CDC
PAPER No. TP2-4

QUASI-NEWTON ALGORITHMS: APPROACHES AND MOTIVATIONS

S. S. Oren

Xerox Palo Alto Research Center and
Stanford University

ABSTRACT

This paper surveys some of the unifying approaches used to derive formulae for updating the inverse Hessian approximations in quasi-Newton algorithms and presents a new approach of this kind based on geometric considerations. The paper discusses the intuitive motivations for these approaches and their potential in providing explanations for observed behavior of such algorithms.

1. INTRODUCTION

The quasi-Newton algorithms, known also as Variable Metric Methods, are considered to be the most sophisicated algorithms for solving the unconstrained minimization problem

$$\min f(x) \quad \text{where} \quad x \in E^n \quad \text{and} \quad f \in C^2 .$$

These methods, which assume the availability of the gradient $g(x)$ for any given x, are based on the recursion

$$x_{k+1} = x_k - \alpha_k D_k g_k . \qquad (1)$$

In this recursion, an analog to the one used in the Newton Raphson method, α_k, is a positive step size parameter selected to satisfy certain descent conditions, while D_k is an $n \times n$ matrix approximating the inverse Hessian $[\nabla^2 f(x)]^{-1}$. The basic principle in these algorithms is to obtain some of the advantages of Newton's method while using only first order information about the function. Thus, the approximations D_k are inferred from the gradients at previous iterations and updated as new gradients become available. The updating is done such that $D_{k+1} q_k = p_k$ where $q_k = g_{k+1} - g_k$ and $p_k = x_{k+1} - x_k$. This condition, which is often referred to as the "quasi-Newton condition," is motivated by the fact that if the function was quadratic, then $[\nabla^2 f(x)]^{-1} q_k = p_k$.

The first algorithm of this type was invented by Davidon [1] and further developed and simplified by Fletcher and Powell [2]. In this method, referred to as the DFP algorithm, D_k is updated by the rank two formula

$$D_k = D_k - D_k q_k q_k' D_k / q_k' D_k q_k + p_k p_k' / p_k' q_k , \qquad (2)$$

and α_k is selected such as to minimize $f(x_k - \alpha D_k g_k)$. When applied to a positive definite quadratic function with Hessian H, the directions p_k generated by the above algorithm are H conjugate implying n-step convergence. Furthermore, the nth approximation D_n equals H. Since the pioneering works by Davidon and by Fletcher and Powell, the field of quasi-Newton algorithms has been a very active research area and the subject of a vast number of publications. Many of these contributions proposed alternative updating formulae, and some of them ([3], [4], [5], [6]) introduced unifying approaches that led to general classes of such formulae. These approaches often offer interesting interpretations to some of the well known formulae that are special cases of the more general classes, and might prove valuable in explaining some of the puzzling behavior of these formulae.

It is the purpose of this paper to survey some of the existing approaches mentioned above and introduce a new approach that leads to some commonly used updating formulae. This derivation offers geometric interpretations to the role of the various terms in these formulae.

2. APPROACHES BASED ON GENERATING CONJUGATE DIRECTIONS

The most common approaches used to derive updating formulae for variable metric algorithms are based on viewing these formulae as means of generating conjugate direction. Thus, the objective in these approaches is to construct general classes of such formulae that will produce, in a quadratic case, conjugate directions of search. In constructing such classes, it is always assumed that the objective function is a positive definite quadratic of the form

$$f(x) = \frac{1}{2} x'Hx + bx + c \qquad (3)$$

and that the step size α_k in (1) is "perfect," i.e., it minimizes $f(x_k - \alpha D_k g_k)$.

Using the notation introduced earlier, the directions of search are defined by $p_k = -\alpha_k D_k g_k$. Suppose the vectors $p_0, \ldots, p_{n-1}$ form a set of mutually conjugate vectors (with respect to H), and the points $x_1, \ldots, x_n$ are obtained by applying (1) with a perfect α_k to (3). Then $p_i' H p_j = 0$ for $i \neq j$ and $g_{i+1}' p_i = 0$ for all i. Consequently,

$$g_i' p_j = (Hx_i + b)' p_j = (g_{j+1} + \sum_{k=j+1}^{i-1} \alpha_k H p_k)' p_j = 0$$
$$\text{for } j \leq i-1. \qquad (4)$$

From the conjugacy condition, we have

$$p_i' H p_j = -\alpha_i g_i' D_i H p_j = 0 \quad \text{for} \quad i \neq j \,. \qquad (5)$$

Satisfying (4) and (5) is a necessary and sufficient condition for the p's to be conjugated. Thus, the desired updating formulae have to generate matrices D_k that satisfy

$$g_k' D_k H p_j = 0 \quad \text{and} \quad g_k' p_j = 0 \quad \text{for} \quad j < k \qquad (6)$$

where $$p_j = -\alpha_j D_j g_j \,.$$

Broyden [3] was the first to investigate such classes of formulae, but he added the quasi-Newton condition $D_k q_k = p_k$ to his requirement. He considered updatings of the form

$$D_{k+1} = D_k + p_k w_k' - D_k q_k z_k' \qquad (7)$$

where w_k and z_k are vectors chosen such as to satisfy (6) and the quasi-Newton condition. This led to a one parameter family of updating formulae.

A more general class that contains Broyden's family was developed by Huang [4]. To satisfy condition (6), Huang required that

$$D_k H p_j = D_k q_j = \rho_j p_j \quad \text{for} \quad j \leq k-1 \qquad (8)$$

and considered updates of the general form,

$$D_{k+1} = D_k + a_k p_k p_k' + b_k D_k q_k q_k' D_k' + c_k p_k q_k' D_k' + d_k D_k q_k p_k' \qquad (9)$$

where ρ_j, a_k, b_k, c_k, d_k are arbitrary scalars. Imposing condition (8) on (9) and requiring that D_k be symmetric leads to a two parameter family of formulae that can be written in the form:

$$D_{k+1} = D_k - D_k q_k q_k' D_k / q_k' D_k q_k + \varphi_k v_k v_k' + \rho_k p_k p_k' / p_k' q_k \qquad (10)$$

where $v_k = p_k / p_k' q_k - D_k q_k / q_k' D_k q_k$ while φ_k and ρ_k are arbitrary scalars. If $\rho_k = 1$, for all k, (10) reduces to Broyden's class mentioned above, and if in addition $\varphi_k = 0$, it becomes the DFP formula. One should point out that though (10) does not satisfy the quasi-Newton condition unless $\rho_k = 1$, it can be modified to satisfy that condition by multiplying the right hand side by $1/\rho_k$ (assuming $\rho_k \neq 0$). Such a modification, which clearly would not affect the directions of search, may be useful when these formulae are used with a predetermined $\alpha_k = 1$ for all k. (See Oren [7]).

Many of the updating formulae proposed in the past were special cases of (10) or its modification mentioned above. Most of these formulae use $\rho_k = 1$ for all k, since this guarantees $D_n = H$. Some of the more recent contributions take advantage of the freedom in choosing φ_k to impose additional requirements such as positive definiteness of D_k and low condition number of D_k (see Spedicato [8] and Shanno [9]). It was shown, however, by Dixon [10] that if ρ_k is fixed, and α_k perfect then the points generated by Huang's algorithm are independent of φ_k even for a nonquadratic function.

Algorithms of Huang's class that use variable ρ_k have been considered only recently and seem to perform consistently better than the ones using fixed values of ρ_k. One algorithm of this type was given by Biggs [11], who adjusts ρ_k to account for nonquadratic terms in the objective function. A different criterion was proposed by Oren and Luenberger [12], who suggested to select ρ_k such as to ensure monotonic decrease in the condition number of $H^{1/2} D_k H^{1/2}$. This approach led to the Self Scaling Variable Metric Algorithm given by Oren, [13], [14].

In spite of its generality, (9) is not the only possible form of updating formulae that will satisfy condition (8). This fact was noticed by Adachi [5], who introduced three more general families of formulae that satisfy (8) with $\rho_j = 1$. Though the restriction on ρ_j is not essential to Adachi's development, it was introduced by the author to ensure $D_n = H$. Adachi has shown that all of the known updating formulae (not including the ones with variable ρ_k) may be derived as special cases of his families. In a later paper, Adachi [15] also extended Dixon's [10] result and derived conditions under which algorithms using different members of his three general classes of formulae generate the same sequence of points in a nonquadratic case.

Though the approaches discussed above are very elegant and unify much of the theoretical work done in this area, they have limited value in terms of explaining observed behavior of the various updating formulae and suggesting ways to improve them. Theories that view variable metric algorithms as special kinds of conjugate direction methods will never be able to explain the observed fact that variable metric methods are superior to regular conjugate direction algorithms such as Fletcher and Reeves [16]. Such observations suggest that perhaps the conjugacy property and the n-step convergence are not the most important properties of variable metric methods. This view is concurred by Fletcher's [17] and Oren's [7] results which indicate that using predetermined step sizes α_k (rather than perfect ones) does not radically increase the number of iterations to convergence and actually reduce the total number of gradient and function evaluations. Such modifications destroy the conjugacy of the search directions even in a quadratic case and with only one exception (the "rank one update") relinquish the n-step convergence feature. It is clear, therefore, that theories hinging on such properties are useless in analyzing the effect of these modifications.

3. GREENSTADT'S VARIATIONAL APPROACH

In contrast with the previous approach which is based on properties of the search directions, Greenstadt's [6] approach is based on properties of the updating formulae. The objective here is to find a symmetric correction E_k to the inverse Hessian approximation D_k such that

$$D_{k+1} = D_k + E_k \qquad (11)$$

and

$$D_{k+1} q_k = p_k \qquad (12)$$

(q and p defined as before). Greenstadt felt that in order to avoid instability one should try to restrict the correction by minimizing some norm of E_k. For convenience reasons he chose the norm $N(E) = T_r(WEWE')$, where W is an arbitrary positive definite matrix. By minimizing $N(E_k)$ subject to (12) and the requirement that E_k be symmetric, Greenstadt arrived at the general updating formula

$$D_{k+1} = D_k + \frac{1}{q_k'M_kq_k}\{p_kq_k'M_k + M_kq_kp_k' - D_kq_kq_k'M_k - M_kq_kq_k'D_k - [(p_kq_k - q_k'D_kq_k)/q_k'M_kq_k]M_kq_kq_k'M_k\} \qquad (13)$$

where $M_k = W_k^{-1}$.

Special cases of this formula may be obtained by particular choices of M_k. It has been shown, for instance, by Goldfarb [18] that if we denote by E_k^1 the correction term in (13) corresponding to $M_k = D_k$ and by E_k^2 the correction corresponding to $M_k = D_{k+1}$, then the class of updating formulae $D_{k+1} = D_k + \theta E_k^1 + (1-\theta)E_k^2$ (where θ is an arbitrary scalar) is equivalent to Broyden's family. One should note, however, that Goldfarb's choice of M_k may violate the assumption that W_k is positive definite.

Except for special cases, (13) will not generate conjugate search direction even with perfect step size. On the other hand, this theory does not depend on whether the step size is perfect or not, which justifies using members of (13) without line search.

4. A GEOMETRIC APPROACH

The approach presented in this section adopts the view that the most important feature in quasi-Newton algorithms is the approximation of the inverse Hessian. One should realize that the fact that an algorithm converges in n steps for a quadratic case and produces the exact inverse Hessian at the n-th step, does not imply that it produces good approximations to the inverse Hessian at each iteration. On the contrary, it has been shown by Luenberger [19] and by Oren and Luenberger [12] that even in a quadratic problem the DFP algorithm may produce bad approximations to the inverse Hessian before n steps are completed. In such cases small perturbations in the objective function or the step size which destroy the conjugacy property cause the DFP method to perform worse than steepest descent.

Poor inverse Hessian approximations are usually caused by a poor initial approximation or a fast changing Hessian (in a nonquadratic case). Since the corrections in most variable metric methods are restricted to the direction of the updating vector p_k, it may take, even in a quadratic case, n iterations to compensate for a poor initial estimate. Furthermore, if the Hessian is changing, the quality of the approximation might deteriorate faster than it is improved so that even a good initial estimate may deteriorate and never recover. Poor inverse Hessian approximations D_k arising in this manner will usually generate poor search directions ($-D_kg_k$), since the gradient g_k is almost orthogonal to the latest updating vectors.

The Self-Scaling Variable Metric Algorithms described in [12] and [13] may be interpreted in this context as methods for correcting D_k through scaling with respect to an n-1 dimensional subspace not including p_k, in addition to the regular update in the direction p_k.

The following approach, which was motivated by the above considerations, is based on an algorithm proposed by Luenberger [19] that uses partial information about the Hessian by taking Newton steps restricted to the subspace over which the Hessian is known followed by steepest descent steps which are orthogonal to that subspace.

Let M be an m-dimensional subspace spanned by the column vectors of the $n \times m$ matrix B. Then minimizing the quadratic approximation of a function $f(x)$ over the linear variety $M + x_k$ yields a point z_k such that

$$z_k = x_k - B(B'\nabla^2 f(x_k)B)^{-1}B'\nabla f(x_k) . \qquad (14)$$

Equation (14) defines a Newton iteration restricted to the subspace M, and by analogy with the formula for Newton's method, $B(B'\nabla^2 f(x_k)B)^{-1}B'$ can be interpreted as the inverse of $\nabla^2 f(x_k)$ restricted to M.

Luenberger [19] also pointed out that in a quadratic case, the difference in gradients along the steepest descent step in his combined algorithm may be used to infer the inverse Hessian over a larger subspace. Implementing this idea in a recursive way yields a quasi-Newton algorithm. This can be done by updating the restricted inverse Hessian such as to expand the subspace M by one dimension at every iteration; then the (n+1)-th iteration will be a full Newton step that yields the minimum. The partial Newton steps in such a procedure will always yield the minimum over manifolds that contain the preceding updating vectors. Hence, finding the minimum along these vectors, i.e., the steepest directions, is no longer necessary. In fact, the difference of gradients q along any vector $p \notin M$ will provide enough information for updating the inverse Hessian restricted to M, and obtain the inverse Hessian restricted to $M + p$. Such an updating formula is provided in the next theorem.

Theorem 1. Let $f(x)$ be a positive definite quadratic function with Hessian H and $D_k = B_k(B_k'HB_k)^{-1}B_k$ be the inverse of H restricted to the subspace M_k where the columns of B_k form a basis for M_k. Let $p_k = x_{k+1} - x_k$ and $q_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ where x_{k+1} is such that $p_k \notin M_k + x_k$. Then

$$D_{k+1} = D_k + \frac{(p_k - D_kq_k)(p_k - D_kq_k)'}{q_k'(p_k - D_kq_k)} \qquad (15)$$

where $D_{k+1} = B_{k+1}(B_{k+1}'HB_{k+1})^{-1}B_{k+1}'$ and B_{k+1} is obtained by augmenting p_k to B_k such that $B_{k+1} = [B_k, p_k]$.

Proof: For simplicity we omit the subscripts and denote the entities corresponding to $k+1$ by $(\hat{\ })$. Define $A = (B'HB)^{-1}$ and $b = B'Hp$. Then $D = BAB'$, and since $q = Hp$, we have $b = B'q$. Using these definitions, one can express A^{-1} in the form

$$\hat{A}^{-1} = [B,p]'H[B,p] = \begin{bmatrix} B'HB & B'Hp \\ p'HB & p'Hp \end{bmatrix} = \begin{bmatrix} A^{-1} & b \\ b' & p'q \end{bmatrix} \quad (16)$$

Applying the standard formulae for inverting a partitioned matrix to (16) results in

$$\hat{D} = [B,p]\hat{A}[B,p]' = B(A^{-1} - \frac{bb'}{p'q})^{-1}(B' - \frac{bp'}{p'q}) + p(p'q-b'Ab)^{-1}(p'-b'AB') \quad (17)$$

Equation (15) can be obtained directly from (17) by using the Householder rank one modification formula and few manipulations.

The updating formula (15) is the well known rank-one formula first proposed by Broyden [3]. However, the above derivation and the way it is used here were originally proposed by Oren and Luenberger and first presented in [19].

The ideas described so far are integrated in the following crude algorithm.

Algorithm 1. Start with $D_o = [0]$, $M_o = \theta$ and $x_0 \in E^n$.

Step 1: Choose $p_k \notin M_k$

Step 2: Obtain $z_k = x_k + p_k$
$\hat{g}_k = \nabla f(z_k)$
$q_k = \hat{g}_k - g$

Step 3: Set

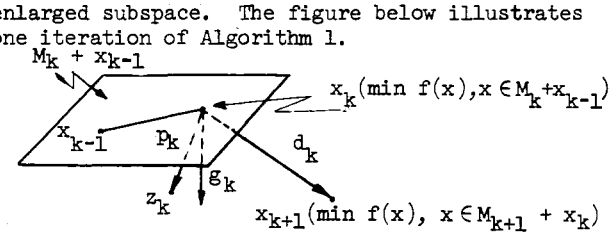
$$D_{k+1} = D_k + \frac{(p_k - D_k q_k)(p_k - D_k q_k)'}{q_k'(p_k - D_k q_k)}$$

$$M_{k+1} = M_k + p_k$$

Step 4: Set $d_k = - D_{k+1} g_k$ and obtain
$x_{k+1} = x_k + \beta_k d_k$ and
$g_{k+1} = \nabla f(x_k)$
(β_k will normally be 1).

Step 5: Add one to k and go to Step 1.

In the above algorithm p_k is an exploratory step that is taken at each iteration so that the updated D will predict the best next move in the enlarged subspace. The figure below illustrates one iteration of Algorithm 1.



For the quadratic case and for $k \le n$, x_{k+1} is the minimum over the linear variety $M_{k+1} + x_k$. This follows from the fact that x_{k+1} is determined by a Newton step restricted to M_{k+1}. Thus the gradient at x_{k+1} is orthogonal to M_{k+1}. This provides us a natural way of choosing $p_k \notin M_k$. Such a choice will be $p_k \in M_k + g_k$, $p_k \notin M_k$. In particular, one can simply take $p_k = g_k$.

It can be proved that if $p_k \in M_k + g_k$, $p_k \notin M_k$ for all k, then Algorithm 1 is a conjugate gradient algorithm. This proof which is omited here due to space limitation is based on the assumption that x_k is the minimum over $M_k + x_k$ for each k. One should note, however, that even if this is not true but the function is quadratic the updating formula will still yield the best possible approximation to the inverse Hessian at every step and $D_n = H^{-1}$. This implies that the (n+1)-th iteration consists of a full Newton step and x_{n+1} is the minimum.

I found out recently that an algorithm based on most of the ideas presented above has been independently developed by Maman and Mayne [20]. This algorithm, which is referred to as "Pseudo Newton-Raphson Method," is in principle similar to Algorithm 1, except for the way D is updated. Instead of updating D directly as done in Step 3 of Algorithm 1, Maman and Mayne update the matrix $P = B'\nabla^2 f(x)B$ and then calculate the new matrix $\hat{D} = BP^{-1}B$. The inversion of $\hat{P}$ is simplified by choosing the column vectors of B such as to make P diagonal. This is accomplished by storing the matrices B and $R = \nabla^2 f(x)B$. These matrices are updated at each step according to: $\hat{B} = [B,v]$ and $\hat{R} = [R,w]$, where $v = p - B(B'R)^{-1}B'q$ and $w = q - R(B'R)^{-1}B'q$. Then, in a quadratic case with Hessian H, $B'Hv = B'q - B'HB(B'R)^{-1}B'q = 0$ and $Hv = w$, so $\hat{P} = \begin{bmatrix} P & 0 \\ 0 & v'w \end{bmatrix}$.

One of the difficulties that would arise if we applied Algorithm 1 to a nonquadratic function follows from the fact that M becomes the entire space after n steps while the minimum is not necessarily reached. Since it is not possible to select a vector $p \notin M_n$ the algorithm cannot be continued. One way to proceed in such a case is to restart the algorithm. However, this is an undesirable approach, since it discards valuable information. An alternative approach described below is to discard only information for which replacement is available. To be more specific, let us consider a quadratic function with Hessian H and let D be a full rank approximation to H.

If D was the correct inverse Hessian we would have had $Dq = p$ for any p. We assume, however, that this is not the case and we choose an updating vector for which the above equality is not satisfied. Clearly, if such a p exists then at least in the direction p, D is the wrong approximation for the inverse Hessian. We wish to correct that discrepancy by replacing the information contained in D with respect to direction p while retaining the rest of the information corresponding to an (n-1)-dimensional subspace not containing p. We assume that over that subspace, D approximates the inverse Hessian correctly.

The replacement is done in two steps. First we obtain a restriction of D to the n-1 dimensional subspace $\tilde{M}$ which does not contain p. This restriction will be denoted by $\tilde{D}$ which is assumed to be the inverse Hessian restricted to $\tilde{M}$. Second we update $\tilde{D}$ by using Eq. (15) with p and q. Since $\tilde{D}$ is assumed to be the inverse Hessian restricted to $\tilde{M}$, we have $\tilde{D} = \tilde{B}(\tilde{B}'H\tilde{B})^{-1}\tilde{B}'$ where

$\tilde{B}$ is a matrix consisting of n-1 columns that span $\tilde{M}$. The only condition on $\tilde{M}$ is that it does not contain p; therefore $\tilde{M}$ can be chosen in an infinite number of ways. A convenient characterization of $\tilde{M}$ in terms of p is obtained by using a positive definite symmetric matrix G such that $\tilde{M} = \{y | y'Gp = 0\}$. Clearly every p and G will define a unique $\tilde{M}$ since $y'Gp = 0$ defines a unique hyperplane through the origin, which is orthogonal to the vector Gp. Furthermore, assuming $p \neq \theta$, p will not be included in $\tilde{M}$ since $p'Gp \neq 0$. In view of the above, for any given p we can choose implicitly $\tilde{M}$ that does not contain p by selecting a positive definite symmetric matrix G. As defined earlier $\tilde{B}$ is a matrix whose columns span $\tilde{M}$, and since $Gp \in M^{\perp}$ then $\tilde{B}Gp = \theta$. This implies $\tilde{D}Gp = \theta$. Another consideration in deriving $\tilde{D}$ is the reversibility of the process. We expect $\tilde{D}$ to be such that updating it, using (15) with p and $\tilde{q}$ where $\tilde{q} = D^{-1}p$, results in D. The following theorem provides a formula for deriving $\tilde{D}$ given D, p, G, that satisfies the conditions stated above.

Theorem 2. Let the matrix $\tilde{D}$ be defined by

$$\tilde{D} = D - \frac{DGpp'GD}{p'GDGp} \tag{18}$$

where p is a given vector and D and G are symmetric $n \times n$ matrices. Then $\tilde{D}Gp = \theta$ and

$$D = \tilde{D} + \frac{(p-\tilde{D}\tilde{q})(p-\tilde{D}\tilde{q})'}{\tilde{q}'(p-\tilde{D}\tilde{q})} \tag{19}$$

Proof: The condition $\tilde{D}Gp = \theta$ follows directly from (18). Equation (19) is proved by using (18) and $\tilde{q} = D^{-1}p$ to substitute $\tilde{D}$ and $\tilde{q}$ in its right hand side, and simplifying it.

After D is obtained by Eq. (18), we can use the new information encoded in p and q and update D using (15) to obtain the next approximation $\hat{D}$

$$\hat{D} = \tilde{D} + \frac{(p-\tilde{D}q)(p-\tilde{D}q)'}{(p-\tilde{D}q)'q} \tag{20}$$

Equations (18) and (20) form a two stage family of formulae for updating a full rank inverse Hessian approximation using the difference of gradients q along any given vector p. Particular selections of G will yield special cases of this family. Of special interest is to choose G such that $Gp = q$. In that case, (18) reduces to $\tilde{D} = D - Dqq'D/q'Dq$ and consequently

$$\hat{D} = D - \frac{Dqq'D}{q'Dq} + \frac{pp'}{p'q} \tag{21}$$

Equation (21) is the familiar DFP formula mentioned in the introduction to this paper. In the above derivation, however, we did not impose any restriction on the updating vector p, which suggests that (21) will improve (in some sense) the approximation to the inverse Hessian with any updating vector p. This conclusion is consistent with Fletcher's [17] observation, which is based on eigenvalue analysis, and justifies to some extent the use of (21) in algorithms with predetermined step size.

The above derivation of (21) also provides a geometric interpretation to the terms in the DFP formula. The first two terms represent the retained information on H^{-1} corresponding to a (n-1)-dimensional subspace that is H orthogonal to p, while the last term represents the updated information in the direction of p. "Self-Scaling" can thus be interpreted as proper weighting of the retained information relative to the new.

The view that (21) is a recursive formula which approximates the inverse Hessian independently of the updating vectors selection, enables us to obtain Broyden's [3] class of formulae by a simple extension of (21). One can argue that if D is an approximation to H^{-1}, then so is $(HDH)^{-1}$. Thus, pre- and post-multiplying (21) by H, substituting $q = Hp$ and then inverting it (by two applications of Householder's rank one modification), yields another updating formula (known as the complementary DFP or the BFS formula). Broyden's family is just a weighted sum of the DFP and BFS formulae, where the weight is the free parameter.

5. CONCLUSION

In this paper we presented a critical survey of unifying approaches to variable metric algorithms and introduced a new approach based on geometric considerations. Each one of these approaches leads to general classes of formulae for updating the inverse Hessian approximation which contain as special cases the commonly used formulae. Each of these approaches differs in the assumption and criteria used in the derivation and hence suggests different interrelations to the resulting updates. The insights obtained in deriving such updates from different approaches may prove valuable in understanding the observed characteristics of various updating formulae such as stability, sensitivity to line-search accuracy, etc. Particularly promising in this respect are approaches that focus on the generation of good inverse Hessian approximations rather than on the search directions. Approaches that focus on the conjugacy of the search directions are based on too restrictive assumptions which are unrealistic and suppress the capability to differentiate between the various updates.

REFERENCES

1. W. C. Davidon, Variable Metric Method for Minimization, A.E.C. Research and Development Rep. ANL-5990 (Rev.), 1959.
2. R. Fletcher and M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization," Comp. Jnl., vol. 6, pp. 163-168, 1963.
3. C. G. Broyden, "Quasi-Newton Methods and Their Application to Function Minimization," Maths. of Comp., vol. 21, pp. 368-381, 1967.
4. H. Y. Huang, "Unified Approach to Quadratically Convergent Algorithms for Function Minimization," J.O.T.A., vol. 5, pp. 405-423, 1970.
5. N. Adachi, "On Variable-Metric Algorithms," J.O.T.A., vol. 7, pp. 391-423, 1971.
6. J. Greenstadt, "Variations on Variable Metric Methods," Maths. of Comp., vol. 24, pp. 1-22, 1970.

7. S. S. Oren, "Self-Scaling Variable Metric Algorithms Without Line-Search for Unconstrained Minimization," Maths. of Comp., vol. 27, (1973).
8. E. Spedicato, Conditioning in Huang's Two Parameter Class of Conjugate Gradients, C.I.S.E-N-154, 1972.
9. D. F. Shanno, "Conditioning of Quasi-Newton Methods for Function Minimization," Maths. of Comp., vol. 24, pp. 647-657, 1970.
10. L. C. W. Dixon, "Variable Metric Algorithms: Necessary and Sufficient Conditions for Identical Behavior on Non-Quadratic Functions," J.O.T.A., vol. 10, pp. 34-40, 1972.
11. M. C. Biggs, "Minimization Algorithms Making Use of Non-Quadratic Properties of the Objective Function," J. Inst. Math. Applics, vol. 8, pp. 315-327, 1971.
12. S. S. Oren and D. G. Luenberger, "Self-Scaling Variable Metric (SSVM) Algorithms I: Criteria and Sufficient Conditions for Scaling a Class of Algorithms," (to appear in Man. Sci.).
13. S. S. Oren, "Self-Scaling Variable Metric (SSVM) Algorithms II: Implementation and Experiments," (to appear in Man. Sci.).
14. S. S. Oren, Self-Scaling Variable Metric Algorithms for Unconstrained Minimization, Ph.D. Thesis, Department of Engineering-Economic Systems, Stanford University, March 1972.
15. N. Adachi, "On the Uniqueness of Search Directions in Variable-Metric Algorithms," J.O.T.A., vol. 11, pp. 590-604, 1973.
16. R. Fletcher and C. M. Reeves, "Function Minimization by Conjugate Gradients," Comp. J., vol. 7, pp. 149-154, 1964.
17. R. Fletcher, "A New Approach to Variable Metric Algorithms," Comp. J., vol. 13, pp. 317-322, 1970.
18. D. Goldfarb, "A Family of Variable Metric Methods Derived by Variational Means," Maths. of Comp., vol. 24, pp. 23-26, 1970.
19. D. G. Luenberger, Introduction to Linear and Nonlinear Programming, Reading, Mass.: Addison-Wesley, 1973.
20. R. Mamen and D. Q. Mayne, "A Pseudo Newton-Raphson Method for Function Minimization," J.O.T.A., Vol. 10, pp. 263-276, 1972.