A Family of Variable-Metric Methods Derived by Variational Means

By Donald Goldfarb

Abstract. A new rank-two variable-metric method is derived using Greenstadt's variational approach [Math. Comp., this issue]. Like the Davidon-Fletcher-Powell (DFP) variable-metric method, the new method preserves the positive-definiteness of the approximating matrix. Together with Greenstadt's method, the new method gives rise to a one-parameter family of variable-metric methods that includes the DFP and rank-one methods as special cases. It is equivalent to Broyden's one-parameter family [Math. Comp., v. 21, 1967, pp. 368-381]. Choices for the inverse of the weighting matrix in the variational approach are given that lead to the derivation of the DFP and rank-one methods directly.

In the preceding paper [6], Greenstadt derives two variable-metric methods, using a classical variational approach. Specifically, two iterative formulas are developed for updating the matrix $H_k$, (i.e., the inverse of the variable metric), where $H_k$ is an approximation to the inverse Hessian $G^{-1}(x_k)$ of the function being minimized.*

Using the iteration formula

$$H_{k+1} = H_k + E_k$$

to provide revised estimates to the inverse Hessian at each step, Greenstadt solves for the correction term $E_k$ that minimizes the norm

$$N(E_k) = \text{Tr} (WE_kW^T)$$

subject to the conditions

(1) $E_k^T = E_k$

and

(2) $E_k y_k = \sigma_k - H_k y_k$.

$W$ is a positive-definite symmetric matrix and Tr denotes the trace.

The first condition is a symmetry condition which ensures that all iterates $H_k$ will be symmetric as long as the initial estimate $H_0$ is chosen to be symmetric. The second condition ensures that the updated matrix $H_{k+1}$ satisfies the equation

$$H_{k+1} y_k = \sigma_k$$

and hence, that the method is of the "quasi-Newton" type [1].

Received June 30, 1969, revised August 4, 1969.

AMS Subject Classifications. Primary 30, Secondary 10.

Key Words and Phrases. Unconstrained optimization, variable-metric, variational methods, Davidon method, rank-one formulas.

* The reader is referred to Greenstadt's paper [6] for a more detailed discussion of variable-metric methods and for definitions of some of the terms used here.
If the function being minimized were quadratic, $H_{k+1}$ would operate on the vector $y_k$ as would the matrix $G^{-1}$. The norm chosen by Greenstadt is essentially a weighted Euclidean norm.

Solving this constrained minimization problem using Lagrange multipliers, Greenstadt obtained the following formula for $E_k$:

$$E_k = \frac{1}{(y^T M y)} \left\{ \sigma y^T M + M y \sigma^T - H y y^T M - M y y^T H - \frac{1}{(y^T M y)} [(y^T \sigma) - (y^T H y)] M y y^T M \right\},$$

(3)

where $M = W^{-1}$.

If the current approximation $H$ to $G^{-1}$ is substituted for $M$, Greenstadt’s first formula is obtained:

$$E_H = \frac{1}{(y^T H y)} \left\{ \sigma y^T H + H y \sigma^T - \left[ 1 + \left( \frac{y^T \sigma}{y^T H y} \right) \right] H y y^T H \right\}.$$

(Throughout the remainder of the note no superscript will indicate the $k$th iterate and a (*) superscript will denote the $(k + 1)$st iterate.)

If, instead, $H^*$ is substituted for $M$ in Eq. (3),

$$E_H^* = \frac{1}{(y^T \sigma)} \left\{ -\sigma y^T H - H y \sigma^T + \left[ 1 + \left( \frac{y^T \sigma}{y^T H y} \right) \right] \sigma \sigma^T \right\},$$

is obtained. The above two correction terms appear to be similar, at least in part, to both the Davidon-Fletcher-Powell (or DFP) rank-2 correction term

$$E_{R2} = \frac{\sigma \sigma^T}{\sigma^T y} - \frac{H y y^T H}{y^T H y},$$

and the rank-1 correction term [1], [3], and [7]

$$E_{R1} = \frac{(\sigma - H y)(\sigma - H y)^T}{(\sigma - H y)^T y}.$$

In fact, all four corrections terms $E_H, E_{H^*}, E_{R1},$ and $E_{R2}$ give rise to algorithms that locate the exact minimum of a strictly convex quadratic objective function of $N$ variables in $N$ steps. They also result in a matrix $H$ which after those $N$ steps is exactly equal to $G^{-1}$. Proofs of this property, which we shall refer to as “exactness” following Broyden [1], were given for $E_{R2}, E_{R1},$ and $E_H$ by Fletcher and Powell [4], Broyden [1], and Bard [6, Appendix], respectively.

It is easy to show that this property also holds for variable-metric algorithms with correction term $E_{H^*}$. For example, Bard’s proof may be followed almost entirely, except for some obvious and trivial changes.

$E_{R2}$ and $E_{H^*}$, moreover, share the additional property of preserving the positive-definiteness of the approximating matrix $H$. This ensures the stability of the corresponding variable-metric algorithms that search for a minimum along the direction $-H g$ at each step. Fletcher and Powell proved this for $E_{R2}$. The proof for $E_{H^*}$ follows from the observation that
A FAMILY OF VARIABLE-METRIC METHODS

\[ x^T (E_{H^*} - E_{R2})x = x^T E_{H^*}x - x^T E_{R2}x = \frac{[y^T H y (x^T \sigma) - (y^T \sigma) (x^T H y)]^2}{(y^T \sigma)^2 (y^T H y)} \geq 0. \]

It may seem then that the iteration scheme \( H^* = H + E_{H^*} \) would be less likely to generate a sequence of matrices \( \{H_i\} \) that tends toward singularity than would the DFP iteration scheme \( H^* = H + E_{R2} \). One should not count this apparent improvement too heavily, for the behavior of a variable-metric algorithm and its convergence to a stationary point depend upon the sequence \( \{H_i\} \) being bounded above as well as being bounded away from singularity [5].

The resemblances between the correction terms \( E_{R2} \), \( E_{R1} \), \( E_{H} \) and \( E_{H^*} \) suggest that each can be written as a linear combination of the others. This is indeed the case: \( E_{R2} \) and \( E_{R1} \) can be expressed directly as weighted sums of \( E_{H} \) and \( E_{H^*} \), and vice versa.

\[ E_{R2} = \frac{(y^T H y) E_H + (y^T \sigma) E_{H^*}}{y^T H y + y^T \sigma} \]
\[ E_{R1} = \frac{(y^T H y)^2 E_H - (y^T \sigma)^2 E_{H^*}}{(y^T H y)^2 - (y^T \sigma)^2} = \frac{(y^T H y)^2 E_H - (y^T H^* y)^2 E_{H^*}}{(y^T H y)^2 - (y^T H^* y)^2}, \]

where

\[ \gamma = \left( \frac{y^T \sigma}{y^T H y} \right). \]

It is especially interesting that the two variationally derived correction terms \( E_{H} \) and \( E_{H^*} \) give rise to a one-parameter family of correction terms \( E = \alpha E_{H} + (1 - \alpha) E_{H^*} \) whose corresponding variable-metric methods are "exact." The DFP-rank-2 and rank-1 correction terms are members of this one-parameter family that correspond to particularly interesting choices for the parameter \( \alpha \). This family includes all symmetric variable-metric correction terms that have been published [1], [2], [3], [4], [6], [7].

In fact, it is equivalent to the one-parameter family given by Broyden's algorithm 2 [1]. The equivalence can be obtained by setting

\[ \alpha = \frac{(1 - \beta y^T \sigma) y^T H y}{y^T H y + y^T \sigma}, \]

where \( \beta \) is Broyden's parameter.

Broyden's algorithm 1 (i.e., the rank-1 algorithm) is just a special case of his algorithm 2 [1], with \( \beta = 1/(y^T H y - y^T \sigma) \); a point that seems to have been overlooked by Broyden himself.

It is also possible to obtain \( E_{R1} \) and \( E_{R2} \) directly from Eq. (3) by choice of a suitable \( M \). For the rank-1 case a choice that works is

\[ M_{R1} = H^* - H = E. \]

However, using \( M_{R1} = M \) in Eq. (3) yields \( E = E_{R1} \) which has rank 1 and, hence, \( M_{R1} \) has no inverse.

** Davidson's variance algorithm [3] multiplies the rank-1 correction term \( E_{R1} \) by a scalar function of \( (g^T H g^*/g^* T H g^*) \) so as to ensure the stability of the method.
Before going further, we note that:

(i) Formula (3) is homogeneous in $M$; therefore, replacing $M$ by $\mu M$, where $\mu$ is a scalar, has no affect on the resultant $E$.

(ii) $M$ always appears in conjunction with $y$ in formula (3) either as $My$ or as $y^T M$; therefore, the replacement of $(y^T H y) H$ by $H y y^T H$ and $(y^T \sigma) H^* = (y^T H^* y) H^*$ by $H^* y y^T H^*$ as terms of $M$ has no affect on the resultant $E$.

Hence the substitution of either

\begin{equation}
M_{R1} = H^* - \frac{H y y^T H}{y^T H y} 
\end{equation}

or

\begin{equation}
M_{R1} = H - \frac{\sigma \sigma^T}{\sigma^T y} 
\end{equation}

for $M$ in Eq. (3) also yields $E_{R1}$.

Substitution of any of the forms of $M_{R2}$ given below in Eq. (3) is sufficient to show that all give rise to the DFP correction term $E_{R2}$.

\begin{equation}
M_{R2} = (y^T H y)^{1/2} H^* - (y^T \sigma)^{1/2} H ,
\end{equation}

\begin{equation}
M_{R2} = (y^T H^* y)^{-1/2} H^* - (y^T H y)^{-1/2} H ,
\end{equation}

\begin{equation}
M_{R2} = H^* - \left( \frac{y^T \sigma}{y^T H y} \right)^{1/2} H y y^T H y^T H y ,
\end{equation}

\begin{equation}
M_{R2} = H - \left( \frac{y^T H y}{y^T \sigma} \right)^{1/2} \frac{\sigma \sigma^T}{\sigma^T y} .
\end{equation}

Although the matrices $M_{R1}$ and $M_{R2}$ given by expressions (7) through (9) are, in general, nonsingular, these choices for $M$ and hence, the corresponding $W$'s are not necessarily positive-definite. Thus, their substitution in Eq. (3) is somewhat contrived. Just what role they play in the variational derivation of the rank-1 and DFP rank-2 methods remains confusing.

City College of the City University of New York
New York, New York 10031