On Some Methods Based on Broyden's Secant Approximation to the Hessian

by

J. E. Dennis, Jr.*

TR 71 - 101

July 1971

Department of Computer Science
Cornell University
Ithaca, New York 14850

*Research supported by NSF GJ-27528.

On Some Methods Based on Broyden's Secant Approximation to the Hessian

by

J. E. Dennis, Jr.

1. Introduction

The secant method is well known to be an excellent computational tool for the solution of nonlinear equations in a single variable. It is therefore quite understandable that much effort has gone into finding multi-variate analogues. Historically this work has been directed toward generalizations which preserve the order of convergence of the scalar method. Most of these methods tend to be computationally inconvenient if not actually unstable, although for fairly wide classes of problems they are quite useful. (See Collatz 1965, Hofmann 1970). The secant method is not really a high order scalar method, instead its main value lies in being fast relative to the small amount of computing it requires. (See Ralston 1965). For this reason, the so-called update methods seem to be the real spiritual heirs of the secant method. They require very little computation and they certainly are eligible for the name since they can be written

$$x_{n+1} = x_n - A_n^{-1}F(x_n), \tag{1}$$

where $F = (f_1,\ldots,f_n)^T$ is the system whose zero we desire and A_n is chosen to satisfy the divided difference relation

$$A_{n+1}(x_{n+1} - x_n) = F(x_{n+1}) - F(x_n). \tag{2}$$

This is sometimes called the "quasi-Newton equation", though of course, Newton's method, ie $A_{n+1} = F'(x_{n+1})$, does not satisfy it. We will call (2) the "secant equation".

Quite often the update methods are implemented not in the form (1) but rather as

$$x_{n+1} = x_n - t_n A_n^{-1} F(x_n) \;, \tag{3}$$

where t_n is chosen by some descent criterion. Our interest here is in performance near a root, when $t_n = 1$. It is certainly granted that to get near enough, t_n or indeed A_n, may need to be chosen differently.

2. A General Theorem.

In this section we present the general theorem and notation which will be useful in the sequal.

Sometimes in an implementation of (1) we wish to leave A_n constant for some number of iterations, and reevaluate it only at irregular intervals in the computation. This can be quite easily incorporated into the theory and so we make the following definition (Dennis 1969a and 1971).

2.1 Definition

A recalculation sequence $\{\alpha_n\}$ is a nondecreasing sequence of nonnegative integers such that $\alpha_0 = 0$ and $\alpha_n = \alpha_{n-1}$ or $\alpha_n = n$. (1) with recalculation sequence $\{\alpha_n\}$ is $x_{n+1} = x_n - A_{\alpha_n}^{-1} F(x_n)$. (4)

Let $||\cdot||$ denote a vector norm and the subordinate matrix norm. D will be a convex set.

2.2 Definition

A method of bounded deteriation is any method (1) such that as long as x_n and $F'(x_n)$ are defined,

$$||A_n - F'(x_n)|| \leq \delta_0 + \gamma \sum_{j=1}^{n} ||x_{j-1} - x_j||. \qquad (4)$$

2.3 Definition

$F' \in Lip_K(D)$ for some set D if and only if there exists some $K \geq 0$ such that for every pair $x,y \in D$,

$$||F'(x) - F'(y)|| \leq K\ ||x-y|| \qquad (5)$$

The following theorem shows that a method of bounded deteriation is generally locally convergent.

2.4 Theorem (Dennis, 1971)

Let $F' \in Lip_K D$ $x_0 \in D$ and let A_0 be a nonsingular NXN matrix with

$$||A_0^{-1}\ F(x_0)|| \leq \eta,\ ||A_0^{-1}|| \leq \beta$$

Assume that there are nonnegative real numbers δ and γ such that for every n for which $x_0,\ldots,x_n$ as defined by (1), are in D,

$$||A_n - F'(x_n)|| \leq \delta_n + \gamma \sum_{j=1}^{n} ||x_j - x_{j-1}||; \qquad (6)$$

$$\delta_n \leq \delta \quad \text{for} \quad n > 0\ .$$

If, in addition

$$1 > \beta\,\delta_0 + 2\beta\delta \tag{7}$$

$$\tfrac{1}{2} \geq h \equiv \frac{(2\gamma+K)\beta\eta}{(1-2\beta\delta\delta-\beta\delta_0)^2} \tag{8}$$

and

$$N(x_0,r_0) \subset D, \text{ where } r_0 = \frac{1-(1-2h)^{\frac{1}{2}}}{\beta(2\gamma+K)}(1 - 2\beta\delta - \beta\delta_0)$$

then $\{x_n\}$ generated by (1) exists in $N(x_0,r_0)$ and converges to x^*. x^* is the unique root of F in

$\overline{N}(x_0, \frac{1-(1-2h')^{\frac{1}{2}}}{\beta K}(1 - \beta\delta_0))$, where $h' \equiv \frac{\beta K\eta}{(1-\beta\delta)^2}$ and if

$h < \frac{1}{2}$, then x^* is unique in $D \cap N(x_0, \frac{1+(1-2h')^{\frac{1}{2}}}{\beta K}(1 - \beta\delta_0))$.

If one applies the previous theorem to the special case of Newton's method then $\delta_n = 0 = \gamma$ and the result is the Kantorovich theorem. Note that the existence of a solution is asserted not assumed.

3. The single rank methods

C. G. Broyden (1965) described a class of secant methods based on the following solutions to (2), the secant equation.

$$A_{n+1} = A_n - F(x_{n+1})\frac{d_n^T A_n}{d_n^T F(x_n)}$$

or

$$A_{n+1}^{-1} = A_n^{-1} - A_n^{-1} F(x_{n+1})\frac{d_n^T}{d_n^T(F(x_{n+1}) - F(x_n))} .$$

d_n must be chosen so that neither denominator is zero. Broyden (1965) suggested $d_n = F(x_{n+1}) - F(x_n)$ and $(A_n^{-1})^T A_n^{-1} F(x_n)$. These approximations are usually denoted by B_n and H_n respectively. We conform to this convention and also generalize the formulae to give a Jacobian approximation at x' given an approximation B and $H = B^{-1}$ at x.

$$B' = B + (F(x') - F(x) - B(x' - x)) \frac{d^T B}{d^T B(x'-x)} \tag{9}$$

$$H' = H - (H(F(x') - F(x)) - (x' - x))\frac{d^T}{d^T(F(x') - F(x))} \tag{10}$$

In (Dennis, 1969b) we proved the following theorem. Let $||\cdot|| = ||\cdot||_2$.

<u>3.1</u> Theorem.

Let $F \in \mathrm{Lip}_K D$ and let $x, x' \in D$ with B a nonsingular matrix. If B' is defined by (9),

$$||B' - F'(x)|| \leq q||B - F'(x)|| + K(1 + q/2)||x' - x||$$

where $q = \dfrac{||x'-x|| \cdot ||d^T B||}{|d^T B(x'-x)|}$.

This allows a straightforward induction proof of the following.

3.2 Theorem (Dennis, 1969b)

Let $F \in Lip_K D$ and let $x_0, \ldots, x_{n+1}$). $B_0, \ldots, B_{n+1}$ be generated by any single rank method. If $\{x_i : i=0,\ldots,n+1\} \subset D$, then

$$||B_{n+1} - F'(x_{n+1})|| \leq (\prod_{i=0}^{n} q_i ||B_0 - F'(x_0)|| + K \sum_{j=0}^{n} (\prod_{i=j+1}^{n} q_i)(1+q_j/2)||x_{j+1} - x_j||, \qquad (11)$$

where q_j is defined as

$$\frac{||x_{j+1} - x_j|| \cdot ||d_j^T B_j||}{|d_j^T B_j (x_{j+1} - x_j)|}, \text{ for } j \geq 0$$

and $\prod_{i=r+1}^{r} q_i = 1; r = 0,1,\ldots$.

Now in order for the single rank method to be of bounded deteriation (according to this analysis) we need $\prod_{i=0}^{n} q_i$ to be uniformly bounded. There is no problem in the case of Broyden's method, $d_j = H_j^T (x_{j+1} - x_j)$ since then $q_i = 1$, its minimum value. Otherwise, it is neccessary and sufficient that $\sum_{j=0}^{\infty} (q_j - 1)$ converge (Knopp, 1947). It is an open problem to characterize d_j in terms of the convergence of $\sum_{j=0}^{\infty} (q_j - 1)$.

Some special cases of interest will be mentioned in the next section. We finish this section with a theorem giving conditions for convergence of Broyden's method, $d_n = H_n^T(x_{n+1}-x_n)$ although it should be clear how to give a theorem for the case $\sum_{j=0}^{\infty}(q_j-1) < \infty$. This theorem strengthens one in Dennis(1969b).

3.3 Theorem.

Let $F \in Lip_K D$, $x_0 \in D$, and H_0 be a nonsingular $N \times N$ matrix bounded in norm by β. Assume also that $\eta \geq ||H_0 F(x_0)||$, and $||B_0 - F'(x_0)|| \leq \delta$. If $1 > 3\beta\delta$, $\frac{1}{8} \geq h \equiv \frac{\beta K \eta}{(1-3\beta\delta)^2}$ and $N(x_0, r_0) \subset D$ for $r_0 = \frac{1-(1-2h)^{\frac{1}{2}}}{4\beta K}(1-3\beta\delta)$ then Broyden's method with arbitrary recalculation sequence converges to x^*, a zero of F in $\bar{N}(x_0, r_0)$.

Proof: The proof consists simply of noticing that by Theorem 3.2, Broyden's method satisfies the hypotheses of Theorem 2.4 with $\delta_n = \delta$ and $\gamma = 3K/2$.

In (Dennis, 1971), we give theoretical justification as well as numerical examples in support of the error bounds

$$\frac{1}{2}||x_{n+1} - x_n|| \leq ||x^* - x_n|| \leq 4\,\frac{||H_0||}{||H_{\alpha_n}||}||x_{n+1}-x_n||$$

for the case when δ is very small and the hypotheses of the previous theorem are satisfied.

4. A class of double rank formulae

Broyden's method works very well for the solution of general nonlinear vector equations but in the unconstrained minimization problem, it suffers from the fact that even though $F'(x) = (\nabla f)'(x) = f''(x)$ is symmetric and even if H_n is, H_{n+1} need not be symmetric. In fact, a glance at (10) shows that the method with $d = \eta$ is the only single rank method with this property.

Powell (1970c) gave the following procedure. Given G, a symmetric nonsingular matrix, update $G = G^{(0)}$ using the Broyden single rank formula to obtain $G^{(1/2)}$. This Hessian approximation satisfies the secant equation (2), but it is not symmetric. Set $G^{(1)} = (G^{(1/2)} + G^{(1/2)T})/2$. $G^{(1)}$ doesn't satisfy (2), so repeat the process. Powell gives the limit of $G^{(k)}$ as

$$G^* = G + \frac{\mu\delta^T + \delta\mu^T}{||\delta||^2} - \frac{\delta^T\mu\delta\delta^T}{||\delta||^4}, \tag{12}$$

where $\delta = x' - x$, $\gamma = \nabla f(x')$, and $\mu = \gamma - G\delta$.

D. Goldfarb in a private communication was the first to notice that the analysis of the previous section could be applied to (12). Before making this analysis, we give the results obtained by applying Powell's procedure to (9) and (10)

$$G^* = G + \frac{\mu d^T G + G d\mu^T}{d^T G\delta} - \frac{\delta^T\mu G d d^T G}{(d^T G\delta)^2} \tag{13}$$

$$H^* = H - \frac{\eta d^T + d\eta^T}{d^T\gamma} + \frac{\gamma^T\eta d d^T}{(d^T\gamma)^2} \tag{14}$$

where $\eta = H\mu$. Notice that even if $GH = I$, unless we consider the special symmetric single rank case, $d = \eta$, $G^*H^* \neq I$ in general. (14) intersects the class of double rank formulae defined in Broyden (1967), only if $d = c_1 H\gamma + c_2\delta$. We could refer to (13) and (14) as "dual" update formulae.

TABLE I

Some examples of dual formulae

d	G^*	H^*
$H\delta$	$G + \frac{\mu\delta^T\delta\mu^T}{\|\|\delta\|\|^2} - \frac{\delta^T\mu\delta\delta^T}{\|\|\delta\|\|^4}$	$H - \frac{\eta\delta^T H + H\delta\eta^T}{\delta^T H\gamma} + \frac{\gamma^T\eta H\delta\delta^T H}{(\delta^T H\gamma)^2}$
$H\gamma$	$G + \frac{\mu\gamma^T + \gamma\mu^T}{\gamma^T\delta} - \frac{\delta^T\mu\gamma\gamma^T}{(\gamma^T\delta)^2}$	$H - \frac{\eta\gamma^T H + H\gamma\eta^T}{\gamma^T H\gamma} + \frac{\gamma^T\eta H\gamma\gamma^T H}{(\gamma^T H\gamma)^2}$
δ	$G + \frac{\mu\delta^T G + G\delta\mu^T}{\delta^T G\delta} - \frac{\delta^T\mu G\delta\delta^T G}{(\delta^T G\delta)^2}$	$H - \frac{\eta\delta^T + \delta\eta^T}{\delta^T\gamma} + \frac{\gamma^T\eta\delta\delta^T}{\delta^T\gamma}$
γ	$G + \frac{\mu\gamma^T G + G\gamma\mu^T}{\gamma^T G\delta} - \frac{\delta^T\mu G\gamma\gamma^T G}{\gamma^T G\delta}$	$H - \frac{\eta\gamma^T + \gamma\eta^T}{\|\|\gamma\|\|^2} + \frac{\gamma^T\eta\gamma\gamma^T}{\|\|\gamma\|\|^4}$

First consider the $H\gamma = d$ row. By straightforward algebra, one finds that $(G^*)^{-1}$ is the Davidon-Fletcher-Powell update. See Powell (1971). It's dual is an update method discovered by Greenstadt and given favorable reports in Greenstadt (1970).

H^* of the $d = \delta$ row is a method found independently by Broyden (1969), Goldfarb (1970), Shanno (1970), and Fletcher (1970). It is interesting to note that the underlying single rank method is due to Pearson. The bottom right hand element was found by Greenstadt (1970) and is the symmetrization of Broyden's poor method. We return to this table in Section 6.

Let us now proceed to analyze Newton-like methods based on (13). Powell (1970c) mentions that for his update G^* for $d = H\delta$, if f is a quadratic with Hessian $\bar{G}$, G^* satisfies

$$G^* - \bar{G} = (I - \frac{\delta\delta^T}{||\delta||^2})(G - G)(I - \frac{\delta\delta^T}{||\delta||^2}) \quad .$$

The importance of this identity is clearly that G^* is at least as good an approximation as G in the quadratic case. One can easily show that (13) satisfies the identify

$$G^* - \bar{G} = (I - \frac{Gd\delta^T}{d^TG\delta})(G - \bar{G})(I - \frac{\delta d^TG}{d^TG\delta}).$$

Note that the ℓ_2 norm of the projectors is $q \geq 1$. The following double rank analogue of Theorem 3.1 generalizes this identify to the nonlinear case.

4.1 Let $f'' \in Lip_K D$ and let G^* be determined by (13) for $x, x' \in D$, then

$$||G^* - f''(x')|| \leq q^2 \cdot ||G - f''(x)|| + (K/2)(2+q+q^2)||x'-x||$$

where $\quad q = \dfrac{||Gd|| \; ||x'-x||}{|d^TG(x'-x)|} \quad .$

<u>Proof</u>: Straightforward algebra gives

$$G^* - f''(x') = G^* - f''(x) + f''(x) - f''(x')$$

$$+ \frac{(\gamma - f''(x)\delta)d^TG + Gd(\gamma - f''(x)\delta)^T}{d^TG\delta}$$

$$- \frac{\delta^T(\gamma - f''(x)\delta)Gdd^TG}{(d^TG\delta)^2}$$

$$- \frac{(G - f''(x))\delta d^TG + G d\delta^T(G - f''(x))}{d^TG\delta}$$

$$+ \frac{\delta^T(G - f''(x))\delta \; Gdd^TG}{(d^TG\delta)^2} \quad .$$

Hence

$$G^* - f''(x') = (I - \frac{Gd\delta^T}{d^TG\delta})(G - f''(x))(I - \frac{\delta d^TG}{d^TG\delta})$$

$$- \frac{(\gamma - f''(x)\delta)d^TG}{d^TG\delta}$$

$$- \frac{Gd}{d^TG\delta}(\gamma - f''(x)\delta)^T(I - \frac{\delta d^TG}{d^TG\delta}) + f''(x) - f''(x').$$

Now $||I - \frac{Gd\delta^T}{d^TG\delta}|| = ||I - \frac{\delta d^TG}{d^TG\delta}|| = q$ (Broyden 1965)

and $||\gamma - f''(x)\delta|| \leq (K/2)||\delta||^2$ so

$$||G^* - f''(x')|| \leq q^2||G - f''(x)|| + (K/2)\frac{||\delta||^2||d^TG||}{|d^TG\delta|}$$

$$+ \frac{||Gd||}{|d^TG\delta|}(K/2)||\delta||^2 q + K||\delta||$$

and the result follows.

This time, in analogy with 3.2 we need $\prod_{i=0}^{n} q_i^2$ be uniformly bounded. Certainly for Powell's update, $q_i \equiv 1$, it is.

<u>4.2</u> Theorem.

Let $f'' \in Lip_K D$, $x_0 \in D$ and H_0 be a symmetric nonsingular NxN matrix bounded in norm by β. Assume also that $\eta \geq ||H_0 \nabla f(x_0)||$ and $||G_0 - f''(x_0)|| \leq \delta$. If $1 > 3\beta\delta$, $.1 \geq h \equiv \frac{\beta K \eta}{(1-3\beta\delta)^2}$ and $N(x_0, r_0) \subset D$ for $r_0 = \frac{1 - (1-2h)^{\frac{1}{2}}}{\delta\beta K}(1-3\beta\delta)$, then Powell's update procedure with arbitrary recalculation sequence converges to x^*, a stationary point of f in $\bar{N}(x_0, r_0)$.

<u>Proof</u>: Since the previous theorem for the case of Powell's update reduces to $||G^* - f''(x')|| \leq ||G' - f''(x)|| + 2K||x' - x||$ it needs only an easy induction argument to show that the analogue of (11) is

$$||G_{n+1} - f''(x_{n+1})|| \leq ||G_0 - f''(x_0)|| + 2K\sum_{j=0}^{n}||x_{j+1} - x_j||.$$

Now apply Theorem 2.4, with $\delta = \delta$, $\gamma = 2K$.

The same error estimates noted at the end of section 3 for Broyden's method should be of value here. They have not been tested.

5. A class of methods for nonlinear least squares.

Suppose we wish to minimize

$$f(x) = \sum_{i=1}^{M} (\phi_i(x))^2$$

where ϕ_i is a scalar function defined on a subset of E^N, $M \geq N$. Let us assume that we are willing to calculate $\nabla f(x) = 2\Phi'(x)\Phi(x)$, where $\Phi = (\psi_1,\ldots,\psi_M)^T$. One way to apply the update methods is the rather obvious way of applying a correction to an approximation to the Hessian of f. A better though less obvious way was given in Brown-Dennis (1970a). We sketch it below.

$$f''(x) = \sum_{i=1}^{M} \psi_i(x)\psi_i''(x) + \Phi'(x)^T\Phi'(x).$$

Notice that we have assumed $\Phi'(x)$ can be computed and so its ith row, $\nabla\psi_i(x)$ is available. Hence an update method can be used to approximate $\psi_i''(x_{n+1})$, given $\Phi'(x_{n+1})$ and $\Phi'(x_n)$. If one of the formulae of section 4 is used then since the $\psi_i''(x_{n+1})$ approximations will be symmetric, the algorithm requires the same storage as Newton's method applied to ∇f but none of the $\frac{NM^2}{2} + NM$ cross partials of that method. In the reference examples are given which show the method to out perform Levenberg-Marquardt when f is large at the solution ie $M>>N$, which justifies the additional arithmetic in that case. A theorem based on the analysis of section 2 is also given which shows the method to be second order when the Powell update is used and f is zero at the minimum. This last will usually correspond to the case $M=N$ and one then has another update method due to Powell (1970b) available which does not require Φ'. One also has the finite difference analogue of Levenberg-Marquardt given in Brown-Dennis (1970b) available which seems to be very good when the minimum of f is small.

6. Numerical results and discussion.

When the results of section 3 where first given it did not really seem very important whether or not Broyden's method was the only single rank method of bounded deteriation. Generally the other work on single rank methods was more concerned with the minimization problem than the solution of vector equations. The relationship between the single and double rank methods shown in section 4 makes this question of much more interest. One certainly cannot dismiss the methods given in Table I.

Whether or not q for say the Davidon-Fletcher-Powell method satisfies $Q \geq \prod_{i=0}^{n} q_i$, $n = 0,1,\ldots,$ it would be interesting to know if the update generates a locally convergent Newton-like method, that is, can one add to Powell's (1971) theorem that t_n in (3) can eventually be taken as (1)? See the computational results in this context. If a counter-example can be found, this will certainly mean that one can not be completely half-hearted in his choice of t_n.

It is also possible to be irreverent and ask if possibly one of these methods could be better than the Davidon-Fletcher-Powell, even when used in connection with iteration (3).

The examples we report are of course subject to error, and even if they are correct, they prove nothing. All the computations were done on the Cornell University IBM 360/65. The programs were written in FØRTRAN and the WATFIV compiler was used. All gradients were obtained analytically.

Problem I refers to Rosenbrock's function:

$$f(x) = 100(x_2 - x_1)^2 + (1 - x_1)^2$$

$$x_0 = (-1.2, 1.0), \; f_0 = 24.2$$

$$x^* = (1,1) \quad \min f = 0 \quad .$$

Problem II is

$$f(x) = x_1^2 + 2x_2^2 + 3x_3^2 + 4x_4^2 + (x_1+x_2+x_3+x_4)^4$$

$$x_0 = (1,-1,-1,1) \qquad f_0 = 10$$

$$x^* = (0,0,0,0) \qquad \min f = 0 \; ,$$

Algorithm I.

Algorithm I is based on Powell's MINFA routine (1971d), Powell's update, (13) with $d = H\delta$, is used to approximate the Hessian and (14) with $d = \gamma$ furnishes the inverse Hessian approximation. The program used did not have a singularity monitor.

Problem I: In 203 iterations and the same number of function and gradient calls, f was reduced to $.24 \times 10^{-10}$. The execution time was 7.79 seconds.

Problem II: In 21 iterations and the same number of function and gradient calls, f was reduced to $.14 \times 10^{-21}$. The execution time was .99 seconds.

Algorithm II

This algorithm used a modification of Powell's MINFA with Powell's Hessian approximation and the dual approximation to the inverse Hessian, ie. the first row of Table 1.

Problem I: After 50 iterations, function value and gradient calls, f was $.2 \times 10^{-9}$. The last 5 iterations were Newton-like and reduced f from about 10^{-2}. The execution time was 1.94 seconds.

Problem II: In 19 iterations f was $.8 \times 10^{-24}$. Again the algorithm seem to exhibit superlinear convergence once it was near the minimum. The execution time was .86 seconds.

Algorithm II could probably be improved by experimenting with the program parameters.

Algorithm III: This algorithm is MINFA using the bottom right hand element of Table I as the inverse Hessian approximation and its inverse matrix as the Hessian approximation.

Problem I: In 95 iteration f was only reduced to 3.6. The Hessian approximation became very nearly singular.

Problem II: In 142 iterations f was 10^{-19} but the algorithm never became Newton-like. The execution time was 5.91 seconds.

Algorithm IV This algorithm was MINFA using row two of Table I, ie, the DFP and Greenstat's methods.

Problem I: In 299 iterations f was .7 with an execution time of 5 seconds.

Problem II: In 22 iterations f was 2.3×10^{-22} with an execution time of .88 seconds.

Algorithm V This algorithm was a slight change in MINFA (2 or 3 cards were changed). It was based on Powell's update.

Problem I: 43 iterations and .71 seconds reduced f to $.36 \times 10^{-11}$ as opposed to MINFA's $f = 0$ after 49 iterations and .88 seconds. f was $.21 \times 10^{-3}$ after 45 iterations. Both versions appeared superlinear.

Problem II: 18 iterations gave $f = .95 \times 10^{-21}$ in .83 seconds.

These five algorithms are based on the assumption that the underlying method is locally convergent. Algorithm II and V both based on Broyden's method seem on the strength on admittedly thin evidence to perform better than Algorithm III which is based on the Davidon-Fletcher-Powell, Greenstadt dual pair. The next question we investigate is how these methods compare in a standard descent algorithm. The updates were all used in iteration (3) where t_n was chosen by subroutine VD02A from the A.E.R.E. Harwell subroutine library.

Powell's update failed almost immediately on both problems. The elements of H^* became about 10^{30}. The formula for H^*, given in (Powell,1970d) seems susceptible to this behavior. The dual of Powell's update performs quite well. Problem I was reduced to 0 after 21 iterations, 121 function values and gradient calculations and 1.26 seconds. Problem II was reduced to $.28\times10^{-34}$ after 12 iterations, 41 each function and gradient evaluations and .74 seconds.

Greenstadt's method gave the same performance on Problem I as the dual Powell but with an execution time of 1.51 seconds. The DFP method gave $f = 0$ in 18 iterations, 96 each function and gradient evaluations and .98 seconds. On Problem II, Greenstadt's method and the DFP reduced f to $.52\times10^{-34}$ in 10 iterations and 34 each function and gradient values in .59 seconds.

Thus the new update,

$$H^* = H - \frac{\eta\delta^T H + H\delta\eta^T}{\delta^T H\gamma} + \frac{\gamma^T\eta H\delta\delta^T H}{(\delta^T H\gamma)^2}$$

shows promise of being a very versital formula worthy of further investigation. The first project probably should be to fit this update more snugly into the MINFA framework.

7. Acknowledgements This work was supported by NSF Grant GJ-27528. I wish to thank J.J. More' for useful suggestions and R. Schumann and A. Chien for programming help. I would also like to thank M.J.D. Powell for making a listing of VD02A available.

REFERENCES

[1] K. M. Brown and J. E. Dennis (1970a), "New computational algorithms for minimizing a sum of squares of nonlinear functions", Yale University technical report, submitted for publication.

[2] ___________(1970b), "Derivative free analogues of the Levenberg-Marquardt and Gauss algorithms for nonlinear least squares approximation", IBM Phila. Scientific Center technical report, to appear in Numerische Mathematik.

[3] C. G. Broyden (1965), "A class of methods for solving nonlinear simultaneous equations, Math. Comp. 19, pp 577-593.

[4] ______________ (1967), "Quasi-Newton methods and their application to function minimization", Math. Comp. 21 pp 368-381.

[5] ______________ (1969), "A new double-rank minimization algorithm", AMS Notices 16 p 670.

[6] _______________ (1970), "The convergence of a class of double-rank minimization algorithms parts I and II", J.I.M.A.

[7] L. Collatz (1966), "Functional Analysis and Numerical Mathematics, Academic Press, New York.

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

[9] J. E. Dennis (1969a), "On the Kantorovich hypothesis for Newton's method", SIAM J. Numer. Anal. 6 pp 493-507.

[10] _____________ (1969b), "On the convergence of Broyden's method for nonlinear systems of equations", Cornell Comp. Sci. rpt 69-48, to appear in Math. Comp.

[11] _____________ (1971) "Toward a unified convergence theory for Newton-like methods" in Nonlinear Functional Analysis and Applications edited by L. B. Rall, Academic Press, New York.

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

[13] D. Goldfarb,(1970), "A family of variable-metric methods derived by variational means", Math. Comp. <u>24</u> pp 23-26.

[14] J. Greenstadt, (1970), "Variations on variable-metric methods", Math. Comp. <u>24</u> pp 1-18.

[15] W. Hofmann (1970), "Regula-falsi Verfahren in Banachräumen", Dissertation Hamburg.

[16] K. Knopp (1947), <u>Theory of functions</u>, <u>part II</u> translated by F. Bagemihl, Dover Publications, New York.

[17] M.J.D. Powell (1970a) "A hybrid method for nonlinear equations" in Numerical Methods for Nonlinear Algebraic Equations edited by Philip Rabinowitz, Gordon and Breach, London.

[18] ____________ (1970b), "A fortran subroutine for solving systems of nonlinear algebraic equations", Ibid.

[19] ____________ (1970c), "A new algorithm for unconstrained optimization" in <u>Nonlinear Programming</u>, edited by J.B. Rosen, O.L. Mangasarian, K. Ritter, Academic Press, New York.

[20] ____________ (1970d) "A Fortran subroutine for unconstrained minimization, requiring first derivatives of the objective functions", A.E.R.E. Harwell rpt. R64-69.

[21] ____________ (1971), "On the convergence of the variable metric algorithm", J.I.M.A.

[22] A. Ralston (1965), <u>A First Course in Numerical Analysis</u>, McGraw-Hill, New York.

[23] D. F. Shanno (1970), "Conditioning of Quasi-Newton methods for function minimization", Math. Comp. <u>24</u> pp 647-656.