

# Variations on Variable-Metric Methods

By J. Greenstadt

(Appendix by Y. Bard)

**Abstract.** In unconstrained minimization of a function  $f$ , the method of Davidon-Fletcher-Powell (a "variable-metric" method) enables the inverse of the Hessian  $H$  of  $f$  to be approximated stepwise, using only values of the gradient of  $f$ . It is shown here that, by solving a certain variational problem, formulas for the successive corrections to  $H$  can be derived which closely resemble Davidon's. A symmetric correction matrix is sought which minimizes a weighted Euclidean norm, and also satisfies the "DFP condition." Numerical tests are described, comparing the performance (on four "standard" test functions) of two variationally-derived formulas with Davidon's. A proof by Y. Bard, modelled on Fletcher and Powell's, showing that the new formulas give the exact  $H$  after  $N$  steps, is included in an appendix.

**1. The DFP Method.** The class of gradient methods for finding the unconstrained minimum of a function  $f(x)^*$  in which the direction  $s_k$  of the next iterative step from  $x_k$  to  $x_{k+1}$  is computed from a formula such as:

$$(1-1) \quad s_k = -G_k^{-1}g_k$$

is called the class of variable-metric methods. Here  $G_k$  is a (preferably) positive-definite  $N \times N$  matrix and  $g_k$  is the gradient  $\nabla f$  evaluated at  $x_k$ .

The reason for this nomenclature is that  $s_k$  is the direction in which the directional derivative of  $f$  is a minimum, i.e., the direction in which

$$(1-2) \quad s_k^T g_k \equiv s_k^T (\nabla f)_k = \text{minimum}$$

subject to the length of  $S_k$  being constant:

$$(1-3) \quad \|s_k\| = \text{constant}.$$

Usually, the length of  $s_k$  is given in terms of a quadratic form involving a metric matrix (or tensor)  $G$  (of order  $N \times N$ ), so that:

$$(1-4) \quad \|s\|^2 = s^T G s.$$

Then, it can easily be shown [2] that the solution to the problem is given by Eq. (1-1). When  $G$  varies from point to point (as in Newton's method), the "metric" is variable, hence the name.

Davidon's well-known method [3] was called by him a *variable-metric* method because it has this feature of using a changing "inverse matrix" from one step to the next. Fletcher and Powell [4] were able to simplify Davidon's method, and to clarify many of its features and characteristics.

---

Received August 27, 1968, revised June 30, 1969.

*AMS Subject Classifications.* Primary 30; Secondary 10.

*Key Words and Phrases.* Variable-metric, Davidon method, unconstrained minimization, variational method.

\* By which we mean  $f(x_1, x_2, \dots, x_N)$ .  $f$  is assumed twice-differentiable.

The Davidon-Fletcher-Powell (or DFP) method is very closely related to Newton's method. If  $G_k$  were equal to the Hessian matrix of  $f$ , viz.,

$$(1-5) \quad G_k = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j} \right]_{x=x_k}$$

then Eq. (1-1) would be the Newton formula used for finding the root of  $\nabla f = 0$ . Since Newton's method is usually by far the most efficient of the gradient methods [5], it would be very nice to have available at each step all of the quantities needed to evaluate  $s_k$  from (1-1). It is usually rather inconvenient, however, and sometimes not feasible, to calculate so many second derivatives.

In the DFP method, a sequence of progressive estimates  $\{H_k\}$  is made of the inverse Hessian  $G^{-1}$ , based only on the first derivatives of  $f$ . The sequence of steps in a cycle is as follows: From the calculated gradient  $g_k$  at  $x_k$ , the next step direction is computed using the current estimate for  $G^{-1}$ , so that:

$$(1-6) \quad s_k = -H_k g_k .$$

Then the minimum of  $f$  is found along the direction  $s_k$ . Let the total step  $\sigma_k$  to this point be a multiple  $\alpha_k$  of  $s_k$ , i.e.,

$$(1-7) \quad \sigma_k = \alpha_k s_k$$

and

$$(1-8) \quad x_{k+1} = x_k + \sigma_k .$$

We then define

$$(1-9) \quad y_k \equiv g_{k+1} - g_k .$$

The correction to  $H_k$ , to form the next estimate,  $H_{k+1}$ , is as follows:

$$(1-10) \quad H_{k+1} = H_k + \frac{\sigma_k \sigma_k^T}{(\sigma_k^T y_k)} - \frac{H_k y_k y_k^T H_k}{(y_k^T H_k y_k)}$$

and, using this new  $H$ , the whole cycle is repeated.

As emphasized by Fletcher and Powell, a full appreciation of the significance of formula (1-10) rests on an analysis of functions  $f$  which are exactly quadratic. For such a function the Hessian  $G$  is a constant matrix, so that certain exact relationships are valid among the various quantities involved. In particular, they show that the first part of the correction in (1-10) follows from the form of the spectral resolution of  $G^{-1}$ . The second part is related to a very important requirement on  $H$ , viz., that it should satisfy a relationship derivable for a function which is an exact quadratic.\*\*

Let us therefore consider a quadratic  $f$  of the form:

$$(1-11) \quad f = f_0 + g_0^T x + \frac{1}{2} x^T G_0 x ,$$

where  $f_0$ ,  $g_0$ , and  $G_0$  are all constants. We then have

$$(1-12) \quad g = \nabla f = g_0 + G_0 x .$$

\*\* Alternative corrections to  $H_k$  have been derived by Davidon [8], Broyden [9] and Wolfe [10]. These differ from the DFP correction in that they are of rank unity ("rank-one" corrections).

Now, since  $g_0$  is not known, but only  $g_k$  (for  $x = x_k$ ), we can eliminate  $g_0$  by differencing two equations of the form of (1-12) which hold at  $x_k$  and  $x_{k+1}$ , respectively. We obtain

$$(1-13) \quad \begin{aligned} g_{k+1} - g_k &= (g_0 + G_0 x_{k+1}) - (g_0 + G_0 x_k) \\ &= G_0(x_{k+1} - x_k) \end{aligned}$$

so that if we define  $y_k \equiv g_{k+1} - g_k$  and  $\sigma_k \equiv x_{k+1} - x_k$  (as in (1-8)), we can write

$$(1-14) \quad y_k = G_0 \sigma_k$$

or in terms of  $H$ , the inverse of  $G_0$

$$(1-15) \quad H y_k = \sigma_k .$$

It was shown by Fletcher and Powell that the second term in the correction in (1-10) is the simplest way of making the  $(k + 1)$ st estimate of  $H$  obey Eq. (1-15), namely:

$$(1-16) \quad H_{k+1} y_k = \sigma_k$$

as well as satisfying the conditions already mentioned. This condition is central to what follows in this paper, and we shall call it the DFP condition.

**2. Variational Formulation.** Let us now rephrase the variable-metric problem as follows: We wish to find a correction  $E_k$  to the estimate  $H_k$  of the inverse Hessian, as follows:

$$(2-1) \quad H_{k+1} = H_k + E_k$$

so that  $H_{k+1}$  will satisfy (1-16). Since  $E_k$  is not thereby rendered unique, we need another principle, or criterion, to define it more precisely.

Let us ask for the "best" correction  $E_k$  in some sense. There are many possible choices to make, but a good one is to ask for the *smallest* correction  $E_k$ , in the sense of some norm. To a certain extent, this would tend to keep the elements of  $H$  from growing too large, which might cause an undesirable instability.

The simplest type of norm, and one which would be expected to lead to simple solutions for  $E$ , is a quadratic form in the components of  $E$ . The most general form of this kind is

$$(2-2) \quad N_G(E) = \sum_{ijkm} C_{ijkm} E_{ij} E_{km} .$$

However, investigation of this general norm [6] yielded rather unsatisfactory, complicated formulas for  $E$ , which seemed to involve an amount of calculation comparable to that of calculating the Hessian directly.

The simplest quadratic norm is, of course, the Euclidean norm, given by

$$(2-3) \quad N_E(E) = \sum_{ij} E_{ij}^2 = \text{Tr} (EE^T)$$

where the symbol  $\text{Tr}$  indicates the trace. Since this is too specialized,  $E$  was first transformed as follows:

$$(2-4) \quad F = AEB$$

and the Euclidean norm of  $F$  was calculated:

$$(2-5) \quad \begin{aligned} N_E(F) &= \text{Tr}(FF^T) = \text{Tr}(AEBB^TE^TA^T) \\ &= \text{Tr}(A^TAE BB^TE^T). \end{aligned}$$

This form also rendered a solution for  $E$  too complicated; however, when  $A = B$ , and  $A^TA$  is denoted by  $W$  (a positive-definite symmetric matrix), the problem turned out to have a rather simple solution.

Hence, we shall take for  $N(E_k)$ :

$$(2-6) \quad N(E_k) = \text{Tr}(WE_kWE_k^T)$$

and we shall try to minimize  $N(E_k)$ , subject to (1-16), and to a symmetry condition:

$$(2-7) \quad E_k^T - E_k = 0$$

which is required because the Hessian is symmetric, and we wish to preserve the symmetry of  $H$  (when we start with a symmetric first guess). We shall rewrite (1-16) in terms of  $E_k$ :

$$(2-8a) \quad \sigma_k = H_{k+1}y_k = (H_k + E_k)y_k$$

which reduces to:

$$(2-8b) \quad E_k y_k = \sigma_k - H_k y_k \equiv r_k.$$

In the remainder of this derivation, we shall ignore the subscript  $k$ .

We shall solve this constrained minimization problem by the use of Lagrange multipliers. We form the composite function  $\Phi$  as follows:

$$(2-9) \quad \Phi \equiv \frac{1}{2} \text{Tr}(WEWE^T) + \lambda^T(Ey - r) + \text{Tr}[\Gamma(E - E^T)].$$

We also note that

$$(2-10) \quad \lambda^T(Ey - r) = \text{Tr}[(Ey - r)\lambda^T].$$

Our next task is to differentiate  $\Phi$  with respect to  $E$ . We note that

$$(2-11a) \quad \frac{\partial}{\partial E} [\text{Tr}(EA)] = \left\{ \frac{\partial}{\partial E_{km}} \sum_{ij} E_{ij} A_{ji} \right\} = \{A_{mk}\} = A^T$$

and

$$(2-11b) \quad \frac{\partial}{\partial E} [\text{Tr}(E^T A)] = \left\{ \frac{\partial}{\partial E_{km}} \sum_{ij} E_{ji} A_{ji} \right\} = \{A_{km}\} = A.$$

Hence, we have

$$(2-12) \quad \frac{\partial \Phi}{\partial E} = WEW + \lambda y^T + \Gamma^T - \Gamma = 0$$

so that

$$(2-13) \quad E = -M[\lambda y^T + \Gamma^T - \Gamma]M$$

where  $M \equiv W^{-1}$ . Transposing  $E$ , we have

$$(2-14) \quad E^T = -M[y\lambda^T + \Gamma - \Gamma^T]M$$

since  $M$  is symmetric. Subtracting  $E^T$  from  $E$  should give zero, so that

$$(2-15) \quad E - E^T = -M\{\lambda y^T - y\lambda^T + 2\Gamma^T - 2\Gamma\}M = 0$$

and we have

$$(2-16) \quad \Gamma^T - \Gamma = \frac{1}{2} (y\lambda^T - \lambda y^T).$$

Substituting this into (2-13) gives

$$(2-17) \quad \begin{aligned} E &= -M\{\lambda y^T + \frac{1}{2} (y\lambda^T - \lambda y^T)\}M \\ &= -\frac{1}{2} M\{y\lambda^T + \lambda y^T\}M. \end{aligned}$$

Now we take note of the DFP condition; Eq. (2-8b):

$$(2-18) \quad Ey - r = -\frac{1}{2} M[y\lambda^T + \lambda y^T]My - r = 0.$$

Premultiplying by  $2W$ , we have

$$(2-19) \quad (y\lambda^T + \lambda y^T)My + 2Wr = 0$$

from which we solve for the  $\lambda$  which is free from the inner product. The result is

$$(2-20) \quad \lambda = -(y^T My)^{-1}[2Wr + y(\lambda^T My)].$$

We now premultiply by  $y^T M$  to obtain:

$$(2-21) \quad y^T M\lambda = -(y^T My)^{-1}[2(y^T r) + (y^T My)(\lambda^T My)]$$

and, since  $y^T M\lambda$  is the same as  $\lambda^T My$ , we can solve for  $\lambda^T My$ . The result is

$$(2-22) \quad \lambda^T My = -(y^T My)^{-1}(y^T r).$$

We now substitute this back into (2-20) to obtain:

$$(2-23) \quad \begin{aligned} \lambda &= -(y^T My)^{-1}[2Wr - (y^T My)^{-1}(y^T r)y] \\ &= (y^T My)^{-2}(y^T r)y - 2(y^T My)^{-1}Wr \end{aligned}$$

and we are in a position to replace  $\lambda$  in Eq. (2-17). We then have for  $E$

$$(2-24) \quad E = \frac{1}{(y^T My)} \left\{ ry^T M + Myr^T - \left( \frac{y^T r}{y^T My} \right) Myy^T M \right\}$$

and, finally, replacing  $r$  by  $\sigma - Hy$ , we obtain

$$(2-25) \quad \begin{aligned} E &= \frac{1}{(y^T My)} \left\{ \sigma y^T M + My\sigma^T - Hyy^T M - Myy^T H \right. \\ &\quad \left. - \frac{1}{(y^T My)} [(y^T \sigma) - (y^T Hy)] Myy^T M \right\} \end{aligned}$$

which is our final formula for  $E$ . The two obvious choices for the weighting matrix  $W$ , both of which lead to relatively simple formulas for  $E$  are

$$(2-26a) \quad W^{-1} \equiv M = H,$$

$$(2-26b) \quad W = I.$$

We obtain for  $E$  respectively

$$(2-27a) \quad E_I = \frac{1}{(y^T H y)} \left\{ \sigma y^T H + H y \sigma^T - \left[ 1 + \left( \frac{y^T \sigma}{y^T H y} \right) \right] H y y^T H \right\}$$

$$(2-27b) \quad E_{II} = \frac{1}{(y^T y)} \left\{ \sigma y^T + y \sigma^T - H y y^T - y y^T H - \frac{1}{(y^T y)} [(y^T \sigma) - (y^T H y)] y y^T \right\}.$$

The DFP correction, for comparison, is

$$(2-28) \quad E_D = \frac{1}{(y^T \sigma)} \sigma \sigma^T - \frac{1}{(y^T H y)} H y y^T H$$

so that  $E_I$ , to some extent, resembles  $E_D$ .

The resemblance between  $E_I$  and  $E_D$  goes deeper than mere appearance. Fletcher and Powell showed that for a quadratic function  $f$ , the DFP formula for  $E$  would lead to an exact solution of the minimization problem in  $N$  steps, and that the value of  $H$  attained at that point would be exactly  $G_0^{-1}$ . The Appendix of this paper contains a proof, by the author's colleague, Dr. Y. Bard, that  $E_I$  also has this desirable property. In the experimental tests, to be described in Section 4, this is borne out.

**3. The Problem of Stability.** The derivative of  $f$  in the direction  $s_k$  at  $x_k$ , is proportional to the expression in Eq. (1-2), i.e.,

$$(3-1) \quad (df/dt)_k = \mu_k s_k^T g_k$$

where  $t$  is any parameter measured along  $s_k$ , and  $\mu$  is some positive number independent of  $s_k$ . When  $s_k$  is found from a formula such as (1-6), we have

$$(3-2) \quad (df/dt)_k = -\mu_k g_k^T H_k g_k$$

which is a quadratic form in the components of  $g_k$ .

Fletcher and Powell showed that, in the DFP method,  $H_k$  is positive definite for any  $k$ , provided that  $H_0$  is positive definite, and that the line search for minimum  $f$  along  $s_k$  is carried out to sufficient accuracy. This is a very good property of the DFP estimates for  $G^{-1}$  since it guarantees that *some* progress can be made at each step in decreasing  $f$ . Fletcher and Powell called this *stability*.

Neither of the correction formulas (2-27) has this desirable property; this shows up in the numerical trials described in Section 4, in which it frequently happened that  $(df/dt)_k > 0$ , when it was necessary to reverse  $s_k$ , i.e., to go backwards in order to make  $f$  decrease.

The question can now be raised: Is it not possible to formulate a "best" correction problem which will have some sort of stability, i.e., some guarantee that  $df/dt < 0$  in the direction calculated from  $H_{k+1}$ ? This can be done, but it is a problem involving an *inequality* constraint, which follows from the condition on  $df/dt$ .

Let us assume that we had made a step  $s_k$  from  $x_k$  to  $x_{k+1}$ , have evaluated  $g_{k+1}$ , and have somehow calculated  $H_{k+1}$ , and, from it

$$(3-3) \quad s_{k+1} = -H_{k+1} g_{k+1}.$$

We wish now to assure that, in accordance with the requirement that  $(df/dt)_{k+1} < 0$ , we have

$$(3-4) \quad -g_{k+1}^T H_{k+1} g_{k+1} < 0 .$$

In order to make the conditions independent of the scale of  $g_{k+1}$ , and to allow a little leeway, we shall instead require:

$$(3-5) \quad -g_{k+1}^T H_{k+1} g_{k+1} \leq -\omega g_{k+1}^T g_{k+1}$$

where  $\omega$  is a small number. (It may not always be possible to achieve (3-5), but this would only occur if  $x_{k+1}$  were at a true stationary point.)

Inequality (3-5) can be changed into an equality constraint by using a device due to Klein [7]. We introduce a new variable  $u$ , and set

$$(3-6) \quad g_{k+1}^T H_{k+1} g_{k+1} - \omega g_{k+1}^T g_{k+1} - u^2 = 0 .$$

From this point on, we shall drop the subscripts, and denote quantities associated with  $x_{k+1}$  by a subscript asterisk. Those associated with  $x_k$  will be unmarked.

We now replace  $H_*$  ( $\equiv H_{k+1}$ ) as follows:

$$(3-7) \quad H_* = H + E$$

and Eq. (3-6) becomes:

$$(3-8) \quad g_*^T E g_* + \kappa - u^2 = 0$$

where

$$(3-9) \quad \kappa = g_*^T (H - \omega I) g_* .$$

The composite function analogous to that in Eq. (2-9) is:

$$(3-10) \quad \Phi \equiv \frac{1}{2} \text{Tr} (W E W E^T) + \text{Tr} [(E y - r) \lambda^T] \\ + \text{Tr} [\Gamma (E - E^T)] + \xi (g_*^T E g_* + \kappa - u^2)$$

and we have

$$(3-11a) \quad \partial \Phi / \partial E = W E W + \lambda y^T + \Gamma^T - \Gamma + \xi g_* g_*^T = 0$$

$$(3-11b) \quad \partial \Phi / \partial u = -2 \xi u = 0$$

together with

$$(3-11c) \quad \partial \Phi / \partial \lambda = E y - r = 0$$

$$(3-11d) \quad \partial \Phi / \partial \Gamma = E^T - E = 0 .$$

By the same sort of manipulations as used before, the solution for  $E$  is

$$(3-12) \quad E = \tau^{-1} \{ r y^T M + M y r^T - \tau^{-1} (y^T r - \xi \epsilon^2) M y y^T M \\ - \xi \epsilon [M g_* y^T M + M y g_*^T M] + \xi M g_* g_*^T M \}$$

where

$$(3-13a) \quad \epsilon \equiv g_*^T M y$$

$$(3-13b) \quad \tau \equiv y^T M y .$$

In order to apply condition (3-11b), we note that if  $u \neq 0$ , then  $\xi$  must vanish. Hence, we first evaluate  $E$  for  $\xi = 0$ , and test for whether

$$(3-14) \quad g_*^T E_0 g_* + \kappa > 0$$

(where  $E_0$  stands for  $E$  calculated with  $\xi = 0$ ). If this is so, then  $E_0$  is already satisfactory, and can be added to  $H$ .

TABLE 1  
Rosenbrock's function (strong search)

Step	DFP		Var. I		Var. II	
	f	NF	f	NF	f	NF
0	24.2	1	24.2	1	24.2	1
2	3.79	23	3.79	23	3.79	23
4	2.89	36	3.04	41	2.11	47
6	2.05	53	2.07	52	1.56	64
8	1.27	71	1.87	74	$9.9 \times 10^{-1}$	80
10	$6.4 \times 10^{-1}$	90	$6.4 \times 10^{-1}$	100	$6.9 \times 10^{-1}$	90
12	$3.7 \times 10^{-1}$	106	$2.9 \times 10^{-1}$	123	$4.2 \times 10^{-1}$	116
14	$2.0 \times 10^{-1}$	128	$1.4 \times 10^{-1}$	141	$1.9 \times 10^{-1}$	134
16	$8.2 \times 10^{-2}$	144	$5.3 \times 10^{-2}$	157	$3.7 \times 10^{-3}$	154
18	$3.6 \times 10^{-2}$	161	$1.5 \times 10^{-2}$	179	$6.4 \times 10^{-4}$	181
20	$3.4 \times 10^{-3}$	177	$1.4 \times 10^{-3}$	204	$1.7 \times 10^{-6}$	198
22	$1.9 \times 10^{-4}$	189	$2.2 \times 10^{-7}$	212	$6.6 \times 10^{-10}$	209
24	$4.8 \times 10^{-8}$	197	$3.2 \times 10^{-14}$	221		
26	(25) $3.2 \times 10^{-12}$	200				
T. B. U.		0		10		8

If, on the other hand, (3-14) does not hold, then  $u$  cannot differ from zero, but must be set equal to zero, and  $\xi$  cannot vanish. The result of substituting (3.12) into (3-8) (with  $u = 0$ ) is

$$(3-15) \quad \tau^{-1} \{ 2\eta\epsilon + \tau^{-1}(\rho - \xi\epsilon^2) - 2\xi\theta\epsilon^2 + \xi\theta^2 \} + \kappa = 0$$

where



$$(3-16) \quad \begin{aligned} \tau &\equiv y^T M y, \quad \eta \equiv g_*^T r, \\ \theta &\equiv g_*^T M g_*, \quad \rho \equiv y^T r. \end{aligned}$$

Solving (3-15) for  $\xi$  gives

$$(3-17) \quad \xi = \frac{\tau\kappa + 2\eta\epsilon + \tau^{-1}\rho\epsilon^2}{\tau^{-1}\epsilon^4 + 2\epsilon^2\theta - \theta^2}.$$

$M$  still must be selected, as previously; the most natural choices are as before. However, this analysis will not be carried further here, since no numerical tests have been made on these formulas.

TABLE 2

Rosenbrock's function (weak search)

Step	DFP		Var. I		Var. II	
	f	NF	f	NF	f	NF
0	2.42	1	2.42	1	2.42	1
4	2.33	16	2.33	16	1.90	17
8	1.90	30	1.45	35	1.50	30
12	1.39	49	$8.4 \times 10^{-1}$	54	$5.4 \times 10^{-1}$	41
16	1.10	62	$2.7 \times 10^{-1}$	75	$3.0 \times 10^{-1}$	56
20	$6.5 \times 10^{-1}$	78	$9.2 \times 10^{-3}$	90	$5.1 \times 10^{-2}$	74
24	$4.8 \times 10^{-1}$	97	$8.0 \times 10^{-5}$	105	$1.3 \times 10^{-2}$	91
28	$3.3 \times 10^{-1}$	113	$8.9 \times 10^{-7}$	115	$9.0 \times 10^{-8}$	112
32	$2.5 \times 10^{-1}$	130	$3.9 \times 10^{-9}$	136	(31)0.0	122
36	$1.0 \times 10^{-1}$	148	(33) $8.7 \times 10^{-13}$	138		
40	$1.2 \times 10^{-2}$	161				
44	$1.5 \times 10^{-3}$	175				
48	$4.3 \times 10^{-8}$	186				
56	(51)0.0	192				
T.B.U.		0		11		7

4. Numerical Experiments. A program was written which enabled a comparison to be made of the three  $H$ -corrections; viz.,  $E_D$ ,  $E_I$  and  $E_{II}$ . This program used the same line-search subroutine for all three methods, with the same set of stopping

thresholds, etc. Provision was made for printing out the step number  $k$ , the values of  $f_k, g_k, H_k$ , etc. at each step. Whenever, a nonnegative starting directional derivative was detected, a notation to this effect was printed out, and the sign of  $s_k$  was reversed. The problem was considered solved when the Euclidean norm of  $g_k$  fell below  $10^{-4}$ . An additional test was made on the magnitude of  $\sigma_k$ ; when this fell below  $10^{-6}$  before the minimum was reached, the method was considered to have failed. The reason for this is that when  $|df/dt|$  was too small, it was impossible to

TABLE 3

Powell's function (strong search)

Step	DFP		Var. I		Var. II	
	f	NF	f	NF	f	NF
0	$2.15 \times 10^{-2}$	1	$2.15 \times 10^{-2}$	1	$2.15 \times 10^{-2}$	1
4	$2.9 \times 10^{-2}$	46	$3.0 \times 10^{-2}$	46	2.2	49
8	$1.9 \times 10^{-3}$	69	$1.6 \times 10^{-3}$	73	$3.5 \times 10^{-3}$	78
12	$1.2 \times 10^{-6}$	102	$1.8 \times 10^{-5}$	103	$1.1 \times 10^{-3}$	101
16	$1.2 \times 10^{-9}$	123	$3.2 \times 10^{-8}$	138	$1.0 \times 10^{-3}$	126
20	(18) $4.9 \times 10^{-11}$	134	(19) $1.5 \times 10^{-10}$	163	$1.2 \times 10^{-3}$	151
24					$1.1 \times 10^{-4}$	178
28					$3.9 \times 10^{-6}$	204
32					$3.4 \times 10^{-7}$	226
36					$3.4 \times 10^{-7}$	257
40					$3.1 \times 10^{-7}$	279
44					$3.1 \times 10^{-7}$	304
48					$2.6 \times 10^{-7}$	341
52					$9.6 \times 10^{-8}$	362
56					$7.5 \times 10^{-8}$	391
60					$5.9 \times 10^{-8}$	431
64					$5.3 \times 10^{-8}$	457
68					$4.7 \times 10^{-8}$	498
72					(69) $4.5 \times 10^{-8}$	501
T. B. U.		0		8		33

TABLE 4

Powell's function (weak search)

Step	DFP		Var. I		Var. II	
	f	NF	f	NF	f	NF
0	$2.15 \times 10^2$	1	$2.15 \times 10^2$	1	$2.15 \times 10^2$	1
4	9.0	16	8.5	16	$1.6 \times 10$	18
8	$3.2 \times 10^{-2}$	29	$4.4 \times 10^{-2}$	28	$2.2 \times 10^{-1}$	26
12	$3.6 \times 10^{-4}$	43	$2.7 \times 10^{-2}$	49	$2.5 \times 10^{-2}$	41
16	$1.6 \times 10^{-4}$	61	$6.1 \times 10^{-7}$	63	$5.0 \times 10^{-4}$	58
20	$1.0 \times 10^{-4}$	75	$1.4 \times 10^{-7}$	74	$1.9 \times 10^{-5}$	72
24	$5.0 \times 10^{-5}$	91	$7.4 \times 10^{-8}$	96	$1.4 \times 10^{-5}$	88
28	$2.5 \times 10^{-6}$	111	$4.7 \times 10^{-10}$	113	$6.5 \times 10^{-6}$	102
32	$8.7 \times 10^{-8}$	125	(29) $2.1 \times 10^{-10}$	118	$5.6 \times 10^{-6}$	119
36	$2.6 \times 10^{-9}$	139			$4.4 \times 10^{-6}$	134
40	(37) $2.4 \times 10^{-10}$	144			$3.1 \times 10^{-6}$	153
44					$5.4 \times 10^{-7}$	174
48					$1.2 \times 10^{-7}$	191
52					$4.8 \times 10^{-8}$	206
56					(53) $9.8 \times 10^{-8}$	211
T. B. U.		0		9		22

obtain a detectable change in  $f$  in the  $s_k$ -direction in any reasonable step. This was the result, usually, of a poorly chosen direction  $s_k$  due to the lack of positive definiteness of  $H_k$ , which pointed up the desirability of this attribute.

There were two line searches used in the tests; in the first, which we shall call the "strong" one, the search was terminated when a certain quantity, estimated from current and past values of  $f$ , fell below  $10^{-2}$ . This quantity is the lowest-order dimensionless ratio associated with minimization, and it is given by:

$$(4-1) \quad \rho \equiv f'''f'/f''^2$$

where the primes denote directional derivatives of the various orders. This ratio is closely connected to the error made in estimating the minimum of a nonquadratic

function by an interpolated parabola, and does not depend on the scale of  $f$  or on that of the independent variable along  $s_k$ , so that this termination criterion is independent of whether the minimum is a sharp one or a flat one, and gave uniformly quite accurate minima.

TABLE 5  
Fletcher-Powell's function (strong search)

Step	DFP		Var. I		Var. II	
	f	NF	f	NF	f	NF
0	$2.5 \times 10^3$	1	$2.5 \times 10^3$	1	$2.5 \times 10^3$	1
2	$1.3 \times 10^2$	19	$1.29 \times 10^2$	19	$1.29 \times 10^2$	19
4	$2.7 \times 10$	40	$2.7 \times 10$	46	$5.6 \times 10$	37
6	$1.1 \times 10$	48	$1.1 \times 10$	67	$1.2 \times 10$	52
8	6.5	68	$1.0 \times 10$	79	2.8	64
10	2.2	84	7.6	99	1.7	85
12	$7.0 \times 10^{-1}$	100	3.1	114	$8.5 \times 10^{-1}$	98
14	$1.7 \times 10^{-1}$	114	1.9	131	$4.4 \times 10^{-1}$	110
16	$2.6 \times 10^{-2}$	131	$6.3 \times 10^{-1}$	144	$3.3 \times 10^{-1}$	124
18	$2.2 \times 10^{-4}$	139	$7.5 \times 10^{-2}$	167	$4.4 \times 10^{-2}$	138
20	$4.2 \times 10^{-10}$	145	$9.3 \times 10^{-3}$	179	$3.1 \times 10^{-2}$	161
22	(21) $2.1 \times 10^{-13}$	148	$7.7 \times 10^{-5}$	187	$1.2 \times 10^{-2}$	181
24			$2.6 \times 10^{-12}$	194	$4.5 \times 10^{-3}$	192
26					$2.8 \times 10^{-5}$	201
28					$9.4 \times 10^{-7}$	210
30					$2.6 \times 10^{-7}$	223
32					$3.4 \times 10^{-8}$	237
34					$1.1 \times 10^{-11}$	245
36					(35) $6.6 \times 10^{-15}$	252
T. B. U.		0		10		10

The second line search—the “weak” one, terminated as soon as a point along  $s_k$  was found at which  $f$  was smaller than the values at the points immediately to

its left and to its right; i.e., as soon as a point was "bracketed," it was taken as the solution point of the line search. The significance of this weakening is twofold: on the one hand, the successive directions  $\{s_k\}$  will usually not be conjugate, since this depends on finding a rather accurate minimum in the search, and this causes the Fletcher-Powell proof of positive definiteness to break down. On the other hand, many fewer time-consuming evaluations are required before the search is terminated. Dr. M. J. D. Powell, who suggested trying the "weak" search, was interested in the outcome of this competition, with regard to overall efficiency.

TABLE 6

Fletcher-Powell's function (weak search)

Step	DFP		Var. I		Var. II	
	f	NF	f	NF	f	NF
0	$2.5 \times 10^3$	1	$2.5 \times 10^3$	1	$2.5 \times 10^3$	1
2	$1.2 \times 10^3$	11	$1.2 \times 10^3$	11	$1.2 \times 10^3$	11
4	$3.2 \times 10$	18	$3.3 \times 10$	18	$9.8 \times 10^2$	18
6	$1.5 \times 10$	24	(5) $3.3 \times 10$	39	$2.1 \times 10^2$	22
8	$1.2 \times 10$	31	<u>failed</u>		$4.0 \times 10$	28
10	9.6	43			$2.7 \times 10$	37
12	8.4	55			$2.3 \times 10$	41
14	6.1	65			$2.2 \times 10$	47
16	4.5	89			$2.1 \times 10$	54
18	4.4	113			$2.0 \times 10$	74
20	4.4	126			$1.9 \times 10$	81
22	4.4	132			$1.8 \times 10$	91
24	4.4	143			$1.7 \times 10$	98
26	4.4	147			$1.7 \times 10$	108
28	4.4	152			$1.6 \times 10$	127
30	4.4	159			$1.2 \times 10$	137
32	<u>failed</u>				$1.0 \times 10$	143
34					$1.0 \times 10$	163
36					<u>failed</u>	
T B. U.		3		0		5

TABLE 7  
 Steps to Decrease  $f$  to Below Set Level  
 Box's 2D function (strong search)

Method	Level	Starting Point				
		I	II	III	IV	V
DFP	1	3	1	1	1	1
	10-1	6	1	26	1	1
	10-2	8	3	39	7	1
	10-4	10	6	42	10	1
	10-8	11	7	44	11	3
Total	f. evals.	69	66	443	91	34
Var. I	1	3	1	1	1	1
	10-1	6	1	failed	1	1
	10-2	7	3		4	1
	10-4	9	5		6	1
	10-8	10	7		7	3
Total	f. evals.	81	66		56	34
Var. II	1	3	1	1	1	1
	10-1	6	1	6	1	1
	10-2	7	3	8	4	1
	10-4	10	5	9	6	1
	10-8	12	7	11	7	3
Total	f. evals.	93	71	88	59	33

The tests were made on four "difficult" functions which have been used previously to test other minimization methods. These are:

(a) Rosenbrock's function (tested in [4])

$$(4-2) \quad f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

with a starting point of  $(-1.2, 1.0)$ .

(b) Powell's function (tested in [4])

$$(4-3) \quad f(x_1, x_2, x_3, x_4) = (x_1 - 10x_2)^2 + 5(x_3 - x_4)^2 \\ + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$$

with a starting point of  $(3, -1, 0, 1)$ .

(c) Fletcher-Powell function (tested in [4])

$$(4-4) \quad f(x_1, x_2, x_3) = 100\{(x_3 - 10\theta)^2 + (r - 1)^2\} + x_3^2$$

where  $r = (x_1^2 + x_2^2)^{1/2}$  and

$$\theta = \frac{1}{2\pi} \arctan\left(\frac{x_2}{x_1}\right), \quad \text{if } x_1 > 0$$

$$\theta = \frac{1}{2\pi} \arctan\left(\frac{x_2}{x_1}\right) + \frac{1}{2}, \quad \text{if } x_1 < 0$$

with a starting point of  $(-1, 0, 0)$ .

(d) Box's function [1]

$$(4-5) \quad f(x_1, x_2, x_3) = \sum_{\mu} \{[e^{\mu x_1} - e^{\mu x_2}] - x_3[e^{-\mu} - e^{-10\mu}]\}^2$$

where the summation is over the values of  $\mu = .1, .2, \dots, 1$ . There are two cases considered:

*Case 1.*  $x_3$  is fixed at the value 1, and the minimum with respect to  $x_1$  and  $x_2$  is sought.

*Case 2.* The minimum in terms of  $x_1, x_2$  and  $x_3$  is sought.

The starting points were those chosen by Box, and were

*Case 1.*

$$(x_1, x_2) = (0, 0); (0, 20); (5, 0);$$

$$(5, 20); (2.5, 10) .$$

*Case 2.*

$$(x_1, x_2, x_3) = (0, 20, 1); (2.5, 10, 10);$$

$$(0, 0, 10); (0, 10, 1);$$

$$(0, 10, 10); (0, 10, 20);$$

$$(0, 20, 0); (0, 20, 10);$$

$$(0, 20, 20) .$$

The results of the trials are shown in Tables 1-10, which are largely self-explanatory. They are modelled after the tables published by Fletcher and Powell, and by Box. A few explanations with regard to certain markings are, however, in order.

(a) Whenever a sequence terminates at a step whose number is not in the table, the actual terminal step number is placed in parentheses just preceding the  $f$ -value reached.

(b) The NF column contains the number times the function  $f$  was evaluated up to the completion of the step in question. The gradient at this point would have been evaluated a number of times equal to the step number.

(c) The entry "T.B.U." refers to the total number of "back-ups" due to  $df/dt$  being positive.

(d) In Tables 7 through 10, the starting point numbers refer to the lists given above for the Box cases.

(e) In Tables 9 and 10, the asterisk indicates that the "wrong" solution was found (see Box [1]). This does not mean that the method failed, but that the starting point was unfortunate.

TABLE 8  
Steps to Decrease  $f$  to Below Set Level  
Box's 2D function (weak search)

Method	Level	Starting Point				
		I	II	III	IV	V
DFP	1	2	2	1	1	1
	10-1	7	4	30	3	1
	10-2	8	77	66	36	2
	10-4	11	137	95	39	4
	10-8	12	142	100	42	30
<b>Total</b>	<b>f. evals.</b>	<b>45</b>	<b>628</b>	<b>441</b>	<b>204</b>	<b>114</b>
Var. I	1	2	2	1	1	1
	10-1	7	4	failed	3	1
	10-2	10	11		9	2
	10-4	12	14		12	4
	10-8	14	16		15	9
<b>Total</b>	<b>f. evals.</b>	<b>54</b>	<b>60</b>		<b>67</b>	<b>38</b>
Var. II	1	2	2	1	1	1
	10-1	4	4	11	3	1
	10-2	7	9	14	9	2
	10-4	10	13	18	11	4
	10-8	16	17	20	16	10
<b>Total</b>	<b>f. evals.</b>	<b>74</b>	<b>69</b>	<b>97</b>	<b>78</b>	<b>39</b>

**5. Conclusions.** There are no very clear-cut implications in the results of the numerical experiments. In certain cases, the DFP method is best, and in others the worst. "Var I" seems to be competitive with the DFP methods, but "Var II" is almost always worse than the others. It is also clear that sometimes the weak search is more economical and sometimes not. It is certainly not as dependable as the strong search.



TABLE 9  
Steps to Decrease f Below Set Level  
Box's 3D function (strong search)

Method	Level	Starting Point									
		I	II	III	IV	V	VI	VII	VIII	IX	
DFP	1	1	1	2	1	1	5	1	2	5	
	10-1	1	1	2	1	1	7	1	4	7	
	10-2	6	2	2	3	4	9	1	14	15	
	10-4	11	4	7	4	11	16	1	23	18	
	10-8	14	9	10	6(10 <sup>-6</sup> )	13	19	1	25	20	
Total	f. evals.	114	60	86*	30	110	127	8*	228	141	
Var. I	1	1	1	2	1	1	5	1	2	5	
	10-1	1	1	2	1	1	7	1	4	7	
	10-2	6	2	2	3	4	9	1	9	12	
	10-4	9	4	5	4	8	13	1	10	15	
	10-8	10	8	10	7(10 <sup>-6</sup> )	11	16	1	13	17	
Total	f. evals.	64	62	92*	45	88	99	8*	108	104	
Var. II	1	1	1	2	1	1	5	1	2	5	
	10-1	1	1	2	1	1	7	1	4	7	
	10-2	31	2	2	3	4	9	1	13	13	
	10-4	failed	3	5	4	7	25	1	17	19	
	10-8		15	8	12(10 <sup>-6</sup> )	45(10 <sup>-7</sup> )	44	1	30	26	
Total	f. evals.		102	77*	83	445	314	8*	237	186	

TABLE 10  
Steps to Decrease f to Below Set Level  
Box's 3D function (weak search)

Method	Level	Starting Point								
		I	II	III	IV	V	VI	VII	VIII	IX
DFP	1	2	60	9	1	11	5	1	9	5
	10-1	4	failed	9	2	11	5	1	10	5
	10-2	41		10	4	45	7	1	45	94
	10-4	79		11	6	68	14	1	82	179
	10-8	85		13	9(10 <sup>-6</sup> )	72	71	1	89	187
Total	f. evals.	375		36*	25	309	322	4*	395	866
Var. I	1	2	6	8	1	8	11	1	6	10
	10-1	4	8	10	2	11	13	1	9	11
	10-2	11	11	10	4	21	21	1	19	20
	10-4	14	12	11	5	25	23	1	23	24
	10-8	17	24	12	8(10 <sup>-6</sup> )	32	28	1	29	31
Total	f. evals.	66	90	35*	26	128	109	4*	129	137
Var. II	1	2	5	13	1	10	7	1	8	6
	10-1	4	6	14	2	15	9	1	11	11
	10-2	37	8	15	4	22	14	1	23	15
	10-4	51	25	16	8	33	24	1	32	22
	10-8	59	28	20	12(10 <sup>-6</sup> )	39	27(10 <sup>-8</sup> )	1	failed	33
Total	f. evals.	235	122*	62*	37	143	119	4*		145

One fact is clear: it is possible to derive efficient DFP-like correction formulas by variational methods. In fact, by suitably choosing  $M$ , it is possible to obtain the DFP formula by variational means. This has been done by Dr. D. Goldfarb, who has also derived other important formulas in this manner. (See his paper [11] which follows the present one.)

This sort of thing suggests that it might be possible to derive other types of correction formulas on variational grounds. The writer has in fact derived a correction to the gradient, based on  $f$  evaluations alone (and an assumed  $H$ ), but it has so far not been tested numerically.

**6. Acknowledgement.** I wish to thank Dr. R. T. Mertz, whose remark, during one of our discussions, about the advisability of looking for the "best"  $H$ -correction, consistent with the DFP condition, started me on the variational path. I am also grateful to Dr. D. Goldfarb, who pointed out various errors in the manuscript and a substantial mathematical error in my attempt to derive Davidon's formula variationally.

**APPENDIX**

**Proof that  $H_k \rightarrow G^{-1}$  for the  $E_I$  Correction**

**By Yonathan Bard**

**THEOREM.** *Let  $f(x) = a + g^T x + \frac{1}{2} x^T G x$  be a quadratic function of the  $N$ -dimensional vector  $x$ ,  $H_0$  any nonsingular symmetric matrix, and  $x_0$  an arbitrary point. Let the following quantities be defined iteratively, for  $i = 0, 1, \dots$*

(A1)  $g_i \equiv \nabla f(x_i),$

(A2)  $\sigma_i \equiv -\alpha_i H_i g_i,$

with  $\alpha_i$  chosen so as to make  $\phi_i(\alpha) = f(x_i + \alpha s_i)$  stationary (line search).

(A3)  $x_{i+1} = x_i + \sigma_i,$

(A4)  $y_i \equiv g_{i+1} - g_i,$

(A5)  $\tau_i \equiv y_i^T H_i y_i,$

(A6)  $A_i \equiv \sigma_i y_i^T H_i,$

(A7)  $B_i \equiv H_i y_i y_i^T H_i,$

(A8)  $H_{i+1} = H_i + \frac{1}{\tau_i} \left\{ A_i + A_i^T - B_i - \left( \frac{y_i^T \sigma_i}{\tau_i} \right) B_i \right\}.$

This is equivalent to Eq. (2-27a) (with  $\tau_i \equiv y_i^T H_i y_i$ ).

Then, if either:

- (a)  $G$  is nonsingular and the  $\sigma_i$  ( $i = 0, 1, \dots, N - 1$ ) are linearly independent, or
- (b)  $G$  is positive definite and the  $\sigma_i$  ( $i = 0, 1, \dots, N - 1$ ) are nonzero, we have

(A9)  $H_N = G^{-1}$

and

$$(A10) \quad g_N = 0, \text{ i.e., } x_N = -G^{-1}g$$

(the stationary point of  $f(x)$ ).

*Proof.* The proof follows exactly the argument presented by Fletcher and Powell for Davidon's method (except for a minor error in Fletcher and Powell's induction argument; their Eq. (10) makes no sense for  $k = 1$ ).

By definition

$$(A11) \quad g_i = g + Gx_i.$$

Hence,  $y_i \equiv g_{i+1} - g_i = G(x_{i+1} - x_i)$ , i.e.,

$$(A12) \quad y_i = G\sigma_i.$$

We have

$$H_{i+1}G\sigma_i = H_{i+1}y_i = H_i y_i + (1/\tau_i) (A_i y_i + A_i^T y_i - B_i y_i - (y_i^T \sigma_i / \tau_i) B_i y_i).$$

But  $A_i y_i = \tau_i \sigma_i$ ;  $A_i^T y_i = y_i^T \sigma_i H_i y_i$ ;  $B_i y_i = H_i y_i$ . Hence, after cancellations:

$$(A13) \quad H_{i+1}G\sigma_i = \sigma_i.$$

From the choice of  $\alpha_i$ , it is clear that

$$(A14) \quad g_{i+1}^T \sigma_i = 0.$$

Since  $\sigma_{i+1}^T G\sigma_i = -\alpha_{i+1} g_{i+1}^T H_{i+1} G\sigma_i = -\alpha_{i+1} g_{i+1}^T \sigma_i$  (from (A13)), it follows that

$$(A15) \quad \sigma_{i+1}^T G\sigma_i = 0.$$

With  $i = 0$ , Eqs. (A13) and (A15) demonstrate that for  $k = 1$ ,

$$(A16) \quad H_k G\sigma_i = \sigma_i, \quad (0 \leq i < k)$$

$$(A17) \quad \sigma_j^T G\sigma_i = 0, \quad (0 \leq i < j \leq k).$$

Let us assume that (A16) and (A17) are true for some value  $k$ . We shall prove that they must then be true for  $k + 1$ . Using (A16) and then (A12), we have

$$y_k^T H_k G\sigma_i = y_k^T \sigma_i = \sigma_k^T G\sigma_i, \quad (0 \leq i < k).$$

Thus, from (A17),

$$(A18) \quad y_k^T H_k G\sigma_i = 0.$$

It follows that

$$A_k G\sigma_i = \sigma_k y_k^T H_k G\sigma_i = 0 \quad (\text{from (A18)}),$$

$$A_k^T G\sigma_i = H_k y_k \sigma_k^T G\sigma_i = 0 \quad (\text{from (A17)}),$$

$$B_k G\sigma_i = H_k y_k y_k^T H_k G\sigma_i = 0 \quad (\text{from (A18)})$$

and

$$(A19) \quad H_{k+1} G\sigma_i = H_k G\sigma_i + \frac{1}{\tau_k} \left\{ A_k G\sigma_i + A_k^T G\sigma_i - B_k G\sigma_i - \frac{y_k^T \sigma_k}{\tau_k} B_k G\sigma_i \right\} = H_k G\sigma_i = \sigma_i$$

(from (A16), and (A13), (with  $i = k$  in the latter), we have

$$(A20) \quad H_{k+1}G\sigma_i = \sigma_i, \quad (0 \leq i < k + 1).$$

Now, for any  $0 \leq i < k$ ,

$$g_{k+1} = g_{i+1} + G(x_{k+1} - x_{i+1}) = g_{i+1} + \sum_{j=i+1}^k G\sigma_j$$

whence

$$(A21) \quad g_{k+1}^T\sigma_i = g_{i+1}^T\sigma_i + \sum_{j=i+1}^k \sigma_j^T G\sigma_i = 0.$$

Therefore, substituting (A20) in (A21),

$$(A22) \quad g_{k+1}^T H_{k+1} G\sigma_i = 0.$$

But, from (2),  $g_{k+1}^T H_{k+1} = -(1/\alpha_{k+1})\sigma_{k+1}^T$ , and (A22) becomes

$$(A23) \quad \sigma_{k+1}^T G\sigma_i = 0, \quad (0 \leq i < k)$$

(assuming  $\alpha_{k+1} \neq 0$ ). Again, combining (A17), (A23) and (A15) (with  $i = k$ ), we have

$$(A24) \quad \sigma_j^T G\sigma_i = 0, \quad (0 \leq i < j \leq k + 1).$$

Equations (A20) and (A24) are equivalent to (A16) and (A17), respectively, with  $k$  replaced by  $k + 1$ . Thus, (A16) and (A17) are proven by induction for all  $k$ .

Consider the matrix  $C_N \equiv H_N G$ . According to (A16),  $\sigma_i$  ( $i = 0, 1, \dots, N - 1$ ) are all eigenvectors of  $C_N$  with eigenvalues 1. These vectors are linearly independent, either from assumption (a) or from assumption (b) combined with Eq. (17), for were, say,  $\sigma_m = \sum_{i \neq m} \mu_i \sigma_i$ , then, with  $j \neq m$ ,

$$\sigma_m^T G\sigma_j = \sum_{i \neq m} \mu_i \sigma_i^T G\sigma_j = \mu_j \sigma_j^T G\sigma_i = 0$$

from (A17), but this is impossible if  $\sigma_j \neq 0$  and  $G$  is definite.

Thus, there exists a nonsingular  $N \times N$  matrix  $\Delta$  (whose  $i$ th column is  $\sigma_{i+1}$ ) such that  $H_N G \Delta = \Delta$ . Postmultiplying by  $\Delta^{-1} G^{-1}$ , we have  $H_N = G^{-1}$  as was to be proven. Also, from (A21),  $g_N$  must be orthogonal to the  $N - 1$  independent vectors  $\sigma_0, \sigma_1, \dots, \sigma_{N-2}$ , and from (A14),  $g_N$  is also orthogonal to  $\sigma_{N-1}$ . Thus,  $g_N = 0$ , and  $x_N$  is the stationary point of  $f(x)$ .

IBM

New York Scientific Center  
New York, New York 10021

1. M. J. BOX, "A comparison of several current optimization methods, and the use of transformations in constrained problems," *Comput. J.*, v. 9, 1966, pp. 67-77. MR 33 #870.

2. J. B. CROCKETT & H. CHERNOFF, "Gradient methods of maximization," *Pacific J. Math.*, v. 5, 1955, pp. 33-50. MR 17, 790.

3. W. C. DAVIDON, *Variable Metric Method for Minimization*, AEC Res. and Develop. Report, ANL-5990 (Rev.), 1959.

4. R. FLETCHER & M. J. D. POWELL, "A rapidly convergent descent method for minimization," *Comput. J.*, v. 6, 1963/64, pp. 163-168. MR 27 #2096.

5. J. GREENSTADT, "On the relative efficiencies of gradient methods," *Math. Comp.*, v. 21, 1967, pp. 360-367. MR 36 #6122.
6. J. GREENSTADT, *Variations on Variable-Metric Methods*, IBM NY Scientific Center Report 320-2901, June, 1967.
7. B. KLEIN, "Direct use of extremal principles in solving certain optimizing problems involving inequalities," *J. Operations Res. Soc. Amer.*, v. 3, 1955, pp. 168-175. MR 16, 937.
8. W. C. DAVIDON, "Variance algorithm for minimization," *Comput. J.*, v. 10, 1968, pp. 406-410. MR 36 #4790.
9. C. G. BROYDEN, "Quasi-Newton methods and their application to function minimisation," *Math. Comp.*, v. 21, 1967, pp. 368-381. MR 36 #7317.
10. P. WOLFE, *Another Variable-Metric Method*, Working Paper, 1967.
11. D. GOLDFARB, "A family of variable-metric methods derived by variational means," *Math. Comp.*, v. 24, 1970, pp. 23-26.