On the Davidon-Fletcher-Powell Method for Function Minimization[1]

R. P. Tewarson[2]

Abstract. A method for improving the computations in the Davidon-Fletcher-Powell method for function minimization is suggested. It utilizes the doubly relaxed generalized inverse of the matrix which is usually obtained from the gradient vectors. The method consists of simple perturbations in the scalar terms of the correction matrix.

1. Introduction

The Davidon-Fletcher-Powell (DFP) method (Refs. 1, 2) for function minimization is one of the most popular methods. However, it has been observed that the DFP method does not always proceed smoothly. For example: Broyden (Ref. 3) remarked that occasionally negative steps had to be taken. McCormick (Ref. 4) observed that periodic reinitialization of the matrix lead to significant improvement. Wolfe (Ref. 5) has reported cases where convergence to non-stationary points had taken place. Bard (Ref. 6) had encountered similar behavior which, he observed, was invariably the result of the matrix turning singular.

In this paper we shall make use of the generalized inverses to give a technique for improving the DFP method. In Section 2 we will introduce the doubly relaxed W-generalized inverse, where W is a positive definite matrix.

[1] Paper received July , 1969. This research was supported in part by the National Aeronautics and Space Administration, Grant No. NGR-33-015-013.

[2] Professor, Applied Analysis Department, State University of New York at Stony Brook, Stony Brook, N. Y., 11790.

In Section 3, we will describe a modification in the DFP method and prove that the associated matrices are positive definite.

2. The doubly relaxed W-generalized inverse

Let B be an m x n matrix of rank r ($r \leq m \leq n$). If X is a matrix satisfying each of the equations

$$B X B = B, \quad X B X = X, \quad (BX)^T = BX \text{ and } (XB)^T = XB, \tag{2.1}$$

(where T denotes the transpose), then X is unique and is called the generalized inverse of B, viz., $X = B^+$, (Ref. 7). If

$$\Delta = BB^T + \varepsilon I_m, \tag{2.2}$$

where I_m is the identity matrix of order m and ε is a small positive number, then $B(\varepsilon)^+$, the doubly relaxed generalized inverse of B, is defined by Rutishauser (Ref. 8) as

$$B(\varepsilon)^+ = B^T(\Delta + \varepsilon\, \Delta^{-1})^{-1}. \tag{2.3}$$

Let $D = \{d_i\}$ be a non-singular diagonal matrix of order r with $d_i > 0$ as its ith diagonal element, then we have

Lemma 2.1.

$$\lim_{\epsilon \to 0} D(D^2 + \varepsilon I_r + \varepsilon (D^2 + \varepsilon I_r)^{-1})^{-1} = D^{-1}. \tag{2.4}$$

Proof. Since $D = \{d_i\}$, $d_i > 0$, we have

$$\lim_{\epsilon \to 0} D(D^2 + \varepsilon I_r + \varepsilon(D^2 + \varepsilon I_r)^{-1})^{-1} = \lim_{\epsilon \to 0} \left\{ \frac{d_i}{d_i^2 + \varepsilon + \dfrac{\varepsilon}{d_i^2 + \varepsilon}} \right\} = \{\frac{1}{d_i}\} = D^{-1}.$$

The following Theorem shows the relation between B^+ and $B(\varepsilon)^+$.

Theorem 2.1 (Rutishauser, Ref. 8).

$$B^+ = \lim_{\epsilon \to 0} B(\varepsilon)^+. \tag{2.5}$$

We give a proof of the above theorem since, in Ref. 8 it is omitted.

Proof. There exist matrices Q and S such that (Ref. 9, p. 10)

$$Q^TQ = QQ^T = I_m \; , \quad S^TS = SS^T = I_n \tag{2.6}$$

and

$$QBS = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.7}$$

where D is a non-singular, diagonal matrix of rank r. The diagonal elements of D are greater than zero and are the non-zero singular values of B. In view of (2.7), (2.2) and (2.6), we have

$$B = Q^T \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} S^T \tag{2.8}$$

$$\Rightarrow \quad \Delta = Q^T \begin{bmatrix} D^2 & 0 \\ 0 & 0 \end{bmatrix} Q + \epsilon \, Q^TQ \; ,$$

$$\Rightarrow \quad \Delta = Q^T \begin{bmatrix} D^2 + \epsilon I_r & 0 \\ 0 & \epsilon \, I_{m-r} \end{bmatrix} Q. \tag{2.9}$$

From (2.9) it follows that

$$(\Delta + \epsilon \, \Delta^{-1})^{-1} = Q^T \begin{bmatrix} \left(D^2 + \epsilon I_r + \epsilon \, (D^2 + \epsilon I_r)^{-1}\right)^{-1} & 0 \\ 0 & (1 + \epsilon)^{-1} I_{m-r} \end{bmatrix} Q$$

and from (2.3), (2.8) and Lemma 2.1, we have

$$B(\epsilon)^+ = B^T(\Delta + \epsilon \, \Delta^{-1})^{-1} = S \begin{bmatrix} D\left(D^2 + \epsilon \, I_r + \epsilon \, (D^2 + \epsilon \, I_r)^{-1}\right)^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q \; ,$$

$$\lim_{\epsilon \to 0} B(\epsilon)^+ = S \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q = B^+,$$

the last equality follows from direct substitution in (2.1) and using (2.8) and (2.6) (Ref. 10). This completes the proof of the theorem.

Rutishauser (Ref. 8) has shown theoretically and also by one numerical example that the doubly relaxed generalized inverse $B(\varepsilon)^+$ leads to better results on a computer than the direct computation of B^+, if the non-singular part of B is ill-conditioned. In order to make use of the above fact in the DFP method, we will need the following.

Since W is a positive definite matrix, there exists a non-singular lower triangular matrix R, such that

$$RR^T = W. \tag{2.10}$$

This is known as the Choleskey decomposition of W (Ref. 11, p. 229). Let A be an m x n matrix of rank r, such that

$$B = AR \; . \tag{2.11}$$

Then the unique solution X of the equations

$$AXA = A,\ XAX = X,\ (AX)^T = AX \text{ and } (XAW)^T = XAW, \tag{2.12}$$

is called the W-generalized inverse of A and is denoted by A^+_W. This definition of A^+_W was given by Herring (Ref. 12) in a slightly more general form. For our pusposes the above definition will suffice. Let the W^{-1} norm of X be defined by

$$\|X\|_{W^{-1}} = \text{trace } X^T W^{-1} X, \tag{2.13}$$

then Herring (Ref. 12) has proved the following theorem.

Theorem 2.2 (Herring). If F is a matrix with m rows, then

$$X = A^+_W F \tag{2.14}$$

is the least squares solution of the matrix equation

$$AX = F \; . \tag{2.15}$$

having the minimum W^{-1} norm.

We will also need the following theorem.

Theorem 2.3. If A^+_W and B^+ are the solutions of (2.12) and (2.1) respectively, then

$$A^+_W = RB^+. \tag{2.16}$$

Proof. By direct substitution in (2.1) and using (2.8) and (2.6), it is easy to verify that (Ref. 10)

$$B^+ = S\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q. \tag{2.17}$$

Also, from (2.11) and (2.8), it follows that

$$A = BR^{-1} = Q^T \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} S^T R^{-1}. \tag{2.18}$$

In view of (2.17), (2.18), (2.6) and (2.10), it is easy to check that RB^+ satisfies (2.12) and therefore (2.16) holds.

We conclude this section with the definition of $A(\varepsilon)^+_W$ as follows.

$$A(\varepsilon)^+_W = RB(\varepsilon)^+ , \tag{2.19}$$

which, in view of (2.3), (2.11), (2.10) and (2.2), implies that

$$A(\varepsilon)^+_W = RB^T(\Delta + \varepsilon\ \Delta^{-1})^{-1} = RR^TA^T(\Delta + \varepsilon\ \Delta^{-1})^{-1} = WA^T(\Delta + \varepsilon\ \Delta^{-1})^{-1}, \tag{2.20}$$

where

$$\Delta = BB^T + \varepsilon\ I_m = ARR^TA^T + \varepsilon\ I_m = AWA^T + \varepsilon\ I_m . \tag{2.21}$$

3. A Modification in the DFP Method.

Let us consider the problem of finding the n element column vector x that minimizes the quadratic function

$$f(x) = \tfrac{1}{2}\, x^TGx + b^Tx + c, \tag{3.1}$$

where G is a positive definite matrix, b is an n element column vector and c a constant (Ref. 13, Chap. 3 and Ref. 14).

We will also need the following theorem.

Theorem 2.3. If A^+_W and B^+ are the solutions of (2.12) and (2.1) respectively, then

$$A^+_W = RB^+. \tag{2.16}$$

Proof. By direct substitution in (2.1) and using (2.8) and (2.6), it is easy to verify that (Ref. 10)

$$B^+ = S\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q. \tag{2.17}$$

Also, from (2.11) and (2.8), it follows that

$$A = BR^{-1} = Q^T \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} S^T R^{-1}. \tag{2.18}$$

In view of (2.17), (2.18), (2.6) and (2.10), it is easy to check that RB^+ satisfies (2.12) and therefore (2.16) holds.

We conclude this section with the definition of $A(\varepsilon)^+_W$ as follows.

$$A(\epsilon)^+_W = RB(\epsilon)^+ , \tag{2.19}$$

which, in view of (2.3), (2.11), (2.10) and (2.2), implies that

$$A(\epsilon)^+_W = RB^T(\Delta + \epsilon\ \Delta^{-1})^{-1} = RR^TA^T(\Delta + \epsilon\ \Delta^{-1})^{-1} = WA^T(\Delta + \epsilon\ \Delta^{-1})^{-1}, \tag{2.20}$$

where

$$\Delta = BB^T + \epsilon\ I_m = ARR^TA^T + \epsilon\ I_m = AWA^T + \epsilon\ I_m\ . \tag{2.21}$$

3. A Modification in the DFP Method.

Let us consider the problem of finding the n element column vector x that minimizes the quadratic function

$$f(x) = \tfrac{1}{2}\, x^TGx + b^Tx + c, \tag{3.1}$$

where G is a positive definite matrix, b is an n element column vector and c a constant (Ref. 13, Chap. 3 and Ref. 14).

Let the i^{th} approximation to the vector which minimizes (3.1) be denoted by x_i. Then in (3.1), the gradient of $f(x)$ at x_i is given by

$$g_i = Gx_i + b, \tag{3.2}$$

which implies that

$$g_{i+1} - g_i = G(x_{i+1} - x_i) . \tag{3.3}$$

If we let

$$g_{i+1} - g_i = y_i^T \text{ and } x_{i+1} - x_i = s_i^T , \tag{3.4}$$

then (3.3) can be written as

$$y_i^T = G\, s_i^T \Rightarrow y_i = s_i\, G. \tag{3.5}$$

Let

$$Y_i = \begin{bmatrix} y_o \\ \cdot \\ \cdot \\ \cdot \\ y_{i-1} \end{bmatrix} \text{and } S_i = \begin{bmatrix} s_o \\ \cdot \\ \cdot \\ \cdot \\ s_{i-1} \end{bmatrix} . \tag{3.6}$$

Pearson (Ref. 14) gives the following algorithm for the minimization of (3.1).

Algorithm 3.1. Let P be a positive definite matrix. Given x_o and $H_o = P$. Solve for H_i, the equation

$$Y_i\, H_i = S_i \tag{3.7}$$

and determine x_{i+1} from the relation

$$f(x_{i+1}) = \min_{\alpha_i} f(x_i + \alpha_i H_i g_i) .$$

Compute g_{i+1} and using x_{i+1}, update Y_i and S_i and (3.7) as follows

$$Y_{i+1} = \begin{bmatrix} Y_i \\ y_i \end{bmatrix} \text{and } S_{i+1} = \begin{bmatrix} S_i \\ s_i \end{bmatrix}, \tag{3.8}$$

$$Y_{i+1} H_{i+1} = S_{i+1} . \tag{3.9}$$

It is proved in Ref. 14, that the above algorithm terminates for $i \le n$, if the solution of (3.7) is taken as

$$H_i = (Y_i)^+_W S_i + (I_n - (Y_i)^+_{\hat{W}} Y_i)^{-1} P \tag{3.10}$$

and $W = P$ or G^{-1} and $\hat{W} = P$ or G^{-1}. In case $W = G^{-1}$ and $\hat{W} = P$, we get the Davidon-Fletcher-Powell method.

We are now in a position to describe a modification to the DFP method. To this end, let

$$H_{i+1} = H_i + C_i \, . \tag{3.11}$$

In view of (3.9), (3.11), (3.8) and (3.7), we have

$$Y_{i+1} C_i = S_{i+1} Y_{i+1} H_i = \begin{bmatrix} S_i \\ s_i \end{bmatrix} - \begin{bmatrix} Y_i \\ y_i \end{bmatrix} H_i = \begin{bmatrix} S_i - Y_i H_i \\ s_i - y_i H_i \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ s_i - y_i H_i \end{bmatrix}. \tag{3.12}$$

Let

$$Y_{i+1} \bar{C}_i = \begin{bmatrix} 0 \\ s_i \end{bmatrix} \text{ and } Y_{i+1} \hat{C}_i = \begin{bmatrix} 0 \\ y_i H_i \end{bmatrix}, \tag{3.13}$$

then from (3.12), it follows that

$$C_i = \bar{C}_i - \hat{C}_i . \tag{3.14}$$

Also, in view of Theorem 2.2, (3.13) implies that

$$\bar{C}_i = (Y_{i+1})^+_W \begin{bmatrix} 0 \\ s_i \end{bmatrix} \text{ and } \hat{C}_i = (Y_{i+1})^+_{\hat{W}} \begin{bmatrix} 0 \\ y_i H_i \end{bmatrix} \tag{3.15}$$

are the least squares solutions of the two equations in (3.13) with the minimum W^{-1} and $\hat{W}^{-1}$ norms respectively. We will need the following theorem.

Theorem 3.1. If in (3.15), $(Y_{i+1})^T_W$ and $(Y_{i+1})^+_{\hat{W}}$ are replaced by $Y_{i+1}(\epsilon)^+_W$ and $Y_{i+1}(\epsilon)^+_{\hat{W}}$ respectively and

$$y_i W Y_i^T = 0 \text{ and } y_i \hat{W} Y_i^T = 0 , \tag{3.16}$$

then

$$\bar{C}_i = W y_i^T s_i (\alpha + \epsilon\, \alpha^{-1})^{-1} \text{ and } \hat{C}_i = \hat{W} y_i^T y_i H_i (\hat{\alpha} + \epsilon\, \hat{\alpha}^{-1})^{-1} , \tag{3.17}$$

where

$$\alpha = y_i W y_i^T + \epsilon \text{ and } \hat{\alpha} = y_i \hat{W} y_i^T + \epsilon \tag{3.18}$$

Proof. From (2.20), (2.21), (3.8) and (3.16), we have

$$Y_{i+1}(\epsilon)^+_W = W Y_{i+1}^T (\Delta_{i+1} + \epsilon\, \Delta_{i+1}^{-1})^{-1} , \tag{3.19}$$

where

$$\Delta_{i+1} = Y_{i+1} W Y_{i+1}^T + \epsilon\, I_{i+1} = \begin{bmatrix} Y_i \\ y_i \end{bmatrix} W \begin{bmatrix} Y_i^T , & y_i^T \end{bmatrix} + \epsilon\, I_{i+1}$$

$$= \begin{bmatrix} Y_i W Y_i^T & Y_i W y_i^T \\ y_i W Y_i^T & y_i W y_i^T \end{bmatrix} + \epsilon\, I_{i+1} = \begin{bmatrix} Y_i W Y_i^T + \epsilon\, I_i & 0 \\ 0 & y_i W y_i^T + \epsilon \end{bmatrix} ,$$

and using (3.18), we get

$$\Delta_{i+1} = \begin{bmatrix} \Delta_i & 0 \\ 0 & \alpha \end{bmatrix} , \tag{3.20}$$

where $\Delta_i = Y_i W Y_i^T + \epsilon\, I_i$. Now, in view of (3.8), equations (3.19) and (3.20) give

$$Y_{i+1}(\epsilon)^+_W = W Y_{i+1}^T \left[\begin{bmatrix} \Delta_i & 0 \\ 0 & \alpha^{-1} \end{bmatrix} + \epsilon \begin{bmatrix} \Delta_i^{-1} & 0 \\ 0 & \alpha^{-1} \end{bmatrix} \right]^{-1}$$

$$= W(Y_i^T, y_i^T) \begin{bmatrix} (\Delta_i + \epsilon\, \Delta_i^{-1})^{-1} & 0 \\ 0 & (\alpha + \epsilon\, \alpha^{-1})^{-1} \end{bmatrix}$$

$$= W\left(Y_i^T (\Delta_i + \epsilon\, \Delta_i^{-1})^{-1} , \; y_i^T (\alpha + \epsilon\, \alpha^{-1})^{-1} \right) . \tag{3.21}$$

Now, from the hypothesis of theorem, (3.15) and (3.21), it follows that

$$\overline{C}_i = Wy_i^T s_i (\alpha + \epsilon\, \alpha^{-1})^{-1}.$$

Replacing W by $\hat{W}$ in (3.21), we get the value of $\hat{C}$ given by (3.17). This completes the proof of the theorem.

The following corollary to the above theorem gives the desired re-sired result.

Corollary 3.1. If, in Theorem 3.1, $W = G^{-1}$ and $\hat{W} = H_i$, then

$$H_{i+1} = H_i + \frac{s_i^T s_i}{y_i s_i^T + \epsilon + \epsilon(y_i s_i^T + \epsilon)^{-1}} - \frac{H_i y_i^T y_i H_i}{y_i H_i y_i^T + \epsilon + \epsilon(y_i H_i y_i^T + \epsilon)^{-1}}. \quad (3.22)$$

Proof: Since, in view of (3.5), $Wy_i^T = G^{-1}y_i^T = s_i^T$, and $\hat{W}y_i^T = H_i y_i^T$; equation (3.22) follows from (3.17), (3.18), (3.14) and (3.11).

It is easy to see that (3.16) is satisfied if $\hat{W} = G^{-1}$ and $\hat{W} = H_i$, because in view of (3.5) and (3.7), we have $y_i G^{-1} Y_i^T = s_i G S_i^T$ and $y_i H_i Y_i^T = s_i G S_i^T$. Therefore, in this case (3.16) implies that $s_i G s_i^T = 0$, $j < i$, which is known to be satisfied (Ref. 13, Chap. 3).

The choice of H_i for $\hat{W}$ in Corollary 3.1 is justified provided that H_i given by (3.22) is positive definite. It is easy to see that for $\epsilon = 0$, equation (3.22) is the usual updating formula for the DFP method and H_i is known to be positive definite (Ref. 13, Chap. 3). For $\epsilon > 0$, we have

Theorem 3.3. If $H_o = P$, then the H_i given by (3.22) are positive definite for all i.

Proof. Since $H_o = P$ is positive definite, we will show that whenever H_i is positive definite H_{i+1} is also positive definite; then by induction the theorem is proved. Let H_i be positive definite, then

$$\delta = \epsilon + \frac{\epsilon}{y_i H_i y_i^T + \epsilon} > 0$$

By the Cauchy-Schwartz inequality for an arbitrary n dimensional row vector u with $u^T u \neq 0$, we have

$$(y_i H_i u^T)^2 \leq (y_i H_i y_i^T)(uH_i u^T) < (y_i H_i y_i^T + \delta)(uH_i u^T),$$

which implies that

$$\frac{(y_i H_i u^T)^2}{y_i H_i y_i^T + \delta} < uH_i u^T. \tag{3.23}$$

But, from (3.22) and (3.23) it follows that

$$u H_{i+1} u^T = \frac{(s_i u^T)^2}{y_i s_i^T + \epsilon + \epsilon(y_i s_i^T + \epsilon)^{-1}} + uH_i u^T - \frac{(y_i H_i u^T)^2}{y_i H_i y_i^T + \delta} .$$

$$> 0, \text{ if } y_i s_i^T + \epsilon > 0 .$$

Since $\epsilon > 0$ and in view of (3.5) and the fact that G is positive definite $y_i s_i^T = s_i G s_i^T > 0$; therefore $y_i s_i^T + \epsilon > 0$, and thus we have proved that if H_i is positive definite, then H_{i+1} is also positive definite and this completes the proof of the theorem.

4. Concluding Remarks.

In this paper we have given a technique (3.22) for improving the computation of the H_i matrix in the DFP method. In the derivation of (3.22), we made use of the doubly relaxed W-generalized inverse $Y_{i+1}(\epsilon)^+_W$. Since R is a non-singular matrix from Ref. 8, (2.16) and (2.19), it follows that, in general, $Y_{i+1}(\epsilon)^+_W$ will give better results than $(Y_{i+1})^+_W$. This is especially true, if due to round-off errors etc., the rows of Y_{i+1} are not linearly independent (Ref. 6). Rutishauser (Ref. 8) observes that the choice $10^{-5} \leq \epsilon \leq 10^{-10}$ lead to good results in a computer with 35 bit mantissa when he computed the boubly relaxed generalized inverse. The proper value for ϵ will have to be determined on the basis of large scale numerical experimentation.

We can also use the doubly relaxed W-generalized inverse in the periodic computation of H_i directly from (3.10). It is known that such periodic direct computation of H_i improves the performance of the DFP method (Ref. 4). Thus in (3.10), replacing $(Y_i)^+_W$ and $(Y_i)^+_{\hat{W}}$ by $Y_i(\epsilon)^+_W$ and $Y_i(\epsilon)^+_{\hat{W}}$ respectively and in view of the fact that $W = G^{-1}$ and $\hat{W} = P$ for the DFP method, we get

$$H_i = Y_i(\epsilon)^+_{G^{-1}} S_i + (I_n - Y_i(\epsilon)^+_P Y_i) P \quad , \tag{3.24}$$

where, in view of (3.19) and (3.5), we have

$$Y_i(\epsilon)^+_{G^{-1}} = S_i^T(\Delta_i + \epsilon \Delta_i^{-1})^{-1} \quad , \tag{3.25}$$

$$\Delta_i = S_i Y_i^T + \epsilon I_i \quad , \tag{3.26}$$

$$Y_i(\epsilon)^+_P = PY_i^T(\hat{\Delta}_i + \epsilon \hat{\Delta}_i^{-1})^{-1} \tag{3.27}$$

and

$$\hat{\Delta}_i = Y_i \, P \, Y_i^T + \epsilon \, I_i \quad . \tag{3.28}$$

Note that (3.24) can be computed even if Y_i does not have full row rank, which would not be possible in Ref. 14. In view of the theoretical results in Ref. 8, equations (3.24)-(3.28) should in general, lead to better H_i.

We conclude this paper with the following remarks on H_i which is given by (3.22). The H_i for the DFP method is

$$H_{i+1} = H_i + \frac{s_i^T \, s_i}{y_i \, s_i^T} - \frac{H_i \, y_i^T \, y_i \, H_i}{y_i \, H_i \, y_i^T} \tag{3.29}$$

If $y_i \, s_i^T$ is small then evidently $\frac{s_i^T \, s_i^T}{y_i \, s_i^T}$ dominates $\frac{H_i \, y_i^T \, y_i \, H_i}{y_i \, H_i \, y_i^T}$, this leads to inaccuracy in H_{i+1} (Ref. 6). Let $y_i \, s_i^T = o(\epsilon) = \lambda \, \epsilon$, where λ is a constant, then

$$\lim_{\epsilon \to 0} \frac{s_i^T \, s_i}{y_i \, s_i^T} = \infty \quad .$$

On the other hand

$$\lim_{\epsilon \to 0} (y_i \, s_i^T + \epsilon + \epsilon(y_i \, s_i^T + \epsilon)^{-1}) = \lim_{\epsilon \to 0} ((\lambda + 1) \, \epsilon + (\lambda + 1)^{-1}) = (\lambda + 1)^{-1} .$$

The perturbation in $y_i \, H_i \, y_i^T$ can be similarly justified.

REFERENCES

1. Davidon, W. C., Variable Metric Methods for Minimization, Argonne National Laboratory, ANL-5990(Rev.), November, 1959.

2. Fletcher, R. and Powell, M.J.D., A Rapidly Convergent Descent Method for Minimization, Computer J., Vol. 6, No. 2, 1963.

3. Broyden, C. G., Quasi-Newton Methods and their Application to Function Minimization, Math. Comp., Vol. 21, No. 99, 1967.

4. McCormick, G., Informal Remarks on Unconstrained Optimization, SIGMAP Workshop, New York, 1967.

5. Wolfe, P., Informal Remarks on Unconstrained Optimization, SIGMAP Workshop, New York, 1967.

6. Bard, Y., On Numerical Instability of Davidon-like Methods, Math. Comp., Vol. 22, No. 103, 1968.

7. Penrose, R., A Generalized Inverse for Matrices, Proc. Cambridge Philos. Soc., Vol. 51, No. 3, 1955.

8. Rutishauser, H., Once Again: The Least Square Problem, Linear Algebra and Appl., Vol. 1, No. 4, 1968.

9. Forsythe, G. E. and Moler, C. B., Computer Solution of Linear Algebraic Systems, Prencite Hall, Inc., Englewood Cliffs, N. J., 1967.

10. Tewarson, R. P., On Some Representations of Generalized Inverses, SIAM Rev., Vol. 11, No. 2, 1969.

11. Wilkinson, J. H., The Algebraic Eigenvalue Problem, Oxford University Press, Oxford, 1965.

12. Herring, G. P., Generalized Interpolation and Pseudoinverse, SIAM J. Numer. Anal., Vol. 4, No. 4, 1967.

13. Kowalik, J., and Osborne, M. R., Methods for Unconstrained Optimization Problems, American Elsevier Pub. Co., New York, 1968.

14. Pearson, J. D., Variable Metric Methods of Minimization, Computer J., Vol. 12, No. 2, 1969.