LEAST CHANGE SECANT UPDATES

FOR QUASI-NEWTON METHODS

by

J.E. Dennis, Jr.
R.B. Schnabel

TR78-344

Department of Computer Science
Cornell University
Ithaca, NY 14853

Department of Computer Science
University of Colorado
Boulder, CO 80309

Research supported by NSF MCS76-0032 and by an NSF Graduate Fellowship.

LEAST CHANGE SECANT UPDATES FOR QUASI-NEWTON METHODS

by

J.E. Dennis Jr.
Cornell University
Ithaca, NY 14853

R.B. Schnabel
University of Colorado
Boulder, CO 80309

Abstract

In many problems involving the solution of a system of nonlinear equations, it is necessary to keep an approximation to the Jacobian matrix which is updated at each iteration. Computational experience indicates that the best updates are those that minimize some reasonable measure of the change to the current Jacobian approximation subject to the new approximation obeying a secant condition and perhaps some other approximation properties such as symmetry.

In this paper we extend the affine case of a theorem of Cheney and Goldstein on proximity maps of convex sets to show that a generalization of the symmetrization technique of Powell always generates least change updates. This generalization has such broad applicability that we obtain an easy unified derivation of all the most successful updates. Furthermore, our techniques apply to interesting new cases such as when the secant condition might be inconsistent with some essential approximation property like sparsity. We also offer advice on how to choose the properties which are to be incorporated into the approximations and how to choose the measure of change to be minimized.

Research supported by NSF MCS76-0032 and by an NSF Graduate Fellowship.

1. Introduction

Many problems require the numerical solution of a system of n nonlinear equations in n unknowns:

$$\text{given } F:\mathbb{R}^n \to \mathbb{R}^n, \text{ find } x_* \in \mathbb{R}^n \text{ such that } F(x_*) = 0. \tag{1.1}$$

For example, (1.1) arises in optimization problems, and in finding equilibrium points of nonlinear systems. The numerical solution of (1.1) is usually iterative, proceeding at each iteration from an estimate x of x_* to a better estimate x_+. In most algorithms, each iteration includes calculation of the Newton step, $s_N = -J(x)^{-1}F(x)$, the amount by which x would differ from x_* if F were linear. Our notation is

$$J(x) \in L(R^n),\ J(x)^{ij} = \left.\frac{\partial f^i}{\partial x^j}\right|_x , \tag{1.2}$$

f^i the ith component function of F, x^j the jth component of the vec x, $J(x)^{ij}$ the component of the matrix $J(x)$ in row i and column j.

In many cases, such as when $F(x)$ is the output from a subroutin with input parameter x, calculation of $J(x)$ or approximation by finite differences is either impossible, prohibitively expensive or very prone to human error. In these cases $J(x)$ (or some portion of it) is replaced by an approximation A, and at each iteration the current A is updated to an approximation A_+ of $J(x_+)$. Such updates are the topic of this paper. To incorporate derivative information the approximations usually are chosen to obey the secan equation,

$$A_+(x_+-x) = F(x_+) - F(x) \tag{1.3}$$

(which holds exactly for linear F), and hence we call them secant updates. We call the resultant algorithms, which replace the Newton step $-J(x)^{-1}F(x)$ by the approximation $-A^{-1}F(x)$, quasi-Newton.

If $n \geq 2$ and $x_+-x \neq 0$, many matrices will obey (1.3). Therefore, if the Jacobian has special properties, such as symmetry (which it does in optimization applications) or a significant number of positions which are always zero (sparsity), then each A_+ is further restricted to the subset $\mathcal{A}$, of matrices which have these desirable approximation properties. It has been found that the most successful updates are the ones that then chose the A_+ which, for some appropriate matrix norm, solves

$$\min_{A_+\in\mathcal{A}} ||A_+-A|| \text{ subject to (1.3).} \tag{1.4}$$

Update selection strategy (1.4) is good because it helps preserve information from previous iterations. The resultant updates could be called least change secant updates (Dennis and Tapia [8]).

The leading Jacobian updates for various forms of problem (1.1) are all of the least change secant form (1.4), although this was not the original motivation behind most of them. In this paper we give a unified treatment of the derivation of all the important updates, using a simple geometric property of affine sets. The updates discussed have previously been shown to be least change secant updates [10,16,17,18,23,24] but the techniques of proof are new and easier, and should facilitate the development of updates for problems where the Jacobian approximation is required to have different or additional properties. Furthermore, we have found this viewpoint to be a very effective teaching device.

In section 2 we give an easy derivation of Broyden's (1965) least change secant update [1], the case when $\mathcal{A}$ in (1.4) is $L(R^n)$. In section 3 we derive a general technique for adding additional properties to the least change secant update (i.e., restricting $\mathcal{A}$ in (1.4)), using the method of iterated projections. In fact, our approach allows for important cases when it is vital to choose our approximations from $\mathcal{A}$ even if this means we can not exactly satisfy (1.3). When this happens we show how to get $A_+ \in \mathcal{A}$ as near as possible to satisfying (1.3). Furthermore, our technique allows the derivation in section 4 of other well-known updates including Powell's (1970) least change symmetric secant [19] , the DFP [6], [13] and BFGS [2], [12], [16], [23] weighted least change symmetric secants, Schubert's (1970) least change sparse secant [21] and Marwil's (1977) [18], Toint's (1977) [24] least change symmetric sparse secant as easy consequence

In section 5 we discuss the application of least change secant updates to various nonlinear problems, pointing out which update properties (i.e., choices of $\mathcal{A}$ in (1.4)) and which norms have proven important for different types of problems, and giving guidance to update selection for future problems.

The norms which appear useful in (1.4) are the Frobenius norm,

$$||M||_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (M^{ij})^2\right)^{1/2}, \; M \in L(R^n),$$

and the weighted Frobenius norm

$$\|W_1 M W_2\|_F, \; W_1, W_2 \in L(R^n) \text{ nonsingular.}$$

We will also use the ℓ_2 vector norm (the Frobenius norm of an $n \times 1$ matrix) and the induced matrix norm. Vector inner products will be denoted by $u^T v$ and $\langle u,v\rangle$. We also use $\langle A,B\rangle =$ trace $(A^T B)$ to denote the inner product of two matrices when stacked by columns and viewed as "long" n^2 vectors. We denote the i^{th} unit vector by ε^i.

2. Deriving the least change secant update

We saw in section 1 that we are interested in updates to matrices in $L(R^n)$ which change the current Jacobian approximation as little as possible while obeying a secant equation and perhaps some other properties. We will write the general secant equation for A_+ as

$$A_+ s = y, \; s,y \in R^n, \; s \neq 0$$

and define the set of matrix quotients of y by s by

$$Q(y,s) = \{M \in L(R^n) \mid Ms = y\}. \tag{2.1}$$

(Recall that often $s = x_+ - x$ and $y = F(x_+) - F(x)$). The simplest least change secant update is the one which minimizes $\|A_+ - A\|_F$ subject to $A_+ \in Q(y,s)$. We give a straightforward derivation of this update in Theorem 2.2, using Lemma 2.1 which gives the analytic solution to the easy constrained optimization problem we encounter. We show how to find the least change secant update in a weighted Frobenius norm in Corollary 2.3.

It is possible to derive the updates of section 4 by solving the appropriate constrained optimization problems (1.4) directly, using the techniques of pseudoinverses for a special type of linearly constrained problem (see [11,23]). However, the technique of section 3 are simpler and probably more easily applicable to new situations.

Lemma 2.1: Let $\alpha \in \mathbb{R}$, $v \in \mathbb{R}^n$, $v \neq \underline{0}$. Then the unique solution t

$$\min_{x \in R^n} ||x||_2 \text{ subject to } v^T x = \alpha$$

is

$$x = \alpha v/\langle v,v\rangle.$$

Proof: If $\alpha = 0$, the lemma is trivially true. If $\alpha \neq 0$, then x must equal $\alpha w/\langle v,w\rangle$ for some $w \in R^n$, $v^T w \neq 0$. Thus $||x||_2{}^2 = \alpha^2 ||w||^2/\langle v,w\rangle^2$, which is greater than or equal to $\alpha^2/\langle v,v\rangle$ by the Cauchy-Schwartz inequality, with equality if and only if w is a scalar multiple of v. Therefore $||x||_2$ is minimized when w is a nonzero multiple of v, which means $x = \alpha v/\langle v,v\rangle$. It is interesting to note that a less elementary proof follows from the fact that $v/\langle v,v\rangle$ is the Moore-Penrose pseudoinverse [20] of v^T

The next theorem characterizes the projector onto the affine set Q(y,s).

Theorem 2.2: Let $A \in L(R^n)$, $s,y \in R^n$, $s \neq 0$, Q(y,s) defined by (2.1). Then the unique solution to

$$\min_{A_+ \in Q(y,s)} \|A_+ - A\|_F \tag{2.2}$$

is

$$A_+ = A + \frac{(y - As)s^T}{\langle s,s \rangle} . \tag{2.3}$$

<u>Proof</u>: Define $C = A_+ - A$, c_i = row i of C. Then (2.2) can be rewritten

$$\min_{C \in L(R^n)} \sum_{i=1}^{n} \|c_i\|_2^2 \text{ subject to } s^T c_i = (y - As)^i,\ i=1,\dots,n.$$

This can be broken up into n disjoint problems

$$\min_{c_i \in R^n} \|c_i\|_2 \text{ subject to } s^T c_i = (y - As)^i, \tag{2.4}$$

$i=1,\dots,n$. By Lemma 2.1, the solution to (2.4) is $c_i = (y - As)^i s^T / s^T s$. Thus the solution to (2.2) is (2.3).

Update (2.3) was introduced by Broyden [1]. It has been the most successful update for approximating the Jacobian when there are no special features of $J(x)$ which A should reflect. Successful updates for various problems are discussed more fully in section 5. There we will see that least change updates in weighted Frobenius norms are sometimes more appropriate, and so we now derive the weighted least change secant update.

<u>Corollary 2.3</u>: Let $A, s, y, Q(y,s)$ be defined as in Theorem 2.2, and let $W, \bar{W} \in L(R^n)$ be non-singular. Then the unique solution to

$$\min_{A_+ \in Q(y,s)} ||\bar{W}(A_+ - A)W||_F \tag{2.5}$$

is

$$A_+ = A + \frac{(y - As)v^T}{\langle v,s\rangle}, \quad v = W^{-T}W^{-1}s \tag{2.6}$$

<u>Proof</u>: Define $B = \bar{W}AW$, $B_+ = \bar{W}A_+W$. Then (2.5) can be rewritten

$$\min ||B_+ - B||_F \text{ subject to } B_+ \in Q(\bar{W}y, W^{-1}s). \tag{2.7}$$

By Theorem 2.2, the unique solution to (2.7) is

$$B_+ = B + \frac{(\bar{W}y - BW^{-1}s)(W^{-1}s)^T}{\langle W^{-1}s, W^{-1}s\rangle}. \tag{2.8}$$

Substituting $B = \bar{W}AW$, $B_+ = \bar{W}A_+W$ into (2.8), pre-multiplying by $\bar{W}^{-1}$ and post-multiplying by W^{-1} gives (2.6).

3. <u>Incorporating additional properties using iterated projections</u>

In regions where the Jacobian matrix always has some special property, such as symmetry or a known sparsity pattern, it seems reasonable that the Jacobian approximation can be improved by incorporating this property. Since there are many instances of problem (1.1) with symmetric and/or sparse Jacobians, we are interested in least change secant updates which choose from among the matrices possessing such properties. Below we will assume that such matrices form an affine set $\mathcal{Q}$ as they do in the cases mentioned.

Suppose we have $A \in \mathcal{Q}$ which approximates $J(x)$, and want the weighted least change secant update A_+ which is also in $\mathcal{Q}$. We might proceed as follows: find the appropriately weighted least change secant update U_1 to A; then find the nearest U_2 to U_1 such that $U_2 \in \mathcal{A}$; then the weighted least change secant update U_3 to U_2; then the nearest U_4 to U_3 such that $U_4 \in \mathcal{A}$; ... (see Figure 3.1).

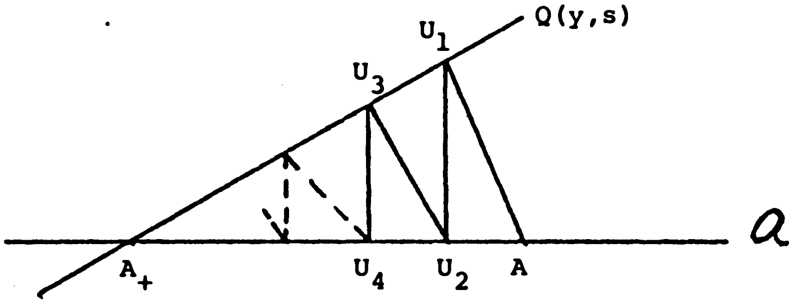

Fig. 3.1

What we are doing is projecting A onto the closest (in an appropriate norm) $U_1 \in Q(y,s)$; then U_1 onto the closest $U_2 \in \mathcal{A}$; then U_2 onto the closest U_3 in $Q(y,s)$,... . We might hope that this two step iterative projection process of elements of $\mathcal{A}$ onto $Q(y,s)$ and back to $\mathcal{A}$ has a limit A_+, and that $A_+ \in \mathcal{A}$ is the weighted least change update to A which is as close as possible to $Q(y,s)$.

In Theorem 3.1 we show algebraically that the limit of the iterated least change Frobenius norm projections from an A in an affine set of approximants $\mathcal{A}$ onto an affine subspace Q, then back to $\mathcal{A}$ is indeed the least change Frobenius norm projection of A onto $Q \cap \mathcal{A}$ or the subset of $\mathcal{A}$ nearest Q. Using this we derive in Theorem 3.2 a closed form for the least change secant update remaining in a general affine set $\mathcal{A}$, in terms of the least change projection operator onto $\mathcal{A}$. This is a powerful theorem which allows for the easy derivation in section 4 of weighted least change secant updates that retain desirable properties which define an affine subset of $L(\mathbb{R}^n)$.

The idea of iterated projections is not new, having been introduced in this context by Powell [19] for the special case of derivin a symmetric secant update from the least change secant update, and generalized by Dennis [7] to show how other symmetric updates could be derived from weighted least change secant dates. In fact, Cheney and Goldstein [4] is an earlier more general reference concerned with convex sets.

<u>Theorem 3.1</u>: Let A_1 and A_2 be affine subsets of $\mathbb{R}^m$ and let P_1, P_2 be their respective projections. If $x \in A_1$ then $\lim_{i\to\infty} (P_1P_2)^i x = x_+$ exists and is the nearest point in A_1 to x for which the distance from that point to A_2 is the distance from A_1 to A_2. Furthermore, P_2x_+ is the nearest point to x from the set of nearest points in A_2 to A_1.

The proof is just an algebraic realization of the idea of figure 3.1. The new parts of the proof are the two statements about x_+, P_2x_+ being nearest x. Cheney and Goldstein [4] had shown that the limit exists and x_+ is a nearest point in A_1 to A_2.

There is no real complication introduced by allowing A_1 and A_2 to be disjoint. We will prove a lemma which makes this clear by showing that the sets of points in A_1 nearest A_2 and vice versa are just a nonreflexive generalization of the intersection of A_1 and A_2.

As usual we take $||A_1 - A_2|| = \min_{u\in A_1, v\in A_2} ||u - v||$.

<u>Lemma</u>: Let $A_1^2 = \{x \in A_1 : ||A_1 - A_2|| = ||x - A_2||\}$ and $A_2^1 = \{y \in A_2 : ||A_1 - A_2|| = ||A_1 - y||\}$. These are parallel affine sets and if $(u,v) \in A_1^2 \times A_2^1$ with $||u - v|| = ||A_1 - A_2||$ then

$$A_1^2 = A_2^1 + (u - v) = A_1 \cap [A_2 + (u - v)].$$

<u>Proof</u>: Note that the lemma remains true for any $(u,v) \in A_1^2 \times A_2^1$, but that we do not require this generality. Let $(u',v') \in A_1^2 \times A_2^1$ with $||u' - v'|| = ||A_1 - A_2||$, and let $0 \le t \le 1$. Then since A_1, A_2 are convex

$$\begin{aligned} ||A_1 - A_2|| &\le ||tu' + (1 - t)u - [tv' + (1 - t)v]|| \\ &= ||t(u' - v') + (1 - t)(u - v)|| \\ &\le t||u' - v'|| + (1 - t)||u - v|| \\ &= ||A_1 - A_2||. \end{aligned}$$

But since the norm is strictly convex, the triangle inequality can only be an equality if $u' - v' = u - v$. Thus A_1^2 and A_2^1 are parallel affine sets and $A_1^2 = A_2^1 + (u - v)$. Also $A_1^2 = A_1 \cap A_1^2 = A_1 \cap [A_2^1 + (u - v)] \subset A_1 \cap [A_2 + (u - v)]$. If $x \in A_1 \cap [A_2 + u - v]$ then $x \in A_1$ and for some $y \in A_2$, $x = y + u - v$. Thus $x - y = u - v$ so $||x - y|| = ||u - v|| = ||A_1 - A_2||$. Therefore $x \in A_1^2$ and so $A_1^2 \supset A_1 \cap [A_2 + (u - v)]$.

<u>Proof of Theorem 3.1</u>: Let u,v) be defined by the statement of the preceeding lemma. Set $S_1 = A_1 - u$, $S_2 = A_2 - v$, $S_1^2 = A_1^2 - u$ and $S_2^1 = A_2^1 - v$. These are four subspaces with the relationships $S_2^1 = S_1^2 = S_1 \cap S_2$ that follow easily from the lemma.

Let Q_1, Q_2 be the projectors onto S_1 and S_2 and let Q, $Q^{\perp}$ respectively be the projectors onto $S_1 \cap S_2$ and $(S_1 \cap S_2)^{\perp}$. We know from standard results (Rao-Mitra Chapter 5, [20]) that $QQ_1 = QQ_2 = Q$ and $Q_1 = Q + Q_1Q^{\perp}$, $Q_2 = Q + Q_2Q^{\perp}$. Hence for any $s \in S_1$, $(Q_1Q_2)s = Q_1(Qs + Q_2Q^{\perp}s) = Q^2s + Q_1Q^{\perp}Qs + QQ_2Q^{\perp}s + Q_1Q^{\perp}Q_2Q^{\perp}s = Qs + Q_1Q^{\perp}Q_2Q^{\perp}s$, and similarly for any $i > 0$, $s^i = (Q_1Q_2)^i s = Qs + (Q_1Q^{\perp}Q_2Q^{\perp})^i s$. Since $S_1 \cap S_2 \cap (S_1 \cap S_2)^{\perp} = \{0\}$, $\lim_{i\to\infty} (Q_1Q^{\perp}Q_2Q^{\perp})^i = 0$, and $\lim_{i\to\infty} s^i =$

Now, let $x \in A_1$ and set $s = x - u$. Let P_1, P_2 and P be the orthogonal projectors on A_1, A_2 and A_1^2 respectively. Since $A_1 = S_1 + u$, $A_2 = S_2 + v$, and $A_1^2 = S_1^2 + u$, we have $P_1z = Q_1(z - u) + u$, $P_2z = Q_2(z - v) + v$, and $Pz = Q(z - u) + u$ for any $z \in R^n$. Thus, since $v - u \in S_1^{\perp} \cap S_2^{\perp}$, we have $P_2x = Q_2(x - v) + v = Q_2(x - u - (v - u)) + v = Q_2s + v$, and $P_1P_2x = P_1(Q_2s + v) = Q_1(Q_2s + v - u) + u = Q_1Q_2s + u$. Similarly, for $x^i = (P_1P_2)^i x$, $x^i = (Q_1Q_2)^i s + u = s^i + u$. Therefore $\lim_{i\to\infty} Qs + u = Q(x - u) + u = P$ That P_2x_+ is the projection of x onto A_2^1 follows in the same way since it is just $Qs + v$.

The next theorem shows that in the case $A_2 = Q(y,s)$ we can reduce the problem of finding x_+ to a particular linear least squares problem. Later, we will show how to solve this problem explicitly for some important special cases of A_1.

<u>Theorem 3.2</u>: Let $s, y \in \mathbb{R}^n$, $s \neq 0$ with $Q(y,s)$ defined by (2.1) and let $\mathcal{A}$ be an affine subspace of $\mathbb{R}^{n\times n}$ with S its parallel subspace. If $P_{\mathcal{A}}$ and P_S are the orthogonal projections into $\mathcal{A}$ and S respecti then for $M \in \mathbb{R}^{n\times n}$ and $A \in \mathcal{A}$, $P_{\mathcal{A}}(M) = A + P_S(M-A)$. Let $\mathcal{P}$ be the $n\times n$ matrix whose j^{th} column is $P_S(\frac{\varepsilon_j s^T}{s^T s})s$ and let $A \in \mathcal{A}$. If v is any solution to

$$\min_{v\in\mathbb{R}^n} \|\mathcal{P}v - (y - As)\|_2 \tag{3.1}$$

or equivalently to

$$\min_{v \in \mathbb{R}^n} \| P_S(\frac{vs^T}{s^Ts})s - (y - As) \|_2 \qquad (3.2)$$

then

$$A_+ = A + P_S(\frac{vs^T}{s^Ts}) \qquad (3.3)$$

is the nearest to A of all the nearest points of $\mathcal{A}$ to Q(y,s). If the minimum is zero then $A_+ \in \mathcal{A} \cap Q(y,s)$.

<u>Proof</u>: The relation between $P_{\mathcal{A}}$ and P_S is the one between P_2 and Q_2 we used in the last proof. It is straightforward and geometrically obvious so we omit the proof. It is stated here as a formal reminder for someone reading only the statements of the theorems. We will also use the other standard properties of a projector. It follows directly from Theorem 2.2 and this identity that if $A \in \mathcal{A}$, then

$$P_{\mathcal{A}} P_Q(A) = A + P_S(\frac{(y-As)s^T}{s^Ts}).$$

This establishes the i = 1 case of the induction hypothesis that for some $v_i \in \mathbb{R}^n$,

$$(P_{\mathcal{A}} P_Q)^i(A) = A + P_S(\frac{v_i s^T}{s^Ts}) \in \mathcal{A}$$

Thus,

$$(P_{\mathcal{A}}P_Q)^{i+1}(A) = (P_{\mathcal{A}}P_Q)[(P_{\mathcal{A}}P_Q)^i(A)] = P_{\mathcal{A}}P_Q[A + P_S(\frac{v_i s^T}{s^T s})]$$

$$= A + P_S(\frac{v_i s^T}{s^T s}) + P_S[\frac{(y - (A + P_S(\frac{v_i s^T}{s^T s}))s)s^T}{s^T s}]$$

$$= A + P_S([v_i + (y - As - P_S(\frac{v_i s^T}{s^T s})s)]\frac{s^T}{s^T s})$$

$$\triangleq A + P_S(\frac{v_{i+1}s^T}{s^T s}) = P_{\mathcal{A}}(A + \frac{v_{i+1}s^T}{s^T s}) \in \mathcal{A}.$$

Now $P_S(\frac{v_i s^T}{s^T s}) = P_S(\sum_{j=1}^{n} \frac{v_i^j \varepsilon_j s^T}{s^T s}) = \sum_{j=1}^{n} v_i^j P_S(\frac{\varepsilon_j s^T}{s^T s})$. Thus $\{(P_{\mathcal{A}}P_Q)^i($ is a sequence in A + {the linear span of $P_S(\frac{\varepsilon_j s^T}{s^T s})$, $j = 1,\ldots,n\}$ and this is a closed set. Since by Theorem 3.1 the sequence converges, it must converge to a point $A_* = A + \sum_{j=1}^{n} v_*^j P_S(\frac{\varepsilon_j s^T}{s^T s}) = A + P_S(\frac{v_* s^T}{s^T s})$. Remember also that A_* is the nearest point in $\mathcal{A}$ to A which is also a nearest point to Q(y,s).

Thus we have the minimum distance from $\mathcal{A}$ to Q(y,s) is the same as the minimum distance to Q(y,s) from the subset of $\mathcal{A}$ of matrices of the form $A + P_S(\frac{ws^T}{s^T s})$. But we can solve this simpler problem by computing that distance as

$$\min_{w \in \mathbb{R}^n} \|P_Q(A + P_S(\frac{ws^T}{s^T s})) - (A + P_S(\frac{ws^T}{s^T s}))\|_F$$

$$= \min_{w \in \mathbb{R}^n} \|\frac{(y - As - P_S(\frac{ws^T}{s^T s})s)s^T}{s^T s}\|_F$$

$$= \min_{w \in \mathbb{R}^n} \frac{\|y - As - P_S(\frac{ws^T}{s^T s})s\|_2}{\|s\|}.$$

The minimum obviously corresponds to the minimum of the numerator which is

$$\min_{w\in\mathbb{R}^n} ||(y - As) - P_S(\frac{\sum_{j=1}^{n}(w_j\varepsilon_j)s^T}{s^Ts})s||_2$$

$$= \min_{w\in\mathbb{R}^n} ||(y - As) - (P_S(\frac{\varepsilon_1 s^T}{s^Ts})s|\ldots|P_S(\frac{\varepsilon_n s^T}{s^Ts})s)w||_2.$$

This establishes that every solution of the linear least squares problem (3.1) or (3.2) put into (3.3) yields a nearest point from $\mathcal{A}$ to $Q(y,s)$ and that v^* must itself solve (3.1) since A^* is a nearest point. It also shows $Q(y,s) \cap \mathcal{A} \neq \{\}$ if the minimum is 0. Now it remains only to show that $P_S(\frac{v_1 s^T}{s^Ts}) = P_S(\frac{v_2 s^T}{s^Ts})$ for any pair v_1, v_2 of solutions to (3.1) or (3.2).

Direct computation or properties of the trace yield the useful identity $\langle uv^T, M\rangle = u^TMv$ where the matrix inner product is the sums of the inner products of respective columns. Now let v_1, v_2 solve (3.1). Then $\mathcal{P}v_1 = \mathcal{P}v_2$ so $\mathcal{P}(v_1 - v_2) = 0 = P_S(\frac{(v_1 - v_2)s^T}{s^Ts})s$. But, since projectors are self-adjoint and idempotent,

$$\left\|P_S\left(\frac{v_1 s^T}{s^T s}\right) - P_S\left(\frac{v_2 s^T}{s^T s}\right)\right\|_F^2 = \left\|P_S\left(\frac{(v_1 - v_2)s^T}{s^T s}\right)\right\|_F^2$$

$$= \left\langle P_S\left(\frac{(v_1 - v_2)s^T}{s^T s}\right), P_S\left(\frac{(v_1 - v_2)s^T}{s^T s}\right)\right\rangle$$

$$= \left\langle \frac{(v_1 - v_2)s^T}{s^T s}, P_S^2\left(\frac{(v_1 - v_2)s^T}{s^T s}\right)\right\rangle$$

$$= \left\langle \frac{(v_1 - v_2)s^T}{s^T s}, P_S\left(\frac{(v_1 - v_2)s^T}{s^T s}\right)\right\rangle$$

$$= \left(\frac{v_1 - v_2}{s^T s}\right)^T P_S\left(\frac{(v_1 - v_2)s^T}{s^T s}\right)s = (v_1 - v_2)^T 0$$

so $P_S\left(\frac{v_1 s^T}{s^T s}\right) = P_S\left(\frac{v_2 s^T}{s^T s}\right)$ and the proof is complete.

Theorems 3.1 and 3.2 remain true when stated for least change updates in a weighted Frobenius norm. However, we do not require this generality here.

4. Deriving restricted least change secant updates

Using Theorem 3.2 we can now derive the least change updates remaining in specific affine sets of approximants such as the spaces of symmetric and/or sparse matrices. In each case we first require a lemma giving the least change Frobenius norm projection operator P onto S, the subspace parallel to $\mathcal{a}$. These P's turn out to be easy to find. The application of Theorem 3.2 then requires only solving the easier of the equivalent linear least squares problems (3.1), (3.2). The respective weighted least change updates follow as easy corollaries in a manner similar to Corollary 2.3.

We begin by deriving the least change symmetric secant update due to Powell (1970), [19].

Lemma 4.1: Let S_1 be the subspace of symmetric matrices in $L(R^n)$. Then the unique solution to

$$\min_{M_+ \in L(R^n)} ||M_+ - M||_F \text{ subject to } M_+ \in S_1$$

is

$$M_+ = (M + M^T)/2 \triangleq P_1(M). \tag{4.1}$$

Proof: Let M^{ij}, M_+^{ij} be the elements of M, M_+ respectively. Then for any $M_+ \in S_1$, $M_+^{ij} = M_+^{ji}$, so that

$$||M_+ - M||_F^2 = \sum_{i=1}^{n} (M_+^{ii} - M^{ii})^2 + \sum_{1 \le i < j \le n} [(M_+^{ij} - M^{ij})^2 + (M_+^{ij} - M^{ji})^2]. \tag{4.2}$$

Simple calculus shows that the minimum of (4.2) occurs when $M_+^{ii} = M^{ii}$, $i=1,\ldots,n$, and $M_+^{ij} = (M^{ij} + M^{ji})/2$, $1 \le i \le j \le n$, so that $M_+ = (M + M^T)/2$.

The next theorem derives the Powell symmetric Broyden method [19] as a least change update.

Theorem 4.2: Let $s, y \in R^n$, $s \neq 0$, $Q(y,s)$ defined by (2.1). Let S_1 be the subspace of symmetric matrices in $L(R^n)$, and let $A \in S_1$. Then the unique solution to

$$\min_{A_+ \in L(R^n)} ||A_+ - A||_F \text{ subject to } A_+ \in Q(y,s) \cap S_1$$

is

$$A_+ = A + \frac{r_A s^T + s r_A{}^T}{\langle s,s\rangle} - \frac{\langle s,r_A\rangle ss^T}{\langle s,s\rangle^2} \tag{4.3}$$

where $r_A = y - As$.

Proof: Let P_1 be given by (4.1). We seek $v \in \mathbb{R}^n$ for which (3.1) is solved. Since $\mathcal{A} = S_1$ is itself a subspace, $P_1 = P_S$ and

$$\begin{aligned}\mathcal{P} &= \left(\frac{\varepsilon_1 s^T + s\varepsilon_1^T}{2s^Ts}s \Big| \ldots \Big| \frac{\varepsilon_n s^T + s\varepsilon_n^T}{2s^Ts}s\right)\\ &= \frac{1}{2s^Ts}(\varepsilon_1 s^Ts + s\varepsilon_1^Ts|\ldots|\varepsilon_n s^Ts + s\varepsilon_n^Ts)\\ &= \frac{1}{2}(I + \frac{ss^T}{s^Ts}).\end{aligned}$$

Thus $\mathcal{P}$ has full rank and $\mathcal{P}^{-1} = \frac{2}{s^Ts} I - \frac{ss^T}{(s^Ts)^2}$ so $v_1 = \mathcal{P}^{-1}r_A = \frac{2r_A}{s^Ts} - \frac{ss^Tr_A}{(s^Ts)^2}$ solves (3.1) uniquely. Hence, from (3.3) $A_+ = A + P_1(\overset{v}{-}$ which, using (4.1) yields (4.3).

Corollary 4.3: Let s,y, Q(y,s), S_1, and A be defined as in Theorem 4.2 and let $W \in L(R^n)$ be symmetric and nonsingular. Then the unique solution to

$$\min_{A_+ \in L(R^n)} ||W(A_+ - A)W||_F \text{ subject to } A_+ \in Q(y,s) \cap S_1 \tag{4.4}$$

is

$$A_+ + A + \frac{r_A v^T + v r_A{}^T}{\langle v,s\rangle} - \frac{\langle s,r_A\rangle vv^T}{\langle v,s\rangle^2},\ v = W^{-2}s. \tag{4.5}$$

Proof: Define $B = WAW$, $B_+ = WA_+W$. Since W is symmetric, we have $B \in S_1$, and $B_+ \in S_1$ if and only if $A_+ \in S_1$. Therefore (4.4) is equivalent to

$$\min_{B_+ \in L(R^n)} \|B_+ - B\|_F \text{ subject to } B_+ \in Q(Wy, W^{-1}s) \cap S_1$$

which by Theorem 4.2 has the solution

$$B_+ = B + \frac{(Wy - BW^{-1}s)(W^{-1}s)^T + (W^{-1}s)(Wy - BW^{-1}s)^T}{\langle W^{-1}s, W^{-1}s\rangle} - \frac{\langle W^{-1}s, Wy - BW^{-1}s\rangle (W^{-1}s)(W^{-1}s)^T}{\langle W^{-1}s, W^{-1}s\rangle^2} . \tag{4.6}$$

Substituting $B = WAW$, $B_+ = WA_+W$ into (4.6) and then pre and post-multiplying by W^{-1} yields (4.5).

Update (4.3) was first derived by Powell [19], using a special case of the method of iterated projections to symmetrize Broyden's update (2.3). It is thus refered to as the Powell-symmetric-Broyden (PSB) update. It has been less successful than a specific case of update (4.5), the Davidon-Fletcher-Powell update [5,13], which will be discussed in section 5.

We now derive the least change update which preserves a specific sparsity pattern. The method was independently discovered by Schubert [21] and Broyden [3] and analyzed by Marwil [18].

Lemma 4.4: Let $Y \in L(R^n)$ be a matrix each of whose entries is 0 or 1 and define the subspace S_2 of matrices in $L(R^n)$ with zero pattern Y as

$S_2 = \{M \in L(\mathbb{R}^n) \mid M^{ij} = 0 \text{ for all } 1 \le i,j \le n \text{ such that } Y^{ij} = 0\}$.

Let Z be the operator

$$Z: L(\mathbb{R}^n) \to L(\mathbb{R}^n), \quad (Z(M))^{ij} = \begin{cases} 0 & Y^{ij} = 0 \\ M^{ij} & Y^{ij} = 1 \end{cases}. \tag{4.7}$$

Then the unique solution to

$$\min \|M_+ - M\|_F \text{ subject to } M_+ \in S_2$$

is $M_+ = Z(M)$.

<u>Proof</u>: Define

$I_0 = \{(i,j) \mid 1 \le i,j \le n,\ Y^{ij} = 0\}$, $I_1 = \{(i,j) \mid 1 \le i,j \le n,\ Y^{ij} = 1\}$.

Then for any $M_+ \in S_2$, $M_+^{ij} = 0$ for all $(i,j) \in I_0$, so that

$$\|M_+ - M\|_F^2 = \sum_{(i,j)\in I_0} (M^{ij})^2 + \sum_{(i,j)\in I_1} (M_+^{ij} - M^{ij})^2. \tag{4.8}$$

The unique minimum to (4.8) occurs when $M_+^{ij} = M^{ij}$ for all $(i,j) \in I_1$, so that $M_+ = Z(M)$.

Below we will use pseudo-inverse notation: $a^+ = a^{-1}$ if $a \ne 0$ and $0^+ = 0$. This should not be confused with the subscript +.

<u>Theorem 4.5</u>: Let $s, y \in \mathbb{R}^n$, $s \ne 0$ and $Q(y,s)$ be defined by (2.1). Let Y and S_2 be defined as in Lemma 4.4. Let $\underline{s_i}$ be the vector formed from s by setting its j^{th} component to 0 if $Y^{ij} = 0$, and let $D^+ = \operatorname{diag}((\underline{s_1}^T s)^+, \ldots, (\underline{s_n}^T s)^+)$. The unique solution to

$$\min_{A_+ \in L(\mathbb{R}^n)} \|A_+ - A\|_F \text{ for } A_+ \in S_2 \text{ nearest } Q(y,s) \tag{4.9}$$

is
$$A_+ = A + Z(D^+(y - As)s^T)$$

$$A_+ = A + \sum_{i=1}^{n} (s_{\underline{i}}^T s)^+ \varepsilon_i^T (y - As)\varepsilon_i s_{\underline{i}}^T. \tag{4.10}$$

Proof: The proof follows simply from Theorem 3.2 and Lemma 4.4. The j^{th} column of $\mathcal{P}$ is $Z(\frac{\varepsilon_j s^T}{s^T s})s = \frac{s_{\underline{j}}^T s}{s^T s}\varepsilon_j$ and so

$\mathcal{P} = \frac{1}{s^T s}\,\mathrm{diag}(s_{\underline{1}}^T s, s_{\underline{2}}^T s, \ldots, s_{\underline{n}}^T s) \triangleq \frac{1}{s^T s} D.$ Thus,

$v = s^T s \sum_{j=1}^{n} (s_{\underline{j}}^T s)^+ \varepsilon_j^T (y - As)\varepsilon_j = (s^T s)D^+(y - As)$ is a least squares solution to (3.1) and $A_+ = A + Z(\frac{vs^T}{s^T s}) = A + Z(\sum_{j=1}^{n} (s_{\underline{j}}^T s)^+ \varepsilon_j^T (y - As)\varepsilon_j s^T)$ from whence (4.10) is immediate.

There are some useful things to note about (4.10). The crucial point is that although it involves a rank n correction to A, this is of no computational significance. The formula is given in this form because it suggests doing the update one row at a time and because it makes clear that the amount of work necessary is proportional to the number of nonzeros.

The other point to be noted concerns the possibility that $S_2 \cap Q(y,x) = \{\}$. If $s_{\underline{j}}^T s \neq 0$ for every $1 \le j \le n$ then $\mathcal{P}$ has full rank and so A_+ given by (4.10) is in $S_2 \cap Q(y,s)$. The intersection is also nonempty in the important case when $y = F(x + s) - F(x)$ for a continuously differentiable F with Y chosen to reflect the sparsity of F' over a convex set containing $[x,x+s]$ since then

$$y = [\int_0^1 F'(x + ts)\,dt]s = Ms \tag{4.11}$$

and $M \in Q(y,s) \cap S_2$. Of course, by Theorem 3.2 (4.10) is in the

Next we derive the least change update which preserves symmetr and a specific sparsity pattern. The method was independently derived by Marwil [18] and Toint [24].

Lemma 4.6: Let $Y_S \in L(\mathbb{R}^n)$ be a symmetric 0 - 1 matrix and define S_3 as the subspace of symmetric matrices in $L(\mathbb{R}^n)$ with zeros in all the positions where Y_S is zero. If the operator Z is defined by (4.8) then for $M \in L(\mathbb{R}^n)$ the unique solution to

$$\min ||M_+ - M||_F \text{ for } M_+ \in S_3$$

is

$$M_+ = \frac{1}{2}Z(M + M^T) \triangleq P_3(M).$$

Proof: The proof is a straightforward combination of the proofs of Lemmas 4.1 and 4.4.

Theorem 4.7: Let $s,y \in R^n$, $s \neq 0$, $Q(y,s)$ be defined by (2.1). Let Y_S, S_3, P_3, and Z be defined as in the statement of Lemma 4.6 and s $D = \text{diag}(\langle s_{\underline{1}},s_{\underline{1}}\rangle,\ldots,\langle s_{\underline{n}},s_{\underline{n}}\rangle)$ using the notation defined in Theorem 4.5. If $A \in Y_S$, and v is any least squares solution to

$$\frac{1}{2s^Ts}(D + Z(ss^T))v = y - As, \quad (4.1$$

then

$$A_+ = A + P_3(\frac{vs^T}{s^Ts}) \quad (4.1$$

solves

$$\min_{A_+} ||A_+ - A||_F \text{ among all } A_+ \in S_3$$

for which the distance from A_+ to $Q(y,s)$ is equal to the distance from S_3 to $Q(y,s)$. If $s_{\underline{i}}^Ts \neq 0 \neq Y_S^{ii}$ for every i from 1 and n, then $D + Z(ss^T)$ is symmetric and positive definite and (4.13) defines

Proof: All that is really required here in light of so many similar proofs, is to show that (4.12) is just the current specific incidence of (3.1). Thus we need to show that the j^{th} column of $\frac{1}{2}[D + Z(ss^T)]$ is $P_3(\varepsilon_j s^T)s$. The former is $[\varepsilon_j s_{\underline{j}}^T s_{\underline{j}} + s_{\underline{j}} s^j]\frac{1}{2}$ since the sparseness structure of Y_S is symmetric and the latter is

$$P_3(\varepsilon_j s^T)s = \frac{1}{2}Z(\varepsilon_j s^T + s\varepsilon_j^T)s$$
$$= \frac{1}{2}(\varepsilon_j s_{\underline{j}}^T + s_{\underline{j}}\varepsilon_j^T)s = \frac{1}{2}(\varepsilon_j s_{\underline{j}}^T s_{\underline{j}} + s_{\underline{j}} s^j).$$

It is very easy to show that $D + Z(ss^T) \triangleq G$ is positive definite.

Let $I_1 \triangleq \{(i,j) : 1 \le i < j \le n \text{ and } Y_S^{ij} = 1\}$ and for each $(i,j) \in I_1$ define the vector s_{ij} to be 0 in all its components except the i^{th} and j^{th} which are s^j and s^i respectively. Thus, since the diagonal of Y_S in all 1 and $Y_S = Y_S^T$, G can be rewritten as

$$G = 2\mathrm{diag}((s^1)^2,\ldots,(s^n)^2) + \sum_{(i,j)\in I_1} s_{ij}s_{ij}^T.$$
$$= \mathrm{diag}(\alpha_1,\ldots,\alpha_n) + \sum_{\substack{(i,j\in I_1)\\ s^i,s^j\neq 0}} s_{ij}s_{ij}^T$$

where $\alpha_i = 2(s^i)^2$ if $s^i \neq 0$ and $s_{\underline{i}}^T s_{\underline{i}}$ otherwise. G is obviously symmetric and since no $s_{\underline{i}}^T s_{\underline{i}}$ is 0, G is positive definite by the interleaving eigenvalue theorem (see e.g. [25]); furthermore the smallest eigenvalue of G is at least $\min_{1\le i\le n} (s^i)^2$ (and double then if no $s^i = 0$).

Note that if we define $I_0 = \{(i,j) : 1 \le i < j \le n \text{ and } Y_S^{ij} = 0\}$ then G can also be written as

$$G = (\langle s_j s\rangle I + ss^T) - \sum_{(i,j)\in I_0} s_{ij}s_{ij}^T,$$

where s_{ij} is defined as above. Thus by the interleaving eigenvalue theorem, the largest eigenvalue of G is less than or equal to the largest eigenvalue of $\langle s,s\rangle I + ss^T$, which is $2\langle s,s\rangle$, and so the ℓ_2 condition number of G is less than or equal to $2\langle s,s\rangle / \min_{1\le i\le n} (s^i)^2$ (or half this bound if no $s^i = 0$).

Update (4.13) was derived by Toint [24], who solved (4.12) by directly considering the constrained optimization problem, and by Marwil [18], who used the symmetrization process on update (4.10) (without showing that this necessarily led to a least change update). Toint also demonstrated that G is positive definite under the stronger assumption that no $s^i = 0$. We feel that our proofs are significantly more simple and may aid in the construction of an efficient algorithm to compute $G^{-1}r_A$ for the sparse symmetric problem.

There has been considerable interest, but no known success, in deriving least change symmetric sparse secant updates in general weighted Frobenius norms. The difficulty stems from the fact that for general symmetric $W \in L(R^n)$, WAW does not have the same zero pattern as A, even if W does. Thus the techniques of Corollaries 2.3 and 3.5 do not seem to apply, and the weighted least change projection operator onto the subspace of matrices with specified zero pattern is hard to find. However, weighting by a diagonal matrix does preserve the zero pattern, and so the diagonally weighted least change sparse secant updates, both symmetric and non-symmetric, follow as easy corollaries to Theorems 4.1 and 4.7.

<u>Corollary 4.8</u>: Let s, y, Q(y,s), Y_S, S_2 and Z be defined as in Theorem 4.5; let W_D and $W_D' \in L(R^n)$ be diagonal and non-singular; let $v = W_D^{-2}s$; and let $D_W = \mathrm{diag}(\langle s_{\underline{1}},v\rangle,\ldots,\langle s_{\underline{n}},v\rangle)$ using the notation

defined in Lemma 4.5. The unique solution to

$$\min_{A_+ \in L(R^n)} \|W_D'(A_+ - A)W_D\|_F \text{ subject to } A_+ \in S_2 \text{ nearest } Q(y,s)$$

is

$$A_+ = A + Z(D_W^+ r_A v^T).$$

If $s_{\underline{i}} \neq 0$ for every $i, 1 \le i \le n$ then $D_W^+ = D_W^{-1}$. If in addition Y is symmetric and S_3 is defined as in Theorem 4.7, then the unique solution to

$$\min_{A_+ \in L(R^n)} \|W_D(A_+ - A) W_D\|_F \text{ subject to } A_+ \in S_3 \text{ nearest } Q(y,s)$$

is

$$A_+ = A + Z(G_W^+ r_A v^T + v r_A^T G_W^+).$$

If also Y^{ii} is not 0 for every $i, 1 \le i \le n$ then G_W is positive definite and symmetric and $G_W^+ = G_W^{-1}$.

Proof: Very similar to the proofs of Corollaries 2.3 and 4.3. D_W is nonsingular because it equals $\text{diag}(\langle t_{\underline{1}}, t_{\underline{1}}\rangle, \ldots, \langle t_{\underline{n}}, t_{\underline{n}}\rangle)$, $t \triangleq W_D^{-1}s$, and $t_{\underline{i}} = W_D^{-1}s_{\underline{i}} \neq 0$, $i=1,\ldots,n$. To show that G_W is nonsingular, express it as $G_W = W_D^{-1}\hat{G}_W W_D$. Then $\hat{G}_W = \text{diag}(\langle t_{\underline{1}}, t_{\underline{1}}\rangle, \ldots, \langle t_{\underline{n}}, t_{\underline{n}}\rangle) + Z(tt^T)$, which is positive definite by the identical proof as for G in Theorem 4.7.

The assumptions made here and elsewhere about nonzero components of Y and s are sufficient to ensure that G is invertible and hence that a sparse and symmetric secant matrix exists but they are quite unsatisfying and so we are pleased that our proof characterize a best update without them. As in the unsymmetric case, (4.11) with $F = \nabla f$, $F' = \nabla^2 f$ is a much more satisfactory way of ascertaining the existence of a zero minimum for (3.1).

Of course, if one actually intended to solve $Gv = r_A$ in order to find A_+ defined by (4.13) then the property of positive definiteness of G would be useful indeed. We wonder if there is any computational or theoretical advantage possessed by (4.13) over $A_1 = A + \frac{1}{2}(C + C^T)$ where C is the sparse correction given in (4.10). In short, if the limit of a particular iterated projection sequence is expensive, perhaps a useful correction could be found by stopping short of the limit. In this case, one might stop after one projectic into the sparse secant matrices and back into the sparse symmetric or

5. Applications of least change secant updates

In practice, least change secant updates seem to be the most successful ones to use when approximating the Jacobian matrix (1.2) in the solution of a system of nonlinear equations (1.1). Furthermore, it seems that it is advantageous to incorporate any special structure of the Jacobian matrix into the Jacobian approximations. For example, in solving the nonlinear optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \quad f: \mathbb{R}^n \to \mathbb{R} \tag{5.1}$$

one attempts to solve the system of nonlinear equations

$$\text{find } x^* \in \mathbb{R}^n \text{ such that } \nabla f(x^*) = 0,\ \nabla f: \mathbb{R}^n \to \mathbb{R}^n. \tag{5.2}$$

The Jacobian matrix for (5.2) is the Hessian matrix of f, $\nabla^2 f(x)$, which is always symmetric (for a twice continuously differentiable f), and so one uses least change symmetric secant updates for this problem.

However, two issues remain in the choice of Jacobian updates. One is when the choice of a weighted Frobenius norm is appropriate. The second is when to favor least change inverse-secant updates, which make the smallest possible change from A^{-1} to A_+^{-1} consistent with the properties required of A_+, reasoning that it is A^{-1} and not A which is used in the calculations of the quasi-Newton step $-A^{-1}F(x)$. Computational experience seems to offer some guidance as to how these two issues should be resolved.

In the standard nonlinear equations problem (1.1), when the Jacobian has no special structure, the plain least change secant update (2.3) seems to work best. It is easy to derive the least change inverse-secant update, the A_+^{-1} which minimizes $||A_+^{-1} - A^{-1}||_F$ subject to $A_+^{-1}y = s$. By direct application of Theorem 2.2,

$$A_+^{-1} = A^{-1} + \frac{(s - A^{-1}y)y^T}{\langle y,y \rangle}$$

so long as A is nonsingular. However, this update, also introduced by Broyden [1], does not work as well as the least change secant update (2.3). Various weighted least change secant updates of form (2.6) have also been tried, but no results clearly superior to the unweighted update have been reported. Thus $||(A_+ - A)||_F$ seems to be the best measure of change for the unstructured nonlinear equation problem.

The situation in solving the unconstrained optimization problem (5.1) appears strongly different. Algorithms for solving this problem use local quadratic approximations to $f(x)$, and thus a transformation of the problem which makes these approximations nicely behaved is desirable. In particular, if the Hessian matrix at the solution, $\nabla^2 f(x^*)$, is positive definite, the transformation $\hat{x} = \nabla^2 f(x^*)^{1/2} x$, which yields a variable space for which the contour curves of the quadratic approximation around x^* are circular, is ideal. This transformation, which for the Hessian matrices corresponds to weighting by $\nabla^2 f(x^*)^{-1/2}$ on either side, can be considered the natura scaling of problem (5.1), and thus it seems desirable that this weighting be used in measuring change in Hessian approximations.

However, solving for the symmetric secant update which minimizes $||\nabla^2 f(x^*)^{-1/2}(A_+ - A)\nabla^2 f(x^*)^{-1/2}||_F$ is impossible, because we do not know x^*. The next best thing seems to be to replace $\nabla^2 f(x^*)$ by some matrix $\bar{A}$ from the feasible set of updates $Q(y,s) \cap S_1$ (S_1 is the subspace of symmetric matrices in $L(R^n)$), because this space is likely to contain our best approximation to $\nabla^2 f(x^*)$ thus far. This means using a weighting matrix $W = \bar{A}^{-1/2}$ in (4.4). While this is possible only if $\bar{A}$ is positive definite, we explain below that this is a valid assumption. Then using Corollary 3.5, we see that the solution to

$$\min_{A_+ \in L(R^n)} ||\bar{A}^{-1/2}(A_+ - A)\bar{A}^{-1/2}||_F \text{ subject to } A_+, \bar{A} \in Q(y,s) \cap S_1, \quad (5.3$$
$$\bar{A} \text{ positive definite}$$

is

$$A_+ = A + \frac{(y - As)y^T + y(y - As)^T}{\langle y,s\rangle} - \frac{\langle s,y - As\rangle yy^T}{\langle y,s\rangle^2}. \qquad (5.4)$$

Update (5.4), first introduced by Davidon [5] and clarified by Fletcher and Powell [13], is known as the Davidon-Fletcher-Powell (DFP) update.

The DFP update was the most successful update for problem (5.1) for a number of years. Another reason besides invariance with respect to linear transformations of the variables which may explain its apparent superiority to the PSB (4.3) is that if $\langle y,s\rangle > 0$, then A_+ is positive definite as long as A is. It seems desirable that each Hessian approximation A to $\nabla^2 f(x)$ be positive definite, because then the local quadratic approximation to f(x) has a unique minimum, and small steps in the quasi-Newton direction $-A^{-1}\nabla f(x)$ are guaranteed to decrease f. Therefore the assumption in (5.3) of a positive definite $\bar{A}$ is warranted, and in practice the DFP update is used to generate a series of positive definite Hessian approximations.

Although the DFP update works reasonably well, since 1970 we have learned that an apparently superior update for unconstrained optimization comes from choosing the symmetric secant update A_+ which minimizes $(A_+^{-1} - A^{-1})$ in the proper weighted Frobenius norm.

Since the ideal weighting of the Hessian approximation is by $\nabla^2 f(x^*)$ the ideal weighting of the inverse-Hessian approximation is by $\nabla^2 f(x^*)^{1/2}$, and so reasoning as above we solve

$$\min_{A_+ \in L(R^n)} ||\bar{A}^{1/2}(A_+^{-1} - A^{-1})\bar{A}^{1/2}||_F \text{ subject to } A_+, \bar{A} \in Q(y,s) \cap S_1, \ \bar{A} \text{ positive definite} \tag{5.5}$$

where A is assumed nonsingular. Straightforward application of Corollary 4.3 shows that the solution to (5.5) is

$$A_+^{-1} = A^{-1} + \frac{(s - A^{-1}y)s^T + s(s - A^{-1}y)}{\langle s,y\rangle} - \frac{\langle y, s - A^{-1}y\rangle ss^T}{\langle s,s\rangle^2} . \tag{5.6}$$

Update (5.6) was introduced by Broyden [2], Fletcher [12], Goldfarb [16] and Shanno [22], and is referred to as the BFGS update. It too has the property that A_+ is positive definite if A is positive de- finte and $\langle y,s\rangle > 0$.

Thus in making secant updates to Hessian approximations, the best measure of change seems to be $W(A_1^{-1} - A^{-1})W$, where W is some symmetric matrix which transforms the variable space x into a new space $\hat{x} = Wx$ for which the local quadratic approximation to the problem near the solution is well-behaved. While such advice may seem ad-hoc, it is supported by computational experience, and probably bears consideration when constructing updates for similar types of problems

Interesting new applications arise in cases where the Jacobian matrix has two components $J_1(x) + J_2(x)$, only one of which need be approximated (the other being calculable). The affine space $\mathcal{Q}$ is the calculated component plus the approximators of the other. The experience of Dennis, Gay, and Welsch [9] on the nonlinear least squares problem seems to show that in the construction of the required approximations A, it is beneficial to take as full advantage of the structure of the matrix being approximated as is possible. This is true in the selection of the space of quotients $Q(y,s)$, and in the inclusion of additional properties, such as symmetry. In these problems the quantity A^{-1} may have no meaning, and so selection of an update which minimizes some norm of $(A_+ - A)$ may be preferable. In an optimization-related problem (such as nonlinear least squares), selection of a weighted norm which corresponds to making the model function well shaped seems appropriate.

Richard McCord and Michael Heath of Stanford University observed to us that in taking full advantage of structure to make the Jacobian matrix approximation better, one might preclude the significant arithmetic savings in solving for the quasi-Newton step that comes from keeping certain of the updates in factored form [15]. The rationality of this observation is completely clear but we illustrate with simple examples.

Generally if we can compute the entire Jacobian matrix, it is worthwhile to do so even though the Newton step will cost $0(n^3)$ arithmetic operations rather than the $0(n^2)$ necessary for the quasi-Newton step when the Broyden sequence $\{A_\ell\}$ is carried along as $\{Q_\ell R_\ell\}$.

The reason we choose Newton's method is that we generally make enough fewer iterations to overcome the handicap of extra work per iteration. If, on the other hand, we can compute one or two elements of the Jacobian matrix, we would probably be better off to ignore this rather than to incur the order of magnitude increase in work per iteration likely to result from straightforward application of the techniques given here. If we could compute all but one or two elements of the Jacobian, the reverse would almost certainly be true. The reader will, of course, recognize all this as the same sort of decision one makes in deciding whether or not to use sparse methods for a specific linear system. As in that case, the correct answer is problem dependent and identifying classes of problems and types of structure for which answers can be given is an interesting and fruitful research area.

We have not included in Section 3 derivations of the updates of Davidon [6] and Gay and Schnabel [14] which have the additional property of preserving past secant information in A_+ (i.e., including in $\mathcal{A}$ past quotient spaces $Q(\mathcal{Y},\mathcal{S})$, or related spaces). These are interesting updates, although it is not yet clear whether they represent an improvement over the corresponding updates discussed above, the BFGS and Broyden's update. They are easily derived using the techniques of section 3. In general, we hope that the techniques of sections 2, 3, and 4 will prove helpful in deriving least change updates for problems with various kinds of structure, and that the advice of section 5 will aid in the choice of which least change updates to use.

References

[1] C.G. Broyden. (1965) "A class of methods for solving nonlinear simultaneous equations", Math. Comp. 19, 577-593.

[2] C.G. Broyden. (1971) "The convergence of a class of double-rank minimization algorithms", Parts I and II, J. Inst. Math. Appl. 6, 76-90, 222-236.

[3] C.G. Broyden. (1971) "The convergence of an algorithm for solving sparse nonlinear systems", Math. Comp. 25, 285-294.

[4] E.W. Cheney, A.A. Goldstein. (1959) "Proximity maps for convex sets", Proceedings of the American Mathematical Society, 10, 448-450.

[5] W.C. Davidon. (1959) "Variable metric method for minimization", Argonne Nat. Labs. Report ANL-5990 Rev.

[6] W.C. Davidon. (1975) "Optimally conditioned algorithms without line searches", Math. Prog. 9, 1-30.

[7] J.E. Dennis Jr. (1972) "On some methods based on Broyden's secant approximation to the Hessian" in Numerical Methods for Non-linear Optimization, edited by F.A. Lootsma, Academic Press, London.

[8] J.E. Dennis Jr., R.A. Tapia. (1976) "Supplementary terminology for nonlinear iterative methods", SIGNUM Newsletter 11, No. 4, 4-6. ?

[9] J.E. Dennis, D.M. Gay and R.E. Welsch. (1977) "An adaptive nonlinear least-squares algorithm", Cornell Computer Science Technical Report TR77-321, (submitted for publication).

[10] J.E. Dennis and J.J. Moré. (1977) "Quasi-Newton methods, motivation and theory", SIAM Review 19, 46-89.

[11] J.E. Dennis Jr. and R.B. Schnabel. In preparation.

[12] R. Fletcher. (1970) "A new approach to variable metric algorithms", Comput. J. 13, 317-322.

[13] R. Fletcher and M.J.D. Powell. (1963) "A rapidly convergent descent method for minimization", Comput. J. 6, 163-168.

[14] D.M. Gay and R.B. Schnabel. (1977) "Solving systems of nonlinear equations by Broyden's method with projected updates", to appear in Proceedings of Nonlinear Programming III, Madison, Wisconsin.

[15] R.E. Gill, G. Golub, W. Murray, M.A. Saunders. (1974) "Methods for modifying matrix factorizations", Math. Comp. 28, 505-536.

[16] D. Goldfarb. (1970) "A family of variable-metric methods derived by variational means", Math. Comp. 24, 23-26.

[17] J. Greenstadt. (1970) "Variations on variable-metric methods", Math. Comp. 24, 1-18.

[18] E.S. Marwil. (1978) "Exploiting sparsity in Newton-like methoc Ph.D. Thesis, Cornell University.

[19] M.J.D. Powell. (1970) "A new algorithm for unconstrained optimization", in Nonlinear Programming, edited by J.B. Rosen, O.L. Mangasarian, K. Ritter, Academic Press, New York.

[20] R.C. Rao and S.K. Mitra. (1971) Generalized Inverse of Matrice and Its Applications, Wiley, New York.

[21] L.K. Schubert. (1970) "Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian", Math. Comp. 24, 27-30.

[22] D.F. Shanno. (1970) "Conditioning of quasi-Newton methods for function minimization", Math. Comp. 24, 647-656.

[23] R.P. Tewarson. (1970) "On the use of the generalized inverses in function minimization", Computing 6, 241-248.

[24] Ph. L. Toint. (1977) "On sparse and symmetric matrix updating subject to a linear equation", Math. Comp. 31, 954-961.

[25] J.H. Wilkinson. (1965) The Algebraic Eigenvalue Problem, Oxfor University Press, London.