mangasarian and Ritter, Academic Press, New York (1970).

46. M. J. D. Powell, On the convergence of the variable metric algorithm, J. Inst. Math. Appl. 7 (1971), 21-36.

47. M. J. D. Powell, Some properties of the variable metric method, in Numerical Methods for Non-linear Optimization, ed. by F. A. Lootsma, Academic Press, London (1972).

48. M. J. D. Powell, Convergence properties of a class of minimization algorithms, A.E.R.E., Harwell (1974), T.R. C.S.S.8.

49. M. J. D. Powell, Some global convergence properties of a variable metric algorithm for minimization without exact line searches, AMS Symposium on Nonlinear Programming, N.Y.C. (1975).

50. G. Schuller, On the order of convergence of certain quasi-Newton methods, Numer. Math. 23 (1974), 181-192.

51. G. Schuller and J. Stoer, Uber die Konvergenzordnung gewisser Rang-2 Verfahren zur Minimierung vor Funktionen, Int'l Ser of Numer. Math. 23 (1974), 125-147, Basel:Birkhauser.

52. D. F. Shanno, Conditioning of quasi-Newton methods for function minimization, Math. Comp. 24 (1970), 647-656.

53. J. Sherman and W. J. Morrison, Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix, Ann. Math. Statist. 20 (1949), p. 621.

54. J. Stoer, On the convergence rate of imperfect minimization algorithms in Broyden's β-class, to appear in Math. Prog. (1975).