MINIMUM NORM SYMMETRIC QUASI-NEWTON UPDATES RESTRICTED TO SUBSPACES

Robert B. Schnabel

TR 77-300

Department of Computer Science
Cornell University
Ithaca, N.Y. 14853

Minimum Norm Symmetric Quasi-Newton Updates Restricted to Subspaces

Robert B. Schnabel

Abstract:

The Davidon-Fletcher-Powell and Broyden-Fletcher-Goldfarb-Shanno updates have been the two most successful quasi-Newton updates for a variety of applications. One reason offered in explanation is that they constitute, in an appropriate norm and metric, the minimum norm change to the matrix, or its inverse, being approximated which preserves symmetry and obeys the quasi-Newton equation. Recent methods have reason to consider updates restricted to certain subspaces. In this paper we derive the general minimum norm symmetric quasi-Newton updates restricted to such subspaces. In the same appropriate norm and metric, the minimum norm change update to the matrix or its inverse is shown to be, respectively, the rank-two update which is a particular projection of the DFP or BFGS onto this subspace.

Computer Science Department
Cornell University
Ithaca, New York 14853
November 1976

This research supported by a National Science Foundation Graduate Fellowship, forms a portion of the author's doctoral thesis at Cornell University under the supervision of Professor J.E. Dennis.

1. Introduction

Quasi-newton updates, first proposed by Davidon (1959), are used in iterative algorithms to approximate matrices which are either impossible or very costly to compute. At the kth iteration, a matrix $\bar{B}$ which approximates M(k) is calculated from a matrix B which approximates M(k-1) by updating B to reflect some additional information about M(k).

When the matrices M(k) are symmetric - as is the case, for example, when they are matrices of second partial derivatives in optimization algorithms - the two most popular updates have involved adding a symmetric matrix of rank two to B to form $\bar{B}$. These two updates, the Davidon-Fletcher-Powell (DFP) (Davidon (1959), Fletcher and Powell (1963)) and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) (Broyden (1970), Fletcher (1970), Goldfarb (1970), Shanno (1970)), are the choices $\phi=0,1$ respectively in the formula

$$\bar{B}=B+\frac{(y-Bs)y^T + y(y-Bs)^T}{\langle y,s\rangle} - \frac{\langle y-Bs,s\rangle yy^T}{\langle y,s\rangle^2} - \phi \langle s,Bs\rangle \left[\frac{y}{\langle y,s\rangle} - \frac{Bs}{\langle s,Bs\rangle}\right]\left[\frac{y}{\langle y,s\rangle} - \frac{Bs}{\langle s,Bs\rangle}\right]^T \quad (1.1)$$

where $B,\bar{B}$ are matrices $\in L(R^n)$, s,y vectors $\in R^n$, and $\langle v,w\rangle$ denotes the scalar product v^Tw.

For any choice of the scalar ϕ in (1.1), $\bar{B}s=y$. This equation is called the quasi-Newton equation, and is the way in which new information is supplied to $\bar{B}$. In the most common application, minimization of a function $f:R^n\to R$, B and $\bar{B}$ represent approximations to the Hessian $\nabla^2 f$ at points x and $\bar{x}$; $s=\bar{x}-x$, and $y=\nabla f(\bar{x})-\nabla f(x)$.

One explanation offered for the success of the DFP and BFGS updates is that, among all possible updates, the DFP is the minimum norm change to the matrix B,

in a very reasonable norm, consistent with preserving symmetry and obeying the quasi-Newton equation; while the BFGS is the analagous minimum norm change from B^{-1} to $\bar{B}^{-1}$ (Dennis and Moré(1977)). This is considered advantageous because, aside from $\bar{B}s=y$, we have no new information about $\bar{B}$; so that consistent with $\bar{B}s=y$ and $\bar{B}$ symmetric we would like to change B as little as possible, to preserve information already in it.

Recently, Davidon (1975) introduced an optimization algorithm which also requires that at each iteration $(\bar{B}-B)$ be orthogonal to some m dimensional subspace Z, $0 \leq m < n$. (m and Z can change at each iteration.) By his choices of this subspace, Davidon's algorithm is able to find the minimum of a quadratic function in a finite number of steps. His updates are chosen from the class

$$\bar{B}=B+\frac{(y-Bs)c^T + c(y-Bs)^T}{\langle c,s\rangle} - \frac{\langle y-Bs,s\rangle\, cc^T}{\langle c,s\rangle^2} - \phi \left[\frac{y-Bs}{\langle y-Bs,s\rangle} - \frac{c}{\langle c,s\rangle}\right]\left[\frac{y-Bs}{\langle y-Bs,s\rangle} - \frac{c}{\langle c,s\rangle}\right]^T \quad (1.2)$$

where (y-Bs) and c are orthogonal to Z. Equation (1.2) still preserves symmetry and obeys the quasi-Newton equation.

Therefore, the question "which update $\bar{B}$ minimizes the norm of $(\bar{B}-B)$, subject to $B,\bar{B}$ symmetric, $\bar{B}s=y$, and $(\bar{B}-B)$ orthogonal to a subspace Z?" is of interest -- for if past experience is a guide, it may indicate good updates to use. In this paper we answer this question for the same general class of norms considered by Dennis and Moré (1977). When we use the same specific norms for which they showed the DFP and BFGS optimal, we now find that the solutions are the "restricted DFP" and "restricted BFGS" updates: formula (1.1) with $\phi=0$ and 1 but where y and Bs are replaced by the projections of these quantities onto the restricted updating subspace in the metric B^{-1}. Davidon's (1975) algorithm uses a one-parameter class of updates which contains both of these updates.

In section 2 we introduce some background material on norms and on projection matrices. Projection matrices are used in this paper as an algebraic tool, and to aid our geometric understanding of the results. In section 3 we first repeat the results of Dennis and Moré (1977). We then show when our restricted update is feasible; what the general minimum norm update is in this case; and that the minimum norm restricted update in our two appropriate norms are the restricted DFP and restricted BFGS. In section 4 we summarize our results, and comment upon how they should influence the choice of update in existing algorithms, including the use of optimal conditioning.

2. Mathematical Background

In this section we introduce the weighted Frobenius norm and the projection matrices which we use in section 3.

Definition 2.1 Let A be a matrix $\in L(R^n)$. The Frobenius norm of A, denoted $||A||_{Fr}$, is defined as the ℓ_2 norm of the elements of A,

$$||A||_{Fr} = \left(\sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij}^2 \right)^{1/2}.$$

Equivalently, $||A||_{Fr}$ can be computed by

$$||A||_{Fr} = \left(\sum_{i=1}^{n} ||Av_i||_2^2 \right)^{1/2} = (\text{trace } A^T A)^{1/2}$$

where $\{v_1, v_2, \ldots, v_n\}$ is any orthonormal set of vectors.

The weighted Frobenius norm of A, $||A||_{Fr,M}$ is defined by

$$||A||_{Fr,M} = ||MAM||_{Fr}$$

where M is a symmetric non-singular matrix $\in L(R^n)$.

Definition 2.2 Let Z be an m dimensional subspace of R^n, $1 \leq m < n$; M a symmetric positive definite matrix $\in L(R^n)$. Let $Z^{\perp}(M) = \{v \mid v^T Mz = 0 \; \forall z \in Z\}$. Then a projection matrix projecting orthogonal to Z in the metric M (or equivalently, onto $Z^{\perp}(M)$ in metric M) is a matrix $P \in L(R^n)$ such that:

$$P^2 = P \; ; \; Pz = 0 \text{ for all } z \in Z; \; Pv = v \text{ for all } v \in P^{\perp}(M).$$

Lemma 2.3 Let Z, M, $Z^{\perp}(M)$ be defined as in Definition 2.2. Let $z_1, \ldots, z_m$ be a basis for Z such that $\langle z_i, Mz_j \rangle = 0$, $1 \leq i < j \leq m$; $v_1, \ldots, v_{n-m}$ a basis for $Z^{\perp}(M)$ such that $\langle v_i, Mv_j \rangle = 0$, $1 \leq i < j \leq n-m$. Then the unique projection matrix projecting orthogonal

to Z in metric M is

$$P = I - \sum_{i=1}^{m} \frac{z_i z_i^T M}{\langle z_i, M z_i \rangle}$$

$$= \sum_{i=1}^{n-m} \frac{v_i v_i^T M}{\langle v_i, M v_i \rangle}$$

and $MP = P^T M$.

Proof: See Schnabel (1977) - proof is straightforward from Definition 2.2.

3. Minimum Norm Symmetric Quasi-Newton Updates

We start this section by stating Dennis and Moré's (1977) theorem giving the $\bar{B}$ which minimizes $||\bar{B}-B||_{Fr,M}$ subject to B, $\bar{B}$ symmetric and $\bar{B}s=y$; their techniques of proof are used in the proof of Theorem 3.4. The Dennis-Moré result shows that $||\bar{B}-B||_{Fr,\bar{B}}$ - 1/2 and $||\bar{B}^{-1}-B^{-1}||_{Fr,\bar{B}}$1/2 are minimized by the DFP and BFGS updates, respectively, and we comment on why these are relevant norms. We then consider updates which in addition require $(\bar{B}-B)$ orthogonal to a subspace Z. Lemma 3.3 states the condition on Z such that this restriction on $\bar{B}$ is consistent with B, $\bar{B}$ symmetric and $\bar{B}s=y$; Theorem 3.4 then gives the $\bar{B}$ which minimizes $||\bar{B} - B||_{Fr,M}$ subject to these restrictions. Corollary 3.5 and Theorem 3.6 show that when the update $\bar{B}$ is so restricted, that $||\bar{B} - B||_{Fr,\bar{B}}$-1/2 and $||\bar{B}^{-1} - B^{-1}||_{Fr,\bar{B}}$ 1/2 are minimized by the restricted DFP and the restricted BFGS, respectively.

Theorem 3.1 (Dennis and Moré (1977)) Let s,y be non-zero vectors $\in R^n$; B, M symmetric matrices $\in L(R^n)$, M non-singular. Then the unique solution to

$$\min \{ ||\bar{B} - B||_{Fr,M} \mid \bar{B} \text{ symmetric}, \bar{B}s=y \} \tag{3.1}$$

is

$$\hat{B} = B + \frac{(y-Bs)v^T + v(y-Bs)^T}{\langle v,s\rangle} - \frac{\langle y-Bs,s\rangle vv^T}{\langle v,s\rangle^2}, \quad v = M^{-2}s \tag{3.2}$$

Proof: a slightly simpler version of the proof of Theorem 3.4.

Corollary 3.2 (Dennis and Moré (1977)). Let s,y,B be defined as in Theorem 3.1. If B is positive definite and $\langle s,y\rangle > 0$, then the DFP update, (3.2) with $v=\bar{B}s=y$, is positive definite and is the unique solution to (3.1) with $M=\bar{B}^{-1/2}$, among all positive definite $\bar{B}$.

If we are dealing with minimization problems where $\bar{B}$ approximates $\nabla^2 f(\bar{x})$, then (especially as $\bar{x}$ approaches the solution x^*) the natural scaling of the variables is by $\nabla^2 f(x^*)^{-1/2}$ (assuming $\nabla^2 f(x^*)$ positive definite.) This scaling, which makes the contour lines at the solution circular, corresponds to using norms weighted by $\nabla^2 f(x^*)^{-1/2}$. Therefore we would really like to solve (3.1) with $M = \nabla^2 f(x^*)^{-1/2}$. Since the best approximation we have so far to $\nabla^2 f(x^*)$ is most likely $\bar{B}$, we are especially interested in the solution to (3.1) with $M = \bar{B}^{-1/2}$. This is why Corollary 3.2 is important, and gives support to the use of the DFP update.

The condition in Corollary (3.2) that B be positive definite and $\langle s,y \rangle$ be > 0 is realistic, as most algorithms require this. The use of the Frobenius norm is natural because it is the obvious measure of the size of the change in a matrix, the ℓ_2 norm of the changes in its elements.

The analagous theorem to Theorem 3.1 can be proven about updates to B^{-1}. In particular, Dennis and Moré show that the unique solution to

$$\min \{ \| B^{-1} - B^{-1} \|_{Fr,B}^{1/2} \mid B,\bar{B} \text{ symmetric and positive definite, } \bar{B}s=y \} \quad (3.3$$

is the BFGS update. Since the "natural" scaling for inverse Hessian approximations is by $\nabla^2 f(x^*)^{1/2}$ on each side, we similarly feel that the above norm is a very desirable one. Because our algorithms actually only use B^{-1} and $\bar{B}^{-1}$, the solution to (3.3) may be of primary interest to us, and specially support use of the BFGS.

Now we wish to consider updates which also require $(\bar{B}-B)$ orthogonal to some m dimensional subspace Z. We first show for which subspaces Z this restriction is consistent with $B,\bar{B}$ symmetric and $\bar{B}s=y$. We are only interested in cases where $Bs \neq y$, for if $Bs=y$ we will set $\bar{B}=B$.

Lemma 3.3 Let B be a symmetric matrix $\in L(R^n)$; s,y non-zero vectors $\in R^n$, $y \neq Bs$; Z an m dimensional subspace of R^n, $m<n$. Define $Z^{\perp} = \{v \mid \langle v,z\rangle = 0 \forall z \in Z\}$. Then the set $S = \{\bar{B} \mid \bar{B}$ symmetric, $\bar{B}s=y$, $(\bar{B}-B)z = 0 \forall z \in Z\}$ is nonempty if and only if $s \notin Z$ and $(y-Bs) \in Z^{\perp}$.

Proof: First assume S is nonempty. Then for any $\bar{B} \in S$ and $z \in Z$, $(\bar{B}-B)z = 0$. Since $(\bar{B}-B)s = y-Bs \neq 0$, $s \notin Z$. Furthermore, since $s^T(\bar{B}-B)z = 0$ for all $z \in Z$ and $(\bar{B}-B)$ is symmetric, $(y-Bs)^T z = 0$ for all $z \in Z$. Thus $(y-Bs) \in Z^{\perp}$.

Now assume $s \notin Z$ and $(y-Bs) \in Z^{\perp}$. If $\langle y-Bs,s\rangle \neq 0$, then the symmetric rank one update $B=B+\frac{(y-Bs)(y-Bs)^T}{\langle y-Bs,s\rangle}$ is in S. If $\langle y-Bs,s\rangle = 0$, then since $s \notin Z$, there exists $v \in Z^{\perp}$ such that $\langle v,s\rangle \neq 0$. Then $\bar{B} = B + \frac{(y-Bs)v^T + v(y-Bs)^T}{\langle v,s\rangle}$ is in S.

By expanding the techniques used by Dennis and Moré (1977) to prove Theorem 3.1 we now find the minimum norm restricted quasi-Newton update in a general weighted Frobenius norm.

Theorem 3.4 Let B,M be symmetric matrices $\in L(R^n)$, M non-singular; s,y non-zero vectors $\in R^n$, $y \neq Bs$. Let Z be an m dimensional subspace of R^n, $m<n$, such that $s \notin Z$ and $(y-Bs) \in Z^{\perp}$ ($Z^{\perp}$ defined as in Lemma 3.3). Then the unique solution to

$$\min \{ \|\bar{B}-B\|_{Fr,M} \mid \bar{B} \text{ symmetric}, \bar{B}s=y, (\bar{B}-B)z = 0 \forall z \in Z\} \tag{3.4}$$

is

$$B = B + \frac{(y-Bs)v^T + v(y-Bs)^T}{\langle v,s\rangle} - \frac{\langle y-Bs,s\rangle vv^T}{\langle v,s\rangle^2} \tag{3.5a}$$

where v is the projection of $M^{-2}s$ orthogonal to $M^{-2}Z$ in M^2 metric,

$$v = M^{-2}s - \sum_{i=1}^{m} \frac{\langle M^{-2}s, z_i\rangle}{\langle M^{-2}z_i, z_i\rangle} M^{-2}z_i, \qquad (3.5b)$$

$\{z_1,\ldots,z_m\}$ a basis for Z such that $\langle z_i, M^{-2} z_j\rangle = 0, 1 \le i < j \le m$.

Proof: Let $S = \{\bar{B} \mid \bar{B}$ symmetric, $\bar{B}s=y$, $(\bar{B}-B)z = 0 \ \forall z\in Z\}$. From (3.5a), $\hat{B}$ is symmetric and $\hat{B}s=y$. From (3.5b), $\langle v,z\rangle = 0$ for any $z\in Z$. Since also $\langle y-Bs,z\rangle = 0$ for all $z\in Z$, $(\bar{B}-B)z = 0$ for all $z\in Z$. Therefore $\hat{B}$ is in S.

Now consider any $\bar{B}$ in S. Define $\hat{E} = M(\hat{B}-B)M$, $\bar{E} = M(\bar{B}-B)M$. Since $y-Bs = (\bar{B}-B)s$ and $(\bar{B}-B)$ is symmetric, subtracting B from both sides of (3.5a) and pre and post-multiplying by M gives

$$\hat{E} = \frac{\bar{E}\,(M^{-1}s)(Mv)^T + (Mv)(M^{-1}s)^T\bar{E}}{\langle Mv, M^{-1}s\rangle} - \frac{\langle M^{-1}s, \bar{E}\,M^{-1}s\rangle\,(Mv)(Mv)^T}{\langle Mv, M^{-1}s\rangle^2} \qquad (3.6)$$

Define $\delta = v-M^{-2}s$; $\delta = -\sum_{i=1}^{m} \frac{\langle M^{-2}s, z_i\rangle}{\langle M^{-2}z_i, z_i\rangle} M^{-2}z_i$. Since $\delta\in M^{-2}Z$, $(\bar{B}-B)M^2\delta = 0$. Therefore $\bar{E}\cdot Mv = M(\bar{B}-B)M^2v = M(\bar{B}-B)s = M(y-Bs)$. Similarly $\hat{E}\cdot Mv = M(y-Bs)$. Thus $\bar{E}\cdot Mv = \hat{E}\cdot Mv$.

Now consider any $w\in R^n$ such that $\langle w, Mv\rangle = 0$. Then (3.6) gives

$$\hat{E}w = \frac{(Mv)(M^{-1}s)^T\bar{E}w}{\langle Mv, M^{-1}s\rangle} \qquad (3.7)$$

Again $(M^{-1}s)^T\bar{E} = (Mv)^T\bar{E}$; and $\langle Mv, M^{-1}s\rangle = \langle Mv, Mv+M\delta\rangle = \langle Mv, Mv\rangle$ as $v\in Z^{\perp}$, $M^2\delta\in Z$.

Therefore from (3.7),

$$\hat{E}w = \frac{(Mv)(Mv)^T\bar{E}w}{\langle Mv, Mv\rangle} \quad ;$$

$$\|\hat{E}w\|_2 \le \left\|\frac{(Mv)(Mv)^T}{\langle Mv, Mv\rangle}\right\|_2 \|\bar{E}w\|_2 = \|\bar{E}w\|_2 .$$

Thus since $||\hat{E}b_i||_2 \leq ||\bar{E}b_i||_2$ for any orthonormal set of vectors $\{b_i\} = (\frac{Mv}{||Mv||}_2, w_1, \ldots, w_{n-1})$, by the definition of Frobenius norm $||\hat{E}||_{Fr} \leq ||\bar{E}||_{Fr}$. $\hat{B}$ is the unique solution to (3.4) because $f:L(R^n)\mapsto R$, $f(\bar{B}) = ||\bar{B}-B||_{Fr,M}$ is strictly convex over all $\bar{B}$ in S.

Corollary 3.5 Let s,y,B,Z be defined as in Theorem 3.4; in addition let B be positive definite. Let P be the projection matrix projecting orthogonal to BZ in the B^{-1} matrix. Then if $\langle s,Py\rangle > 0$, the restricted DFP update,

$$\hat{B} = B + \frac{(\hat{y}-B\hat{s})\hat{y}^T + \hat{y}(\hat{y}-B\hat{s})^T}{\langle\hat{y},\hat{s}\rangle} - \frac{\langle\hat{y}-B\hat{s},\hat{s}\rangle\,\hat{y}\hat{y}^T}{\langle\hat{y},\hat{s}\rangle^2} \tag{3.8}$$

$\hat{y} = Py$, $\hat{s} = P^Ts$, is positive definite and is the unique solution to problem (3.4) with $M = \bar{B}^{-1/2}$.

Proof: If any positive definite $\bar{B}$ exist, Theorem 3.4 shows that the unique solution to problem (3.4) with $M = B^{-1/2}$ is

$$\hat{B} = B + \frac{(y-Bs)v^T + v(y-Bs)^T}{\langle v,s\rangle} - \frac{\langle y-Bs,s\rangle\, vv^T}{\langle v,s\rangle^2}, \tag{3.9a}$$

$$v = \bar{B}s - \sum_{i=1}^{m} \frac{\langle \bar{B}s,z_i\rangle\ \bar{B}z_i}{\langle \bar{B}z_i,z_i\rangle}, \tag{3.9b}$$

where $\{z_i\}$ is a basis for Z such that $z_i^T\bar{B}z_j = 0$, $1 \leq i<j \leq m$. Since $\bar{B}s=y$, and $\bar{B}z_i = Bz_i$ for all $z_i\epsilon Z$, (3.9b) is equivalent to

$$v = y - \sum_{i=1}^{m} \frac{\langle y,z_i\rangle\ Bz_i}{\langle Bz_i,z_i\rangle}$$

= the projection of y orthogonal to BZ in the B^{-1} metric.

Therefore $v = Py = \hat{y}$.

By Lemma 2.3, $PB = BP^T$; and since $y-Bs$ is orthogonal to BZ in the B^{-1} metric, $y-Bs = P(y-Bs)$, $= Py - BP^Ts = \hat{y} - B\hat{s}$. Also, $\langle v,s\rangle = \langle \hat{y},s\rangle = \langle Py,s\rangle = \langle P^2y,s\rangle = \langle Py,P^Ts\rangle = \langle \hat{y},\hat{s}\rangle$; similarly, $\langle y-Bs,s\rangle = \langle \hat{y}-B\hat{s},\hat{s}\rangle$. Therefore (3.9a,b) is equivalent to (3.8).

Finally, from (Fletcher (1970)), $\bar{B}$ is positive definite if and only if $\langle \hat{s},\hat{y}\rangle = \langle s,Py\rangle > 0$. If $\langle s,Py\rangle \leq 0$, then there are no positive definite $\bar{B}$ feasible to (3.4) with B positive definite (Schnabel (1977)).

So we see that in the same norm for which the DFP is the minimum norm symmetric update obeying the quasi-Newton equation, the restricted DFP -- the projection of the DFP update onto the restricted feasible set of updates -- is the minimum norm restricted update. This gives us confidence in the use of the restricted DFP.

Again, a theorem analagous to Theorem 3.4 can be derived, giving the minimum norm restricted update of B^{-1} to $\bar{B}^{-1}$. For many algorithms this is really the relevant problem, as we are only concerned with approximating the inverse of the Hessian matrix. The theorem shows that the restricted BFGS update is the optimal restricted change to the inverse in the same norm for which the regular BFGS is the optimal unrestricted update.

Theorem 3.6 Let B,M be symmetric non-singular matrices $\in L(R^n)$; s,y non-zero vectors R^n, $y \neq Bs$. Let Z be an m-dimensional subspace of R^n, $m<n$, such that $s \notin Z$ and $(y-Bs) \in Z^{\perp}$. Then the unique solution to

$$\min \{ ||\bar{B}^{-1} - B^{-1}||_{Fr,M} \mid \bar{B} \text{ symmetric}, \bar{B}s = y, (\bar{B}-B)z = 0 \ \forall z \in Z \} \quad (3.10)$$

is

$$\bar{B}^{-1} = B^{-1} + \frac{(s-B^{-1}y)w^T + w(s-B^{-1}y)^T}{\langle w,y\rangle} - \frac{\langle s-B^{-1}y,y\rangle\, ww^T}{\langle w,y\rangle^2} ,$$

where w is the projection of $M^{-2}y$ orthogonal to $M^{-2}(BZ)$ in the M^2 me

If in addition B is positive definite, and P is the projection matrix projecting orthogonal to BZ in the B^{-1} metric, then if $\langle s,Py\rangle > 0$, the restricted BFGS update,

$$\bar{B}^{-1} = B^{-1} + \frac{(\hat{s}-B^{-1}\hat{y})\,\hat{s}^T + \hat{s}(\hat{s}-B^{-1}\hat{y})^T}{\langle \hat{s},\hat{y}\rangle} - \frac{\langle \hat{s}-B^{-1}\hat{y},\hat{y}\rangle\,\hat{s}\hat{s}^T}{\langle \hat{s},\hat{y}\rangle^2}$$

$\hat{y} = Py$, $\hat{s} = P^T s$, is positive definite and is the unique solution to (3.10) with $M = \bar{B}^{1/2}$.

The proof of Theorem 3.6 is analagous to that of Theorem 3.4 and Corollary 3.5, given the observation that the condition "$(\bar{B}-B)z = 0$ for all $z \in Z$" is equivalent to "$(\bar{B}^{-1}-B^{-1})u = 0$ for all $u \in BZ$".

The restrictions on the update in Corollary 3.5 and its analog in Theorem 3.6 -- that $\bar{B}$ be symmetric and positive definite, obey $\bar{B}s=y$, and have $(\bar{B}-B)z = 0$ for all z in some subspace Z -- correspond to Davidon's conditions upon his algorithm when it is applied to a quadratic function. In this case, Z contains past values of s, and grows by one dimension at each iteration. Davidon selects his update from that family of updates with one free parameter which includes the restricted DFP and BFGS -- so he has selected the one parameter family containing our two "least change" updates.

For non-quadratic problems, Davidon's algorithm restricts $\bar{B}-B$ at each iteration to a two dimensional subspace spanned by $y-Bs$ and some other vector c (which changes at each iteration). This is equivalent to defining Z in the above theorems as the $n-2$ dimensional subspace orthogonal to $y-Bs$ and c. Theorems 3.4-6 apply perfectly well to this case. However, this value of Z, in the non-quadratic case, is not necessarily the one which Davidon's own motivation would seem to recommend. We comment upon this briefly in section 4.

4. Summary and Conclusions

We have considered quasi-Newton algorithms which at each iteration, update a symmetric matrix approximation B into a new symmetric approximation $\bar{B}$, subject also to the quasi-Newton equation $\bar{B}s=y$, and the restriction that $\bar{B}-B$ be orthogonal to some m dimensional subspace Z. Dennis and Moré (1977) have shown which $\bar{B}$ minimizes $||\bar{B}-B||_{Fr,M}$ subject only to $\bar{B}$ symmetric and $\bar{B}s=y$. The additional restriction involving Z comes originally from Davidon's (1975) algorithm, which showed how the choice of Z could produce algorithms with attractive additional properties. So in this paper we have shown which $\bar{B}$ minimizes $||\bar{B}-B||_{Fr,M}$ (or $||B^{-1}-\bar{B}^{-1}||_{Fr,M}$) subject also to this restriction -- after first showing for which Z the restriction is consistent with the other conditions upon $\bar{B}$. We find that in the same norms for which the popular DFP and BFGS updates minimize $||\bar{B}-B||$ and $||\bar{B}^{-1}-B^{-1}||$, respectively, subject to symmetry and the quasi-Newton equation, that the restricted DFP and restricted BFGS updates -- the projections of the DFP and BFGS onto the restricted updating space -- minimize the same norms among all updates obeying in addition the condition involving Z.

Our results, and our way of viewing Davidon's work, generally support his choice of update, but they do suggest two possible changes to his algorithm: one in the selection of Z, the second in the selection of the restricted update. In the quadratic case, Davidon shows that he in effect selects Z to contain all past values of s. This is desirable, as it means that past quasi-Newton equations remain valid; and it leads to $n+1$ step termination. In the non-quadratic case values of s from previous iterations (call them z) may not be orthogonal to $y-Bs$ (as Z must be by Lemma 3.3), but Davidon's reasoning suggests letting Z contain

the components $\hat{z}$ of several past z's, which are orthogonal to y-Bs. Davidon, partly for algebraic convenience, instead implicitly chooses an $n-2$ dimensional Z which is only guaranteed to contain the last $\hat{z}$. Since this paper shows how to choose minimum norm updates for more general Z without too much algebraic trouble, we therefore plan to modify Davidon's algorithm to use what seems to us a more natural restricting space Z.

Secondly, once Davidon has chosen his restricting space Z, he doesn't necessarily use the restricted DFP or BFGS updates (call them $\bar{B}_1$ and $\bar{B}_2$), but rather a positive definite combination $\phi\bar{B}_1 + (1-\phi)\bar{B}_2$ chosen to minimize a bound on the growth in the condition number of $\bar{B}$. Mei (1977) however suggested that the restricted BFGS may be a better choice, and tests of Mei (1977) and Schnabel (1977) have borne this out. Theoretical work of Schnabel (1976) also supports choosing the restricted BFGS mostly, and otherwise a convex combination of the restricted DFP and BFGS, $\phi \in 0,1$. The results of this paper, which show the restricted BFGS and DFP to have attractive minimum norm change properties, support the suggestions of Mei and Schnabel.

Acknowledgement

My thanks to the Atomic Energy Research Establishment, Harwell England, for providing the facilities to perform a part of this research; and to Professor John Dennis, who suggested extending the minimum norm change theory to restrcted updates.

Bibliography

Broyden, C.G. (1970), "The convergence of a class of double-rank minimization algorithms," Journal of the Institute of Mathematics and its Applications 6, 76-90.

Davidon, W.C. (1959), "Variable metric algorithm for minimization," Argonne National Laboratory Report ANL-5990 (Rev.).

Davidon, W.C. (1975), "Optimally conditioned optimization algorithms without line searches," Mathematical Programming 9, 1-30.

Dennis, J.E. and Moré, J. (1977), "Quasi-Newton methods, motivation and theory," to appear in SIAM Review, 1977.

Fletcher, R. (1970), "A new approach to variable metric methods," The Computer Journal 13, 317-322.

Fletcher, R. and Powell, M.J.D. (1963), "A rapidly convergent descent method for minimization," The Computer Journal 6, 163-168.

Goldfarb, D. (1970), "A family of variable metric updates derived by variational means," Mathematics of Computation 24, 23-26.

Mei, H. H-W (1977) Ph.D. thesis, Cornell University.

Schnabel, R.B. (1976), "Optimal conditioning in the convex class of rank-two updates," Cornell University Computer Science TR 76-288.

Schnabel, R.B. (1977), Ph.D. thesis, Cornell University.

Shanno, D.F. (1970), "Conditioning of quasi-Newton methods for function minimization," Mathematics of Computation 24, 647-656.