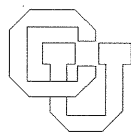


Quasi-Newton Methods Using Multiple Secant Quotations

Robert B. Schnabel*

CU-CS-247-83 June 1983



University of Colorado at Boulder

DEPARTMENT OF COMPUTER SCIENCE

*This research was supported by ARO contract DAAG 29-81-K-0108

ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS PUBLICATION ARE THOSE OF THE AUTHOR(S) AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE AGENCIES NAMED IN THE ACKNOWLEDGMENTS SECTION.

Abstract

We investigate quasi-Newton methods for unconstrained optimization and systems of nonlinear equations where each approximation to the Hessian or Jacobian matrix obeys several secant equations. For systems of nonlinear equations, this work is just a simplification and generalization of previous work by Barnes [1965] and Gay and Schnabel [1978]. For unconstrained optimization, the desire that the Hessian approximation obey more than one secant equation may be inconsistent with the requirement that it be symmetric. We present very simple necessary and sufficient conditions for there to exist symmetric, or symmetric and positive definite, updates that obey multiple secant equations. If these conditions are satisfied, one can derive generalizations of all the standard symmetric updates, including the PSB, DFP, and BFGS, that satisfy multiple secant equations. We show how to successfully specify multiple secant equations for unconstrained optimization, and that algorithms using these secant equations and the generalized PSB, DFP, or BFGS updates are locally and q -superlinearly convergent under standard assumptions.

1. Introduction

Secant approximations to finite dimensional matrices are used in many computational algorithms. These approximations are matrices $A_+ \in R^{m \times n}$ that satisfy a *secant equation*

$$A_+ s = y$$

for some $y \in R^m$ and $s \in R^n$. The most common applications, reviewed briefly below, are in solving square or rectangular systems of nonlinear equations, and in solving unconstrained and constrained optimization problems. In this paper we consider more general approximations $A_+ \in R^{m \times n}$ that satisfy *several* secant equations

$$A_+ S = Y$$

for some $S \in R^{n \times p}$ that has full column rank and $Y \in R^{m \times p}$, and the use of such approximations in solving systems of nonlinear equations and unconstrained optimization problems.

The most basic use of secant approximations is in quasi-Newton algorithms for the square systems of nonlinear equations problem,

$$\text{given } F : R^n \rightarrow R^n, \quad \text{find } x_* \in R^n \text{ such that } F(x_*) = 0.$$

These algorithms generate a sequence of iterates $\{x_k\}$, $x_k \in R^n$, $k = 0, 1, \dots$, that are increasingly good approximations to x_* . The $k+1^{\text{st}}$ iteration is based on an affine model of $F(x)$ around x_{k+1} ,

$$M_{k+1}(x) = F(x_{k+1}) + A_{k+1}(x - x_{k+1}) \quad (1.1)$$

where $A_{k+1} \in R^{n \times n}$ is a *secant approximation* to $F'(x_{k+1})$ that obeys the secant equation

$$A_{k+1} s_k = y_k \quad (1.2a)$$

where

$$s_k = x_{k+1} - x_k, \quad y_k = F(x_{k+1}) - F(x_k). \quad (1.2b)$$

Equations (1.1-2) cause $M_{k+1}(x)$ to interpolate $F(x)$ at $x = x_k$ as well as at

$x = x_{k+1}$. Many matrices $A_{k+1} \in R^{n \times n}$ satisfy (1.2); the standard way to choose A_{k+1} is to *update* the previous approximation A_k by Broyden's update

$$A_{k+1} = A_k + \frac{(y_k - A_k s_k) s_k^T}{s_k^T s_k} \tag{1.3}$$

(Broyden [1965]). This update was shown by Dennis and Moré [1977] to be the solution to

$$\underset{A \in R^{n \times n}}{\text{minimize}} \|A - A_k\|_F \quad \text{subject to } A s_k = y_k$$

where $\|\cdot\|_F$ denotes the Frobenius norm,

$$\|A\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n A[i,j]^2.$$

That is, A_{k+1} is the *least change secant update* to A_k . Broyden, Dennis, and Moré [1973] showed that the sequence of iterates generated by the quasi-Newton method

$$x_{k+1} = x_k - A_k^{-1} F(x_k)$$

with $\{A_k\}$ generated by (1.3) converges q-superlinearly to a root x_* of $F(x)$ provided x_0 and A_0 are sufficiently close to x_* and $F'(x_*)$, respectively, $F'(x_*)$ is nonsingular, and $F'(x)$ is Lipschitz continuous in an open neighborhood containing x_* . For further review of secant methods for nonlinear equations, see Dennis and Moré [1977] or Dennis and Schnabel [1983].

In Section 2 we generalize all the results stated in the last paragraph to methods where each approximation A_{k+1} in the affine model (1.1) satisfies $p \leq n$ secant equations

$$A_{k+1} S_k = Y_k \tag{1.4}$$

for $S_k, Y_k \in R^{n \times p}$. The obvious choices of S_k and Y_k are

$$S_k e_j = x_{k+1} - x_{k+1-j}, \quad Y_k e_j = F(x_{k+1}) - F(x_{k+1-j}), \quad j=1, \dots, p \tag{1.5}$$

where e_j denotes the j^{th} unit vector. If A_{k+1} satisfies (1.4-5), then the affine model (1.1) interpolates $F(x)$ at $x_{k+1-p}, \dots, x_{k+1}$. In Section 2 we give the generalization of Broyden's update that satisfies (1.4) and show that it is the least

change update satisfying these equations. We also give conditions on $\{S_k\}$ and $\{Y_k\}$ under which the quasi-Newton method using this generalized Broyden's update is locally q-superlinearly convergent. The material in Section 2 is only a modest generalization of Gay and Schnabel [1978]. It is included because the proofs are simpler and clearer, and to motivate the material in Section 3.

The other problem considered in this paper is the unconstrained minimization problem,

$$\underset{x \in R^n}{\text{minimize}} \quad f(x) : R^n \rightarrow R . \quad (1.6)$$

The first order necessary condition for x_* to be a solution of (1.6) is $\nabla f(x_*) = 0$, so (1.6) can be considered a special case of the nonlinear equations problem where $F(x) = \nabla f(x)$. While this viewpoint has limitations, it is useful in motivating secant methods for unconstrained minimization. In particular, secant methods for (1.6) base the $k+1^{\text{st}}$ iteration on a model $m(x)$ of $f(x)$ around x_{k+1} ,

$$m_{k+1}(x) = f(x_{k+1}) + \nabla f(x_{k+1})^T(x - x_{k+1}) + \frac{1}{2}(x - x_{k+1})^T H_{k+1}(x - x_{k+1})$$

where $H_{k+1} \in R^{n \times n}$ is an approximation to $\nabla^2 f(x_{k+1})$. If

$$H_{k+1}s_k = y_k \quad (1.7)$$

where

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

then $\nabla m_{k+1}(x)$ interpolates $\nabla f(x)$ at x_k and x_{k+1} . The major difference between secant methods for nonlinear equations and unconstrained minimization is that in unconstrained minimization $\nabla^2 f(x)$ is symmetric so the approximations $\{H_k\}$ should be too.

Powell [1970] introduced a symmetrized version of Broyden's update that satisfies (1.7),

$$H_{k+1} = H_k + \frac{(y_k - H_k s_k) s_k^T + s_k (y_k - H_k s_k)^T}{s_k^T s_k} + \frac{1}{2} \frac{((y_k - H_k s_k)^T s_k) s_k s_k^T}{(s_k^T s_k)^2} \quad (1.8)$$

and this update is known as the Powell symmetric Broyden (PSB) update.

Dennis and Moré showed that (1.8) is the solution to

$$\underset{H \in R^{n \times n}}{\text{minimize}} \|H - H_k\|_F \text{ subject to } H \text{ symmetric, } H s_k = y_k \quad (1.9)$$

provided that H_k is symmetric; that is, (1.8) is the least change symmetric secant update to H_k . Broyden, Dennis and Moré [1973] showed that the sequence of iterates generated by the quasi-Newton method

$$x_{k+1} = x_k - H_k^{-1} \nabla f(x_k) \quad (1.10)$$

with $\{H_k\}$ generated by (1.8) is locally q-superlinearly convergent to a minimizer x_* of $f(x)$ under appropriate assumptions.

Two other symmetric secant approximations to $\nabla^2 f(x)$, however, have been more successful in practice. They are the BFGS and DFP updates. The BFGS update, named after its proposers Broyden [1970], Fletcher [1970], Goldfarb [1970], and Shanno [1970], is

$$H_{k+1} = H_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k}. \quad (1.12)$$

The DFP update, named after its originators Davidon [1959] and Fletcher and Powell [1963], is

$$H_{k+1} = H_k + \frac{(y_k - H_k s_k) y_k^T + y_k (y_k - H_k s_k)^T}{y_k^T s_k} + \frac{1}{2} \frac{((y_k - H_k s_k)^T s_k) y_k y_k^T}{(y_k^T s_k)^2} \quad (1.13)$$

Both updates obey (1.7), and have the additional desirable property that if H_k is symmetric and positive definite and

$$y_k^T s_k > 0, \quad (1.14)$$

then H_{k+1} is well-defined, symmetric and positive definite. In practice, H_0 is chosen symmetric and positive definite and (1.14) is enforced by the line search, so each H_k is symmetric and positive definite. Dennis and Moré [1977] showed that both the BFGS and DFP updates are least change symmetric secant updates in an appropriate weighted Frobenius norm, provided that H_k is symmetric and (1.14) holds. The DFP update is the solution to

$$\underset{H \in R^{n \times n}}{\text{minimize}} \quad \|W^{-T}(H - H_k)W^{-1}\|_F \quad \text{subject to } H \text{ symmetric, } Hs_k = y_k \quad (1.15)$$

and the BFGS update is the solution to

$$\underset{H \in R^{n \times n}}{\text{minimize}} \quad \|W(H^{-1} - H_k^{-1})W^T\|_F \quad \text{subject to } H \text{ symmetric, } Hs_k = y_k \quad (1.16)$$

where in both cases $W \in R^{n \times n}$ is any nonsingular matrix for which $W^T W s_k = y_k$.

Broyden, Dennis, and Moré [1973] showed that the iterates generated by (1.10) using either the BFGS or DFP update to generate $\{H_k\}$ converge locally and q-superlinearly to a minimizer x_* of $f(x)$ under reasonable assumptions. Algorithms using the BFGS update have proven to be the most robust and efficient secant algorithms for unconstrained minimization in practice. For more information on secant methods for unconstrained minimization, see Dennis and Moré [1977], Fletcher [1980], Gill, Murray, and Wright [1981], or Dennis and Schnabel [1983].

Section 3 of this paper considers methods for unconstrained minimization where the Hessian approximation H_{k+1} is asked to satisfy $p \leq n$ secant equations

$$H_{k+1}S_k = Y_k \quad (1.17)$$

for $S_k, Y_k \in R^{n \times p}$. If S_k and Y_k are chosen in the obvious way,

$$S_k e_j = x_{k+1} - x_{k+1-j}, \quad Y_k e_j = \nabla f(x_{k+1}) - \nabla f(x_{k+1-j}), \quad j=1, \dots, p, \quad (1.18)$$

then the new quadratic model would interpolate the p most recent previous gradients, i.e.

$$\nabla m_{k+1}(x_{k+1-j}) = \nabla f(x_{k+1-j}), \quad j=0, \dots, p. \quad (1.19)$$

However, (1.19) may be inconsistent with the requirement that H_{k+1} be symmetric. Section 3.1 gives very simple necessary and sufficient conditions for there to be symmetric, or symmetric and positive definite, H_{k+1} satisfying (1.17). If these conditions are satisfied, then all the results about the PSB, BFGS, and DFP updates mentioned in the two previous paragraphs can be generalized to symmetric updates that satisfy (1.17), and to minimization algorithms that use these updates. Section 3.2 gives the generalizations of the PSB,

DFP, and BFGS updates that satisfy multiple secant equations, and shows that they are least change symmetric updates in the same norms used in (1.9), (1.15), and (1.16) respectively. Section 3.3 considers a special case of symmetric updates satisfying multiple secant equations that has received considerable attention, the "projected" updates introduced by Davidon [1975] and subsequently considered by Dennis and Schnabel [1981], Nazareth [1976], Schnabel [1977, 1978], and others. Here one assumes that H_k already satisfies $p-1$ of the p secant equations imposed upon H_{k+1} . We show that several of the projected updates derived by these authors are special cases of the generalized PSB, DFP, and BFGS updates given in section 3.2. Section 3.4 gives conditions on $\{S_k\}$ and $\{Y_k\}$ under which the iterates generated by (1.10), using the generalized PSB, DFP, or BFGS updates, converge locally and q-superlinearly to a minimizer x_* of $f(x)$. The proofs require only minor modification of the techniques of Broyden, Dennis, and Moré [1973] and Dennis and Moré [1974]. Finally in Section 3.5 we propose several ways for the preceding material on unconstrained minimization to have practical application, by suggesting several reasonable modifications of Y_k given by (1.18) that would allow symmetric (and positive definite) updates satisfying $H_{k+1}S_k = Y_k$ to exist. These modifications to Y_k do not alter the current secant equation $H_{k+1}s_k = y_k$, and alter the other secant equations in a reasonable way. The resultant algorithms obey the conditions of Section 3.4 for q-superlinear convergence.

2. Multiple secant equations for nonlinear equations

The most basic use of secant approximations is in quasi-Newton methods for solving systems of nonlinear equations. The approximation problem underlying the standard methods is to find an $A_+ \in R^{m \times n}$ for which $A_+ s = y$, where $s \in R^n$ and $y \in R^m$. As we mentioned in Section 1, the most successful practical method is based on choosing the A_+ that solves

$$\underset{A_+ \in R^{m \times n}}{\text{minimize}} \|A_+ - A\|_F \quad \text{subject to } A_+ s = y$$

where $A \in R^{m \times n}$. The generalization of this approximation problem to multiple secant equations is

$$\underset{A_+ \in R^{m \times n}}{\text{minimize}} \|A_+ - A\|_F \quad \text{subject to } A_+ S = Y \quad (2.1)$$

where $S \in R^{n \times p}$, $Y \in R^{m \times p}$. The solution to (2.1) is given in Theorem 2.1.

The remainder of this section discusses methods for solving square systems of nonlinear equations where at each iteration, the update given in Theorem 2.1 is used to calculate a Jacobian approximation $A_{k+1} \in R^{n \times n}$ that satisfies $A_{k+1} S_k = Y_k$ for some $S_k, Y_k \in R^{n \times p_k}$. A special case, considered by Barnes [1965] and Gay and Schnabel [1978], is when each update enforces the new secant equation and preserves some old secant equations satisfied by A_k . Updates with this property are sometimes called "projected secant updates". The least change projected secant update, a simple corollary of Theorem 2.1, is given in Corollary 2.2. Theorem 2.5 then gives general conditions on $\{S_k\}$ and $\{Y_k\}$ for a quasi-Newton method based on least change multiple secant updates to be q-linearly, or q-superlinearly, convergent. It uses a generalization of the Broyden, Dennis, and Moré [1973] bounded deterioration theorem that we state in Theorem 2.3, and the Dennis-Moré [1974] characterization of q-superlinear convergence that we state in Theorem 2.4. Corollary 2.6 shows that the q-superlinear result of Theorem 2.5 applies to a class of methods that enforce the

current and some past secant equations, including the method of Gay and Schnabel. This class of methods also includes some algorithms not considered by Gay and Schnabel that may be of practical interest.

Theorem 2.1. Let $p \leq n$, $A \in R^{m \times n}$, $S \in R^{n \times p}$, $Y \in R^{m \times p}$, $\text{rank}(S) = p$. Then the unique solution to (2.1) is

$$A_+ = A + (Y - A S) (S^T S)^{-1} S^T . \quad (2.2)$$

Proof: It is straightforward to derive (2.2) by regarding (2.1) as m linear least squares problems $S^T b_i = z_i$, $i=1, \dots, m$ in the variables b_i , where $b_i = \text{row } i$ of $A_+ - A$ and $z_i = \text{row } i$ of $(Y - AS)$. A different proof, given below, uses techniques of Dennis and Moré [1977] that are more closely related to the techniques we will use in Section 3.

Clearly A_+ given by (2.2) is well-defined and satisfies $A_+ S = Y$. Now let $B \in R^{m \times n}$ be any matrix satisfying $B S = Y$. Substituting BS for Y in (2.2) gives

$$A_+ - A = (B - A) S (S^T S)^{-1} S^T = (B - A) P$$

where $P = S(S^T S)^{-1} S^T$ is a Euclidean projection matrix and thus $\|P\|_2 = 1$.

Therefore

$$\|A_+ - A\|_F \leq \|B - A\|_F \|P\|_2 = \|B - A\|_F .$$

The solution is unique because (2.1) is a minimization problem in a strictly convex norm over a convex set. ■

The use of secant updates in solving systems of nonlinear equations was reviewed in Section 1. The standard secant update for nonlinear equations, Broyden's update, causes the affine model

$$M_{k+1}(x) = F(x_{k+1}) + A_{k+1} (x - x_{k+1}) \quad (2.3)$$

of $F(x)$ around x_{k+1} to interpolate $F(x)$ at x_k and x_{k+1} . An obvious use for multiple secant equations in solving systems of nonlinear equations is to cause (2.3)

to interpolate $F(x)$ at additional past iterates. For example, if $\{x_{l_{jk}}\}$ is a sequence of p_k past iterates satisfying

$$k = l_{1k} > l_{2k} > \cdots > l_{p_k k} \geq 0 , \quad (2.4)$$

and $A_{k+1}S_k = Y_k$ where

$$S_k e_j = x_{k+1} - x_{l_{jk}} , \quad Y_k e_j = F(x_{k+1}) - F(x_{l_{jk}}) , \quad (2.5)$$

then $M_{k+1}(x)$ interpolates $F(x)$ at $x_{l_{jk}}$, $j=2, \dots, p_k$ as well as at x_k and x_{k+1} .

Conditions for a method based on the above secant equations to be q -superlinearly convergent are given in Corollary 2.6. (Clearly, S_k must have full column rank to guarantee the existence of A_{k+1} .)

A special case of the above is when all but one of the function values that we ask $M_{k+1}(x)$ to interpolate already are interpolated by $M_k(x)$. Barnes [1965] and Gay and Schnabel [1978] consider a strategy that has this property. They ask the model $M_{k+1}(x)$ to interpolate $F(x)$ at p_k consecutive past iterates, as well as at x_{k+1} . In the notation of the previous paragraph, this means that $l_{jk} = k+1-j$, $j=1, \dots, p_k$. Thus

$$S_k e_j = x_{k+1} - x_{k+1-j} , \quad Y_k e_j = F(x_{k+1}) - F(x_{k+1-j}) . \quad (2.6)$$

Due to the linearity of the model (2.3), it is equivalent to define

$$\hat{S}_k e_j = x_{k+2-j} - x_{k+1-j} , \quad \hat{Y}_k e_j = F(x_{k+2-j}) - F(x_{k+1-j}) . \quad (2.7)$$

Barnes and Gay and Schnabel also assume that $p_k \leq p_{k-1}+1$, meaning that any previous function values that $M_{k+1}(x)$ should interpolate already are interpolated by $M_k(x)$. If the secant conditions are defined by (2.7), this implies that

$$(\hat{Y}_k - A_k \hat{S}_k) e_j = 0 , \quad j=2, \dots, p_k$$

so that

$$\hat{Y}_k - A_k \hat{S}_k = (\hat{Y}_k - A_k \hat{S}_k) e_1 e_1^T = (y_k - A_k s_k) e_1^T$$

where $A_{k+1}s_k = y_k$ is the current secant equation, i.e.

$$s_k = x_{k+1} - x_k , \quad y_k = F(x_{k+1}) - F(x_k) .$$

If the secant equations are defined by (2.6), then it is easy to show that

$$(Y_k - A_k S_k) e_j = (y_k - A_k s_k) , \quad j=1, \dots, p_k$$

so that

$$Y_k - A_k S_k = (y_k - A_k s_k) (1, 1, \dots, 1)^T .$$

In either case $(Y - A S)$ is a rank one matrix. Corollary 2.2 shows that the least change multiple secant update is a rank one update in this case.

Corollary 2.2. Let the conditions of Theorem 2.1 be satisfied, and let $Y - AS = (y - As) v^T$ where $v \in R^n$ is nonzero. Then the unique solution to (2.1) is

$$A_+ = A + (y - As) w^T$$

where

$$w = S (S^T S)^{-1} v .$$

Proof: immediate from Theorem 2.1.

If $v = e_1$ as in the methods of Barnes and Gay and Schnabel, then it is straightforward to show that w is a multiple of the Euclidean projection of the first column of S onto the linear subspace orthogonal to the remaining columns of S . The term "projected secant update" comes from this relationship.

A local method based on the multiple secant updates discussed above is to select each x_{k+1} to be the root of $M_k(x)$,

$$x_{k+1} = x_k - A_k^{-1} F(x_k) , \quad (2.8)$$

then choose $S_k, Y_k \in R^{n \times p_k}$, and update A_k to

$$A_{k+1} = A_k + (Y_k - A_k S_k) (S_k^T S_k)^{-1} S_k^T . \quad (2.9)$$

Theorem 2.5 gives necessary conditions on $\{S_k\}$ and $\{Y_k\}$ for the sequence of iterates generated by (2.8-9) to be locally q-linearly, or q-superlinearly, convergent to a root x_* of $F(x)$ where $F'(x_*)$ is nonsingular. The linear result is based on Theorem 2.3, a slight generalization of the bounded deterioration Theorem

3.2 in Broyden, Dennis, and Moré [1973], which differs only in that $q_k = 0$. The proof of Theorem 2.3 is omitted; see Theorem 9.2.2 of Schnabel [1977] for a proof of a slightly more general theorem. The superlinear result is based on the well known theorem of Dennis and Moré [1974] which we restate in Theorem 2.4.

In the remainder of this paper, $\|\cdot\|$ denotes the l_2 vector or matrix norm. For any $S \in R^{n \times p}$ with full column rank, $K(S)$ denotes the l_2 condition number of S , $K(S) = \|S\| \|(S^T S)^{-1} S^T\|$. For any $x \in R^n$, we define $N(x, \eta)$ to be the set $\{z \in R^n : \|z - x\| < \eta\}$.

Theorem 2.3. (Broyden, Dennis, and Moré [1973], Schnabel [1977])

Let $F : R^n \rightarrow R^n$ be continuously differentiable in an open convex set D , and assume there exists $x_* \in D$, $\eta > 0$, and $\gamma \geq 0$ satisfying $N(x_*, \eta) \subset D$, $F(x_*) = 0$, $F'(x_*)$ is nonsingular, and $\|F'(z) - F'(x)\| \leq \gamma \|z - x\|$ for all $x, z \in N(x_*, \eta)$. Consider the sequence $\{x_0, x_1, \dots\}$ of points in R^n generated by (2.8), where the sequence $\{A_0, A_1, \dots\}$ of matrices in $R^{n \times n}$ satisfies

$$\|A_{k+1} - F'(x_*)\|_F \leq \|A_k - F'(x_*)\|_F (1 + \alpha_1 \mu_k) + \alpha_2 \mu_k, \quad (2.10)$$

$$\mu_k = \max \{\|x_{k+1} - x_*\|, \|x_k - x_*\|, \dots, \|x_{k-q_k} - x_*\|\} \quad (2.11)$$

$k=0, 1, \dots$ for some fixed $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, with $q_k = \min\{k, q\}$ for some fixed $q \geq 0$.

Then for each $r \in (0, 1)$, there exist positive constants $\varepsilon(r)$, $\delta(r)$ such that if $\|x_0 - x_*\| \leq \varepsilon(r)$ and $\|A_0 - F'(x_*)\|_F \leq \delta(r)$, the sequence $\{x_0, x_1, \dots\}$ is well-defined and converges to x_* with

$$\|x_{k+1} - x_*\| \leq r \|x_k - x_*\|$$

for all k . Furthermore, $\{A_k\}$ and $\{A_k^{-1}\}$ are uniformly bounded.

Theorem 2.4. (Dennis and Moré [1974])

Let the assumptions of Theorem 2.3 hold. Let $\{A_k\}$ be a sequence of nonsingular matrices in $R^{n \times n}$, and suppose for some $x_0 \in R^n$ the sequence of points generated by (2.8) remains in D and converges to x_* , with $x_k \neq x_*$ for any k . Then

$\{x_k\}$ converges q-superlinearly to x_* if and only if

$$\lim_{k \rightarrow \infty} \frac{\|(A_k - F'(x_*))(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0 . \quad (2.12)$$

Theorem 2.5. Let the assumptions of Theorem 2.3 hold. Consider the sequences $\{x_k\}$ and $\{A_k\}$ generated from $x_0 \in R^n$ and $A_0 \in R^{n \times n}$ by (2.8-9) where $S_k, Y_k \in R^{n \times p_k}$ with each $p_k \in [1, n]$. Suppose there exist $c_1 \geq 0, c_2 \geq 1, q \geq 0$, such that for $k = 0, 1, \dots$,

$$\|Y_k - F'(x_*) S_k\|_F \leq c_1 \|S_k\| \max\{\|x_{k-i} - x_*\|\}, \quad i = -1, 0, \dots, q_k \quad (2.13)$$

and

$$K(\|S_k\|) \leq c_2 \quad (2.14)$$

where each $q_k \leq \max\{k, q\}$. Then there exist $\varepsilon \geq 0, \delta \geq 0$ such that if $\|x_0 - x_*\| \leq \varepsilon$ and $\|A_0 - F'(x_*)\| \leq \delta$, the sequence $\{x_k\}$ is well-defined and converges q-linearly to x_* , and $\{A_k\}, \{A_k^{-1}\}$ are uniformly bounded. If in addition, for each k there exists $v_k \in R^{p_k}$ for which

$$S_k v_k = x_{k+1} - x_k , \quad (2.15)$$

then the rate of convergence is q-superlinear.

Proof: Let $J_* = F'(x_*)$. From (2.9),

$$(A_{k+1} - J_*) = (A_k - J_*) (I - S_k (S_k^T S_k)^{-1} S_k^T) + (Y_k - J_* S_k) (S_k^T S_k)^{-1} S_k^T . \quad (2.16)$$

Define $P_k = S_k (S_k^T S_k)^{-1} S_k^T$, and recall that $\|P_k\|, \|I - P_k\| \leq 1$. Then using also (2.13), with μ_k defined by (2.11), in (2.16) gives

$$\begin{aligned} \|A_{k+1} - J_*\|_F &\leq \|A_k - J_*\|_F \|I - P_k\| + \|Y_k - J_* S_k\|_F \|(S_k^T S_k)^{-1} S_k^T\| \\ &\leq \|A_k - J_*\|_F + c_1 K(S_k) \mu_k . \end{aligned}$$

Therefore from (2.14), A_{k+1} satisfies (2.10) with $\alpha_1 = 0$ and $\alpha_2 = c_1 c_2$, which proves q-linear convergence.

To prove q-superlinear convergence, define $E_k = (A_k - J_*)$. Since P_k is a Euclidean projection matrix,

$$\|E_k (I - P_k)\|_F = (\|E_k\|_F^2 - \|E_k P_k\|_F^2)^{1/2} \leq \|E_k\|_F - \frac{\|E_k P_k\|_F^2}{2 \|E_k\|_F} . \quad (2.17)$$

Then from (2.16), (2.17), (2.13), and (2.14),

$$\|E_{k+1}\|_F \leq \|E_k\|_F - \frac{\|E_k P_k\|_F^2}{2 \|E_k\|_F} + c_1 c_2 \mu_k$$

which implies

$$\|E_k P_k\|_F^2 \leq 2 \|E_k\|_F (\|E_k\|_F - \|E_{k+1}\|_F + c_1 c_2 \mu_k) . \quad (2.18)$$

From the proof of linear convergence, there exist $\rho, \beta \in (0, \infty)$ such that $\sum_{i=0}^{\infty} \mu_k \leq \rho$

and $\|E_k\|_F \leq \beta$ for all k . Using these bounds and summing (2.18) from $k=0$ to j gives

$$\sum_{k=0}^j \|E_k P_k\|_F^2 \leq 2 \beta (\|E_0\|_F - \|E_{j+1}\|_F + c_1 c_2 \sum_{k=0}^j \mu_k) \leq 2 \beta (\beta + \rho)$$

which proves that

$$\lim_{k \rightarrow \infty} \|E_k P_k\|_F = 0 . \quad (2.19)$$

Finally we show that, if (2.15) is true, then (2.19) implies the Dennis-Moré condition (2.12) for superlinear convergence. Define $s_k = (x_{k+1} - x_k)$. Then from (2.15),

$$E_k P_k s_k = E_k S_k (S_k^T S_k)^{-1} S_k^T S_k v_k = E_k S_k v_k = E_k s_k ,$$

so that by the definition of an induced matrix norm

$$\|E_k P_k\| \geq \frac{\|E_k P_k s_k\|}{\|s_k\|} = \frac{\|E_k s_k\|}{\|s_k\|} ,$$

and from (2.19)

$$\lim_{k \rightarrow \infty} \frac{\|E_k s_k\|}{\|s_k\|} \leq \lim_{k \rightarrow \infty} \|E_k P_k\| = 0 .$$

Thus the method (2.8-9) satisfies condition (2.12) of Theorem 2.4 and is q -superlinearly convergent. ■

Theorem 2.5 says, roughly, that if $A_{k+1} S_k = Y_k$ are reasonable secant equations in that $F'(x_*) S_k$ is close enough to Y_k , and if the columns of S_k are sufficiently linearly independent, then the method (2.8-9) will be locally q -

linearly convergent; if in addition the most recent secant equation $A_{k+1}S_k = y_k$ always is included, the method will be q-superlinearly convergent. Corollary 2.6 shows that the choices of S_k and Y_k given by (2.4-5), which cause $M_{k+1}(x)$ to interpolate $F(x)$ at p_k not necessarily consecutive past iterates including the most recent, satisfy these criteria as long as the past iterates are chosen so that each S_k is sufficiently linearly independent, and there is some upper bound on how many iterations back the secant equations can go.

Corollary 2.6. Let the assumptions of Theorem 2.3 hold, and let $q \geq 1$ be fixed. Consider the sequences $\{x_k\}$ and $\{A_k\}$ generated from $x_0 \in R^n$ and $A_0 \in R^{n \times n}$ by (2.8-9), where for each k , $1 \leq p_k \leq \min\{k+1, n, q\}$, $S_k, Y_k \in R^{n \times p_k}$ with

$$K(S_k) \leq c$$

for some fixed $c \geq 1$, and

$$S_k e_j = x_{k+1} - x_{l_{jk}}, \quad Y_k e_j = F(x_{k+1}) - F(x_{l_{jk}}), \quad j=1, \dots, p_k$$

where

$$k \geq l_{1k} > l_{2k} > \dots > l_{p_k k} \geq \max\{0, k+1-q\} .$$

Then there exist $\varepsilon, \delta > 0$ such that if $\|x_0 - x_*\| \leq \varepsilon$ and $\|A_0 - F'(x_*)\| \leq \delta$, the sequence $\{x_k\}$ is well-defined and converges q-linearly to x_* . Furthermore, $\{A_k\}$ and $\{A_k^{-1}\}$ are uniformly bounded. If $l_{1k} = k$ for all k , the rate of convergence is q-superlinear.

Proof: By a well known lemma (see for example Section 3.2.5 of Ortega and Rheinboldt [1970]),

$$\|(Y_k - F'(x_*)S_k) e_j\| \leq \gamma \|x_{k+1} - x_{l_{jk}}\| \cdot \max\{\|x_{k+1} - x_*\|, \|x_{l_{jk}} - x_*\|\} \leq \gamma \|S_k e_j\| \mu_k ,$$

where the last inequality uses only the definitions of S_k , and of μ_k from (2.11).

Thus

$$\|Y_k - F'(x_*) S_k\|_F \leq \gamma \|S_k\|_F \mu_k \leq \sqrt{n} \gamma \|S_k\| \mu_k ,$$

so (2.13) is satisfied and q-linear convergence is established by Theorem 2.5. If

$l_{1k}=k$ for all k , then q -superlinear convergence follows trivially since (2.15) is true with $v_k=e_1$ for all k . ■

The strategies covered by Corollary 2.6 for choosing the past iterates whose function values the model will interpolate include the strategy implemented by Gay and Schnabel [1978], as well as the strategy used by Schnabel and Frank [1983] in their "tensor method" for nonlinear equations. Schnabel and Frank always select $S_k e_1 = (x_{k+1} - x_k)$. Then they consider, in order, the steps from x_{k+1} to $x_{k-1}, \dots, x_{k+1-q}$; they include $x_{k+1} - x_{k-i}$ as a column of S_k if and only if it makes an angle of more than 45° with the linear subspace spanned by the already selected columns of S_k . Their experience is that the best results are obtained using only information from fairly recent past iterates; they restrict p_k , and q , to be at most \sqrt{n} . This strategy allows considerably more flexibility in choosing past iterates than the strategy tested by Gay and Schnabel; it would be interesting to test a secant algorithm for nonlinear equations that uses it.

3. Multiple secant equations for unconstrained optimization

Now we turn to the unconstrained minimization problem (1.6), which we reviewed briefly in Section 1. The standard quadratic model of the objective function,

$$m_{k+1}(x) = f(x_{k+1}) + \nabla f(x_{k+1})^T(x - x_{k+1}) + \frac{1}{2}(x - x_{k+1})^T H_{k+1}(x - x_{k+1}) \quad (3.1)$$

could interpolate several past gradient values if the symmetric approximation to the Hessian H_{k+1} obeyed several secant equations

$$H_{k+1} S_k = Y_k \quad (3.2)$$

where $S_k, Y_k \in R^{n \times p}$ are given by (1.18). Several authors, starting with Schnabel [1977], have noted that (3.2) may be inconsistent with the symmetry of H_{k+1} . In Section 3.1 we show that there exists a symmetric, or symmetric and positive definite, H_{k+1} satisfying (3.2) if and only if $Y_k^T S_k$ is symmetric, or symmetric and positive definite, respectively. While the natural choices (1.18) of S_k and Y_k satisfy these conditions if $f(x)$ is a positive definite quadratic, for general $f(x)$ $Y_k^T S_k$ usually is not even symmetric. In Section 3.5 we attempt to remedy this difficulty by proposing several reasonable ways to perturb Y_k to a \tilde{Y}_k for which $\tilde{Y}_k^T S_k$ is symmetric and positive definite. The preceding sections, 3.2-3.4, discuss the updates and methods that may be used if the conditions for symmetric (and positive definite) multiple secant updates to exist are satisfied. Section 3.2 introduces generalizations of the PSB, DFP, and BFGS updates that satisfy (3.2) and shows that they are the least change updates in the appropriate norms. In Section 3.3 we show that several "projected secant updates" that have been proposed for unconstrained minimization are special cases of the updates discussed in Section 3.2. Section 3.4 shows that quasi-Newton methods based on our generalizations of the PSB, DFP, or BFGS updates are locally q-superlinearly convergent under standard assumptions. The methods proposed in Section 3.5 satisfy the conditions for q-superlinear convergence.

3.1. Necessary and Sufficient Conditions for Symmetric Multiple Secant Updates

Theorem 3.1. Let $p \leq n$, $S, Y \in R^{n \times p}$, $\text{rank}(S) = p$. Then there exist symmetric $H_+ \in R^{n \times n}$ such that $H_+ S = Y$ if and only if $Y^T S$ is symmetric. There exist symmetric and positive definite $H_+ \in R^{n \times n}$ such that $H_+ S = Y$ if and only if $Y^T S$ is symmetric and positive definite.

Proof : *Only if :* Suppose there exists a symmetric H_+ for which $H_+ S = Y$. Then $S^T H_+ S = Y^T S$ is symmetric. Similarly, if H_+ is symmetric and positive definite, then $S^T H_+ S = Y^T S$ is symmetric and positive definite.

If : Suppose $Y^T S$ is symmetric. Then

$$H_1 = Y(S^T S)^{-1} S^T + S(S^T S)^{-1} Y^T - S(S^T S)^{-1} (Y^T S) (S^T S)^{-1} S^T \quad (3.3)$$

is well-defined, symmetric, and obeys $H_1 S = Y$. Now suppose $Y^T S$ is symmetric and positive definite. Then

$$H_2 = Y (Y^T S)^{-1} Y^T$$

is well-defined, symmetric, obeys $H_2 S = Y$, and is at least positive semi-definite. Also $\text{rank}(Y) = p$ from $Y^T S$ nonsingular. Thus if $p = n$, H_2 is positive definite. If $p < n$, let $Z \in R^{n \times m}$ be any matrix whose columns all are in, and together span, the null space of S ; that is, $Z^T S = 0$, $m \geq n - p$, and $\text{rank}(Z) = n - p$. Then

$$H_3 = Y (Y^T S)^{-1} Y^T + Z Z^T \quad (3.4)$$

is well-defined, symmetric, obeys $H_3 S = Y$, and is at least positive semi-definite. Now let $U \in R^{n \times (n-p)}$ be an orthonormal basis for the null space of S . Then $Z = U N^T$ where $N \in R^{m \times (n-p)}$ has full column rank, i.e. $N^T N$ is nonsingular. Then from (3.4),

$$H_3 = M_1 M_2 M_1^T$$

where $M_1, M_2 \in R^{n \times n}$,

$$M_1 = \begin{bmatrix} Y & U \end{bmatrix}, \quad M_2 = \begin{bmatrix} (Y^T S)^{-1} & \\ & N^T N \end{bmatrix}.$$

Clearly M_2 is nonsingular, and since

$$M_1^T \begin{bmatrix} S & U \end{bmatrix} = \begin{bmatrix} Y^T S & Y^T U \\ 0 & I \end{bmatrix},$$

M_1 is nonsingular. Therefore H_3 is nonsingular and hence, positive definite. ■

Note that the above proof could be simplified slightly by defining $Z=U$, however the more general definition of Z will be useful to us in Section 3.2.

Now let us consider whether the conditions of Theorem 3.1 are likely to be satisfied in the context of an unconstrained minimization algorithm. Suppose, as in Section 2, that $\{x_{i_{jk}}\}$ is a sequence of past iterates satisfying (2.4) and $S_k, Y_k \in R^{n \times p_k}$ are defined by

$$S_k e_j = x_{k+1} - x_{i_{jk}}, \quad Y_k e_j = \nabla f(x_{k+1}) - \nabla f(x_{i_{jk}}), \quad j=1, \dots, p_k. \quad (3.5)$$

If $f(x)$ is quadratic, then $Y_k = \nabla^2 f(x) S_k$, so $Y_k^T S_k$ is symmetric for any $\{x_{i_{jk}}\}$, and $Y_k^T S_k$ is positive definite if $\nabla^2 f(x)$ is positive definite and S_k has full column rank. When $f(x)$ is not quadratic, however, it is unlikely that $Y_k^T S_k$ is symmetric, as illustrated by the following example.

Example 3.1. Let $x \in R^2$, $f(x) = \frac{1}{2}(x[1])^2 + \frac{1}{2}(x[2])^2 + \frac{1}{4}(x[2])^4$, and suppose some algorithm generates $x_0 = (-2, -2)$, $x_1 = (-1, -1)$, $x_2 = (-1, 0)$. If, in the notation of the preceding paragraph, $x_{i_{j1}} = x_{2-j}$, $j=1,2$, then

$$S_1 = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}, \quad Y_1 = \begin{bmatrix} 0 & 1 \\ 2 & 10 \end{bmatrix}$$

and

$$Y_1^T S_1 = \begin{bmatrix} 2 & 4 \\ 10 & 21 \end{bmatrix}.$$

Since the natural secant equations for unconstrained minimization, $H_{k+1}S_k = Y_k$ with S_k and Y_k defined by (3.5), rarely will satisfy the conditions of Theorem 3.1 when $p_k > 1$, it might seem that the topic of multiple secant equations for unconstrained minimization is fruitless. In Section 3.5, however, we will show how a practical algorithm for unconstrained minimization might generate multiple secant equations that satisfy the conditions of Theorem 3.1, without changing the current secant equation. Sections 3.2-3.4 investigate updates and methods that are possible when the conditions of Theorem 3.1 are satisfied.

3.2. Least change symmetric multiple secant updates

The reader may have noticed that the equation (3.3) used in the proof of Theorem 3.1 reduces, in the case when $p=1$, to the PSB update of $H = 0$. The corresponding update to a nonzero H would be

$$H_{PSBg} = H + (Y-HS)(S^T S)^{-1}S^T + S(S^T S)^{-1}(Y-HS)^T - S(S^T S)^{-1}(Y-HS)^T S(S^T S)^{-1}S^T. \quad (3.6)$$

Equation (3.6) is a generalization of the PSB update (1.8) to multiple secant equations, hence the name "PSBg". H_{PSBg} is well-defined and $H_{PSBg}S = Y$ as long as S has full column rank; if H is symmetric, then it is easy to see that H_{PSBg} is symmetric if and only if $Y^T S$ is symmetric. The rank of $H_{PSBg} - H$ is at most $2p$. We show in Theorem 3.2 that if $Y^T S$ and H are symmetric, then H_{PSBg} is the least change symmetric update to H , in the Frobenius norm, that satisfies $H_{+}S = Y$.

Correspondingly, the DFP update (1.13) may be generalized to

$$H_{DFPg} = H + (Y-HS)(Y^T S)^{-1}Y^T + Y(Y^T S)^{-1}(Y-HS)^T - Y(Y^T S)^{-1}(Y-HS)^T S(Y^T S)^{-1}Y^T. \quad (3.7)$$

H_{DFP_g} is well-defined and $H_{DFP_g} S = Y$ whenever $Y^T S$ is nonsingular; it is symmetric if H and $Y^T S$ are symmetric. Again, $H_{DFP_g} - H$ has rank at most $2p$. We also show in Theorem 3.2 that if H and $Y^T S$ are symmetric and positive definite, then H_{DFP_g} is the solution to

$$\underset{H_+ \in R^{n \times n}}{\text{minimize}} \quad \|W^{-T}(H_+ - H)W^{-1}\|_F \quad (3.8)$$

subject to H_+ symmetric and positive definite, $H_+ S = Y$, where $W \in R^{n \times n}$ is any nonsingular matrix that satisfies $W^T W S = Y$.

The reader also may have noticed that the matrices H_2 and H_3 used in the proof of Theorem 3.1 are related to the BFGS update. In fact, if H is symmetric and positive definite and

$$Z = H^{\frac{1}{2}} - HS (S^T HS)^{-1} S^T H^{\frac{1}{2}} \quad (3.9)$$

then the matrix H_3 given by (3.4) is

$$H_{BFGS_g} = H + Y (Y^T S)^{-1} Y^T - HS (S^T HS)^{-1} S^T H, \quad (3.10)$$

a generalization of the BFGS update (1.12). If $S^T HS$ and $Y^T S$ are nonsingular, H_{BFGS_g} is well-defined and $H_{BFGS_g} S = Y$; H_{BFGS_g} is symmetric if H and $Y^T S$ are symmetric. $H_{BFGS_g} - H$ also has rank at most $2p$. Theorem 3.2 also shows that if H and $Y^T S$ are symmetric and positive definite, then H_{BFGS_g} is the solution to

$$\underset{H_+ \in R^{n \times n}}{\text{minimize}} \quad \|W(H_+^{-1} - H^{-1})W^T\|_F \quad (3.11)$$

subject to H_+ symmetric and positive definite, $H_+ S = Y$ for any nonsingular $W \in R^{n \times n}$ that satisfies $W^T W S = Y$.

Theorem 3.2. Let $p \leq n$, $H \in R^{n \times n}$ symmetric, $S, Y \in R^{n \times p}$, $Y^T S$ symmetric, $\text{rank}(S) = p$. Then the unique solution to

$$\underset{H_+ \in R^{n \times n}}{\text{minimize}} \quad \|H_+ - H\|_F \quad \text{subject to } H_+ \text{ symmetric, } H_+ S = Y \quad (3.12)$$

is H_{PSB_g} given by (3.6). If in addition H and $Y^T S$ are positive definite, and $W \in R^{n \times n}$ is any nonsingular matrix that satisfies $W^T W S = Y$, then the unique

solutions to (3.8) and (3.11) are H_{DFP_g} given by (3.7) and H_{BFGS_g} given by (3.10), respectively.

Proof : If S has full column rank and $H, Y^T S$ are symmetric, then clearly H_{PSB_g} given by (3.6) is symmetric and satisfies $H_{PSB_g} S = Y$. Now let $H_+ \in R^{n \times n}$ be any symmetric matrix satisfying $H_+ S = Y$, and define $E_{PSB_g} = H_{PSB_g} - H, E = H_+ - A$. Then substituting $H_+ S$ for each occurrence of Y in (3.6) gives

$$E_{PSB_g} = EP + PE - PEP = EP + PE(I-P) \quad (3.13)$$

where $P = S(S^T S)^{-1} S^T$ is a Euclidean projection matrix. Recall that $\|P\| \leq 1, \|I-P\| \leq 1$, and $P(I-P) = 0$. We also use the fact that for any $M_1, M_2 \in R^{n \times n}$,

$$\begin{aligned} \|M_1 P + M_2(I-P)\|_F^2 &= \|M_1 P\|_F^2 + \|M_2(I-P)\|_F^2 + 2 \text{trace}(M_1 P(I-P) M_2^T) \\ &= \|M_1 P\|_F^2 + \|M_2(I-P)\|_F^2 . \end{aligned} \quad (3.14)$$

Thus from (3.13) and (3.14),

$$\begin{aligned} \|E_{PSB_g}\|_F^2 &= \|EP\|_F^2 + \|PE(I-P)\|_F^2 \\ &\leq \|EP\|_F^2 + \|P\|^2 \|E(I-P)\|_F^2 \\ &= \|EP\|_F^2 + \|E(I-P)\|_F^2 = \|E\|_F^2 \end{aligned}$$

with the last equality coming from another application of (3.14). This shows that (3.6) is a solution to (3.12). The solution is unique because (3.12) is a minimization problem in a strictly convex norm over a convex set.

If H and $Y^T S$ are symmetric and positive definite, it is straightforward to verify that the generalized DFP update (3.7) is (3.4) with

$$Z = H^k - Y(Y^T S)^{-1} S^T H^k .$$

Clearly $Z^T S = 0$, and since H^k is nonsingular and $Y, S \in R^{n \times p}$, Z has rank $n-p$. Thus from the proof of Theorem 3.1, H_{DFP_g} is symmetric, positive definite, and satisfies $H_{DFP_g} S = Y$. The proof that H_{DFP_g} is the solution to (3.8) then follows from applying the standard transformation of variables technique to the above proof for the generalized PSB; see for example Dennis and Schnabel [1979]. Finally, since the generalized BFGS update (3.10) is (3.4) with Z given by (3.9),

the same argument shows that it is symmetric, positive definite, and satisfies $H_{BFGSg} S = Y$. The proof that H_{BFGSg} is the solution to (3.11) also is obtained in the standard way. First the dual DFP result is obtained, that is, it is shown that the solution to (3.11) is

$$H_+^{-1} = H^{-1} + (S - H^{-1}Y)(Y^T S)^{-1} S^T + S(Y^T S)^{-1}(S - H^{-1}Y)^T \quad (3.15) \\ - S(Y^T S)^{-1}(S - H^{-1}Y)^T S(Y^T S)^{-1} S^T .$$

Then it is straightforward to show that for H_+^{-1} given by (3.15), $H_+ = H_{BFGSg}$. ■

Many algebraic properties of standard symmetric secant updates can be extended to symmetric multiple secant updates. For example, the analog of the Broyden one parameter class (Broyden [1970]) is

$$H_+(M) = H_{BFGSg} + V M V^T, \quad V = (HS(S^T HS)^{-1} - Y(Y^T S)^{-1})(S^T HS)^{\frac{1}{2}}$$

where $M \in R^{p \times p}$ is any symmetric matrix; $H_+(M)$ is positive definite if M is positive definite, and $H_{DFPg} = H_+(I)$. Also, the Cholesky factorization of H_{DFPg} or H_{BFGSg} may be obtained in $O(n^2 p)$ operations from the Cholesky factorization of H ; for example if $H = LL^T$, then $H_{BFGSg} = JJ^T$ where

$$J = L + (YG - HS)(S^T HS)^{-1} S^T L$$

for $G \in R^{p \times p}$ satisfying

$$G^T (Y^T S) G = S^T HS .$$

J can be calculated in $O(n^2 p)$ operations, and its LQ factorization can be obtained in an additional $O(n^2 p)$ operations.

3.3. Projected symmetric secant updates

Davidon [1975] proposed a quasi-Newton algorithm that finds the minimizer of a positive definite quadratic in at most $n+1$ iterations. To accomplish this, each quadratic model interpolates the gradients at current and all past iterates;

that is, for each k ,

$$\nabla m_k(x_j) = \nabla f(x_j), \quad j=0,1,\dots,k. \quad (3.16)$$

Equation (3.16) implies that

$$H_k S_{k-1} = Y_{k-1}$$

where $S_{k-1}, Y_{k-1} \in R^{n \times k}$ are defined by

$$S_{k-1} e_j = x_{k+1-j} - x_{k-j} = s_{k-j}, \quad Y_{k-1} e_j = \nabla f(x_{k+1-j}) - \nabla f(x_{k-j}) = y_{k-j}, \quad j=1,\dots,k.$$

Similarly at the next iteration, Davidon's method requires

$$H_{k+1} S_k = Y_k$$

where $S_k, Y_k \in R^{n \times (k+1)}$,

$$S_k e_j = s_{k+1-j}, \quad Y_k e_j = y_{k+1-j}, \quad j=1,\dots,k+1.$$

It is straightforward from the above definitions that

$$Y_k - H_k S_k = (y_k - H_k s_k) e_1^T, \quad (3.17)$$

that is, H_k already satisfies k of the $k+1$ secant equations imposed upon H_{k+1} .

Symmetric secant updates that satisfy $H_{k+1} S_k = Y_k$ when $f(x)$ is quadratic and (3.17) is true were investigated by Davidon [1975], and subsequently by many authors including Dennis and Schnabel [1981], Nazareth [1976], and Schnabel [1977, 1978]. They often are called "projected secant updates".

Corollary 3.3 shows that the necessary and sufficient conditions for symmetric secant updates to satisfy (3.17) for general $f(x)$ follow immediately from Theorem 3.1. If these conditions are satisfied, the updates discussed in Section 3.2 reduce to rank two updates.

Corollary 3.3. Let $p \leq n$, $H \in R^{n \times n}$ symmetric, $S, Y \in R^{n \times p}$, $\text{rank}(S) = p$, $s = S e_1$, $y = Y e_1$,

$$Y - HS = (y - Hs) e_1^T. \quad (3.18)$$

Then there exist symmetric H_+ for which $H_+ S = Y$ if and only if

$$S^T (y - Hs) = \sigma e_1 \quad (3.19)$$

where $\sigma = s^T (y - Hs)$. In this case, the generalized PSB update (3.6) is a rank

two update of H . If in addition H is positive definite, then there exist symmetric and positive definite H_+ for which $H_+S = Y$ if and only if (3.19) is satisfied and

$$1 + \sigma \tau > 0, \quad (3.20)$$

where $\tau = e_1^T (S^T H S)^{-1} e_1$. In this case, both the generalized DFP update (3.7) and the generalized BFGS update (3.10) are positive definite and rank two updates of H .

Proof : Define $t = y - Hs$. From Theorem 3.1, there exist symmetric H_+ satisfying $H_+S = Y$ if and only if $Y^T S$ is symmetric. From (3.18),

$$Y^T S = S^T H S + e_1^T (S^T t)^T.$$

Since H is symmetric, $Y^T S$ is symmetric if and only if $S^T t$ is some multiple of e_1 . Since $(S^T t)[1] = s^T t = \sigma$, this is possible if and only if (3.19) is true. If (3.19) is satisfied then the generalized PSB update is symmetric, and substituting (3.18) and (3.19) into (3.6) shows that in this case it is the rank two update

$$H_+ = H + (y - Hs) \hat{s}^T + \hat{s} (y - Hs)^T - \sigma \hat{s} \hat{s}^T$$

where $\hat{s} = S(S^T S)^{-1} e_1$.

Also from Theorem 3.1, there exist positive definite H_+ for which $H_+S = Y$ if and only if $Y^T S$ is symmetric and positive definite. If H is symmetric and positive definite and (3.19) holds, then from the above

$$Y^T S = S^T H S + \sigma e_1 e_1^T \quad (3.21)$$

is symmetric. Since $S^T H S$ is positive definite, it is easy to show from (3.21) that $Y^T S$ is positive definite if and only if (3.20) is true. In this case, substituting (3.18) and (3.19) into (3.7) gives the generalized DFP update in this case to be

$$H_+ = H + t \hat{y}^T + \hat{y} t^T - \sigma \hat{y} \hat{y}^T \quad (3.22)$$

where $\hat{y} = Y(Y^T S)^{-1} e_1$. Substituting (3.18) and (3.21) into (3.10) and using the Sherman-Morrison-Woodbury formula for the inverse of (3.21) gives the generalized BFGS update in this case to be

$$H_+ = H + \frac{\bar{y}\bar{y}^T}{\tau(1+\sigma\tau)} - \frac{H\bar{s}\bar{s}^T H}{\tau} \quad (3.23)$$

where $\bar{s} = S(S^T H S)^{-1} e_1$, $\bar{y} = Y(S^T H S)^{-1} e_1 = H \bar{s} + \tau t$. ■

The necessary and sufficient conditions for projected symmetric secant updates to interpolate several past gradients have been discussed by several authors, starting with Schnabel [1977]. As we already have indicated, they rarely are satisfied if $f(x)$ is nonquadratic, even if (3.17) is true. In our opinion, this is the fundamental reason why projected symmetric secant updates have not been an improvement over the BFGS in practice. The projected DFP update (3.22) was proposed by Schnabel [1977] and an algorithm that uses it was shown to be q-superlinearly convergent. If $f(x)$ is quadratic, (3.22) is the dual of the update originally proposed by Davidon [1975]. The projected BFGS update (3.23) is derived by Dennis and Schnabel [1981].

3.4. Superlinear convergence of quasi-Newton methods using symmetric multiple secant updates

A local method for unconstrained minimization based on the symmetric multiple secant updates discussed in Section 3.2 is to select each iterate x_{k+1} to be the critical point of the current quadratic model,

$$x_{k+1} = x_k - H_k^{-1} \nabla f(x_k) ,$$

then choose $S_k, Y_k \in R^{n \times p_k}$ such that $Y_k^T S_k$ is symmetric, and update H_k by the generalized PSB update (3.6), or if $Y_k^T S_k$ also is positive definite, by the generalized DFP (3.7) or BFGS (3.10) update. (When we refer to updates 3.6, 3.7, or 3.10 in this section, we assume that the symbols H_{PSBg} , H_{DFPg} , and H_{BFGSg} in these formulas have been converted to H_{k+1} , and that all other symbols in these formulas have been given the subscript k .) In this section we show that if $\{S_k\}$ and

$\{Y_k\}$ obey the same conditions (2.13-14) as were required for the local convergence of the multiple secant method for nonlinear equations, then any of the aforementioned methods for unconstrained minimization is locally and q-superlinearly convergent to a minimizer x_* of $f(x)$, under standard assumptions. Theorem 3.4 proves the local q-superlinear convergence of the method that uses the generalized PSB update. The proof is based on Broyden, Dennis, and Moré [1973] and Dennis and Moré [1974], and is very similar to the proof of Theorem 2.5. Theorem 3.5 states the analogous result for methods using the generalized DFP, or BFGS, updates. The proofs would follow from the proof for the PSB method. Since these proof techniques are so well established, we omit the proof of Theorem 3.5 and just make a few comments about it.

Theorem 3.4. Let $F : R^n \rightarrow R^n$ be continuously differentiable in an open convex set D , and assume there exists $x_* \in D$, $\eta > 0$, and $\gamma \geq 0$ satisfying $N(x_*, \eta) \subset D$, $F(x_*) = 0$, $F'(x_*)$ is symmetric and nonsingular, and $\|F'(z) - F'(x)\| \leq \gamma \|z - x\|$ for all $x, z \in N(x_*, \eta)$. Consider the sequences $\{x_k\}$ and $\{H_k\}$ generated from $x_0 \in R^n$ and a symmetric $H_0 \in R^{n \times n}$ by

$$x_{k+1} = x_k - H_k^{-1} F(x_k)$$

and the generalized PSB update (3.6), where $\{S_k\}, \{Y_k\} \in R^{n \times p_k}$, with each $p_k \in [1, n]$ and each $Y_k^T S_k$ symmetric. Suppose there exist $c_1 \geq 0, c_2 \geq 1, q \geq 0$, such that for $k = 0, 1, \dots$,

$$\|Y_k - F'(x_*) S_k\|_F \leq c_1 \|S_k\| \max\{\|x_{k-i} - x_*\|\}, \quad i = -1, 0, \dots, q_k \quad (3.24)$$

and

$$K(\|S_k\|) \leq c_2 \quad (3.25)$$

where each $q_k \leq \max\{k, q\}$. Then there exist $\varepsilon \geq 0, \delta \geq 0$ such that if $\|x_0 - x_*\| \leq \varepsilon$ and $\|H_0 - F'(x_*)\| \leq \delta$, the sequence $\{x_k\}$ is well-defined and converges q-linearly to x_* , and $\{H_k\}, \{H_k^{-1}\}$ are uniformly bounded. If in addition, for each k there exists $v_k \in R^{p_k}$ for which

$$S_k v_k = x_{k+1} - x_k ,$$

then the rate of convergence is q-superlinear.

Proof : Let $H_* = F'(x_*)$, $E_k = (H_k - H_*)$, $P_k = S_k(S_k^T S_k)^{-1} S_k^T$. Then from (3.6) it is straightforward to obtain

$$\begin{aligned} E_{k+1} &= E_k - E_k P_k + (Y_k - H_* S_k)(S_k^T S_k)^{-1} S_k^T - P_k E_k + S_k(S_k^T S_k)^{-1}(Y_k - H_* S_k) \\ &\quad + P_k E_k P_k - S_k(S_k^T S_k)^{-1}(Y_k - H_* S_k)^T P_k \\ &= (I - P_k)E_k(I - P_k) + (Y_k - H_* S_k)(S_k^T S_k)^{-1} S_k^T \\ &\quad + S_k(S_k^T S_k)^{-1}(Y_k - H_* S_k)^T (I - P_k) . \end{aligned} \quad (3.26)$$

Thus using $\|I - P_k\| \leq 1$, (3.24), and the definition (2.11) of μ_k ,

$$\begin{aligned} \|E_{k+1}\|_F &\leq \|E_k\|_F \|I - P_k\|^2 + \|Y_k - H_* S_k\|_F \|(S_k^T S_k)^{-1} S_k^T\| (1 + \|I - P_k\|) \\ &\leq \|E_k\|_F + 2 c_1 K(S_k) \mu_k . \end{aligned}$$

Therefore from (3.25), H_{k+1} satisfies (2.10) with $\alpha_1 = 0$ and $\alpha_2 = 2c_1 c_2$, which proves q-linear convergence from Theorem 2.3. To prove q-superlinear convergence, derive from (3.26)

$$\|E_{k+1}\|_F \leq \|E_k (I - P_k)\|_F + 2 c_1 c_2 \mu_k .$$

The remainder of the q-superlinear proof then is identical to the q-superlinear proof in Theorem 2.5. \blacksquare

Theorem 3.5. Let the assumptions of Theorem 3.4 hold, and assume in addition that $F'(x_*)$ is positive definite. Then Theorem 3.4 remains true if the generalized PSB update (3.6) is replaced by the generalized DFP update (3.7), or by the generalized BFGS update (3.10).

The convergence proof for the generalized DFP method is very similar to the proof of Theorem 3.4. The modifications required are similar to the modifications Broyden, Dennis, and Moré [1973] use to convert their proof for the PSB method into a proof for the DFP method. Bounded deterioration is proven using the weighted Frobenius norm

$$\bar{E}_k = \|H_*^{-1/2} (H_k - H_*) H_*^{-1/2}\|_F .$$

It is straightforward to show from (3.7) that

$$\bar{E}_{k+1} = (I - \bar{P}_k) \bar{E}_k (I - \bar{P}_k)^T + (\bar{Y}_k - \bar{S}_k) (\bar{Y}_k^T \bar{S}_k)^{-1} \bar{Y}_k^T + \bar{Y}_k (\bar{Y}_k^T \bar{S}_k)^{-1} (\bar{Y}_k - \bar{S}_k)^T (I - \bar{P}_k)^T$$

where

$$\bar{Y}_k = H_*^{-1/2} Y_k , \quad \bar{S}_k = H_*^{1/2} S_k , \quad \bar{P}_k = I - \bar{Y}_k (\bar{Y}_k^T \bar{S}_k)^{-1} \bar{S}_k^T ,$$

and from (3.24),

$$\|I - \bar{P}_k\| \leq 1 + O(\mu_k) , \quad \|\bar{P}_k\| \leq 1 + O(\mu_k) .$$

Linear convergence follows easily from these relations and Theorem 2.3, and q-superlinear convergence from the same techniques used in the proof of Theorem 3.4. The convergence proof for the generalized BFGS method is essentially the dual of the DFP proof, as in Broyden, Dennis, and Moré. Note that $Y_k^T S_k$ positive definite is implied by (3.24) and $F'(x_*)$ positive definite.

The crucial question is whether there exist reasonable choices of $\{S_k\}$ and $\{Y_k\}$ that satisfy the conditions of Theorems 3.4 and 3.5. The following section provides a positive answer to this question.

3.5. Forming multiple secant equations for unconstrained optimization

The obvious use of multiple secant equations in an unconstrained minimization algorithm would be to allow the quadratic model (3.1) of $f(x)$ around x_{k+1} to interpolate gradients at $p_k > 1$ past iterates $\{x_{l_j k}\}$, $j=1, \dots, p_k$, where

$$k = l_{1k} > l_{2k} > \dots > l_{p_k k} \geq 0 . \quad (3.27)$$

This would require the model Hessian H_{k+1} to satisfy p_k secant equations

$$H_{k+1} S_k = Y_k \quad (3.28)$$

where $S_k, Y_k \in R^{n \times p_k}$ are defined by (3.5). Unfortunately, Theorem 3.1 shows that (3.28) is consistent with H_{k+1} symmetric (and positive definite) only if $Y_k^T S_k$ is

symmetric (and positive definite), and Example 3.1 indicates that this is unlikely for nonquadratic $f(x)$. In this section we discuss several ways to perturb Y_k to $\tilde{Y}_k = (Y_k + \Delta Y_k)$ so that $\tilde{Y}_k^T S_k$ is symmetric (and positive definite). These methods all yield $(\Delta Y_k)e_1 = 0$, that is, the standard secant equation is unchanged, and they all generate sequences $\{S_k\}$ and $\{\tilde{Y}_k\}$ that satisfy the conditions of Theorems 3.4-5 for local q-superlinear convergence. The general aim of these methods is to perturb Y_k as little as possible consistent with $\tilde{Y}_k^T S_k$ symmetric, and to change more recent secant equations less than less recent secant equations.

For the remainder of this section, we assume that $\{S_k\}$ and $\{Y_k\}$ are defined by (3.5, 3.27), with $\{l_{jk}\}$ chosen by a procedure that guarantees $K(S_k)$ sufficiently small; a suitable procedure is described at the end of Section 2. We also drop the subscripts k for the remainder of this section. Now we describe our first strategy for calculating ΔY .

It is trivial to calculate the lower triangular matrix $L \in R^{p \times p}$ for which

$$Y^T S - S^T Y = -L + L^T . \quad (3.29)$$

Note that the diagonal of L is zero. From (3.29), $(Y^T S + L)$ is symmetric. Our first strategy is to choose ΔY such that

$$(\Delta Y)^T S = L . \quad (3.30)$$

Equation (3.30) implies that for each column $(\Delta Y)e_j$ of ΔY , only $((\Delta Y)e_j)^T (S e_i)$, $1 \leq i < j$, need be nonzero. Thus we may choose $(\Delta Y)e_1 = 0$, leaving the standard secant equation intact. This choice is guaranteed if we choose the smallest ΔY that satisfies (3.30), in the Frobenius norm. From Theorem 2.1, it is

$$\Delta Y = S (S^T S)^{-1} L^T . \quad (3.31)$$

The above choice of ΔY guarantees that $(Y + \Delta Y)^T S$ is symmetric, but not necessarily that it is positive definite. An easy modification that assures positive definiteness is to first choose a subset of the rows and columns of $(Y^T S + L)$ that

is positive definite, and restrict the past points used to this subset. This selection is easily accomplished using a modification of the Cholesky decomposition of $(Y^T S + L)$; the normal decomposition is attempted, but if the addition of the j^{th} row and column would cause the matrix not to be positive definite, then the j^{th} past point (and the j^{th} row and column of $(Y^T S + L)$) is eliminated. If the normal line search condition in a quasi-Newton method for minimization, $(Ye_1)^T (Se_1) > 0$, is satisfied, then this strategy always retains the current secant equation.

In Example 3.2 we apply this strategy to Example 3.1.

Example 3.2. Let $S, Y \in R^{2 \times 2}$ be the matrices S_1 and Y_1 from Example 3.1. Then

$$Y^T S - S^T Y = \begin{bmatrix} 0 & -6 \\ 6 & 0 \end{bmatrix}$$

so $L = \begin{bmatrix} 0 & 0 \\ -6 & 0 \end{bmatrix}$. Since $(Y^T S + L) = \begin{bmatrix} 2 & 4 \\ 4 & 21 \end{bmatrix}$ is positive definite, both past points can be retained. From (3.31)

$$\Delta Y = S (S^T S)^{-1} L^T = \begin{bmatrix} 0 & 12 \\ 0 & -6 \end{bmatrix}$$

so that

$$\tilde{Y} = \begin{bmatrix} 0 & 13 \\ 2 & 4 \end{bmatrix}.$$

It is easy to show that under the assumptions of Theorem 3.4, there exists $c > 0$ for which $\|\tilde{Y} - F'(x_*)S\| \leq c \|S\| \mu$, μ defined by (2.11). Let $H_* = F'(x_*)$. We showed in the proof of Corollary 2.6 that

$$\|Y - H_* S\|_F \leq \sqrt{n} \gamma \|S\| \mu,$$

so that

$$\|Y^T S - S^T H_* S\|_F \leq \sqrt{n} \gamma \|S\|^2 \mu.$$

Therefore

$$\begin{aligned} \|L\|_F &= (1/\sqrt{2}) \|-L + L^T\|_F = (1/\sqrt{2}) \|Y^T S - S^T Y\|_F \\ &\leq (1/\sqrt{2}) \|(Y^T S - S^T H_* S) - (S^T Y - S^T H_* S)\|_F \leq \sqrt{2n} \gamma \|S\|^2 \mu \end{aligned}$$

and

$$\|\Delta Y\|_F \leq \|S(S^T S)^{-1}\| \|L\|_F \leq \sqrt{2n} \gamma K(S) \|S\| \mu .$$

Thus

$$\|\tilde{Y} - H_* S\|_F \leq \|Y - H_* S\|_F + \|\Delta Y\|_F \leq c \|S\| \mu \quad (3.32)$$

where $c = \sqrt{n} \gamma (\sqrt{2} K(S) + 1)$. From Theorems 3.4-5, this implies that a generalized PSB, DFP, or BFGS algorithm that chooses $\{S_k\}$ and $\{Y_k\}$ to satisfy the conditions of Corollary 2.6, and modifies Y_k by (3.31), will be locally q-superlinearly convergent to a minimizer x_* where $\nabla^2 f(x_*)$ is nonsingular. Sufficiently close to x_* , (3.32) guarantees that $\tilde{Y}^T S$ will be positive definite.

When using multiple secant equations in conjunction with a generalized DFP or BFGS update, it may be more reasonable to find the smallest ΔY , in a weighted Frobenius norm, that satisfies (3.30). It is straightforward to show that, for $W \in R^{n \times n}$ nonsingular, the solution to

$$\underset{\Delta Y \in R^{n \times p}}{\text{minimize}} \|W^{-T} \Delta Y\|_F \quad \text{subject to} \quad (\Delta Y)^T S = L$$

is

$$\Delta Y = W^T W S (S^T W^T W S)^{-1} L^T .$$

If we assume that the past points have been restricted, if necessary, so that $(Y^T S + L)$ is positive definite, then a reasonable choice is W for which $W^T W S = (Y + \Delta Y)$; it is easy to show that this choice results in

$$\bar{\Delta Y} = Y (S^T Y)^{-1} L^T .$$

$\tilde{Y} = (Y + \bar{\Delta Y})$ also can be shown to satisfy the conditions of Theorems 3.4-5.

We briefly describe a second strategy for perturbing Y that may come closer to the goal of changing recent information as little as possible. It is to change each column of Y only as much as necessary to meet the symmetry requirements imposed by more recent (already revised) secant equations.

Algebraically, this means

Algorithm 3.1.

1. Set $(\Delta Y)e_1 = 0$.
2. For $j=2, \dots, p$ do
 - 2.1. Select $\delta \in R^n$ to minimize $\|\delta\|$,
 subject to $(Ye_j + \delta)^T Se_i = (Se_j)^T (Y + \Delta Y)e_i, \quad i=1, \dots, j-1$.
 - 2.2 Set $(\Delta Y)e_j = \delta$.

That is, column j of ΔY is chosen to be as small as possible subject to the j^{th} column of the $j \times j$ principal submatrix of $S^T(Y + \Delta Y)$ equaling the j^{th} row of this submatrix. The first two columns of ΔY generated by Algorithm 3.1 are the same as are those generated by (3.31); the remaining columns would, in general, be different.

$S^T(Y + \Delta Y)$ generated by Algorithm 3.1 also might not be positive definite. It is easy to modify Algorithm 3.1, however, to generate $S^T(Y + \Delta Y)$ positive definite, by generating iteratively the Cholesky factorization of the current $j \times j$ principal submatrix, and, if the j^{th} step fails to keep the submatrix positive definite, eliminating this point from the set of past points used at that iteration.

Algorithm 3.1 has a close relationship to our first strategy for choosing ΔY . From step 2.1, $\Delta Y = S\bar{L}^T$, where \bar{L} is lower triangular with zero diagonal. Thus $(Y + \Delta Y)^T S = (Y^T S + \bar{L}S^T S)$ is symmetric, so Algorithm 3.1 is equivalent to finding the unique lower triangular (with zero diagonal) \bar{L} for which

$$Y^T S - S^T Y = \bar{L}S^T S - S^T \bar{L}^T, \quad ,$$

and then choosing ΔY to solve

$$\underset{\Delta Y \in R^{n \times p}}{\text{minimize}} \|\Delta Y\|_F \quad \text{subject to} \quad \Delta Y^T S = \bar{L}S^T S.$$

$(Y + \Delta Y)$ generated by Algorithm 3.1 also obeys the conditions of Theorems 3.4-5, since it can be shown that

$$\|\Delta Y\|_F \leq \sqrt{n} \gamma (1 + K(S))^{p-1} \|S\|_\mu.$$

Since p and $K(S)$ will be small in practice, the constant in the above equation is not too large. Finally, a weighted version of Algorithm 3.1 can be obtained by changing the norm in step 2.1 to a weighted norm.

The strategies given in this section may not be the best ways to generate multiple secant equations for minimization. They do show, however, that reasonable choices of $\{S_k\}$ and $\{\tilde{Y}_k\}$ exist that satisfy both the existence conditions of Theorem 3.1 and the local q -superlinear convergence conditions of Theorem 3.4-5. Maybe they will lead to successful computational algorithms. We do think there is a significant difference between the strategies of this section and algorithms that have used projected updates such as those discussed in Section 3.3. While both approaches interpolate multiple past gradients when $f(x)$ is quadratic, the strategies of this section should come closer to interpolating past gradients for non-quadratic functions, because they do not compound the interpolation errors of previous updates. The cost is a higher rank update.

4. References

- J. Barnes [1965], "An algorithm for solving nonlinear equations based on the secant method", *Computer Journal* 8, pp. 66-72.
- C. G. Broyden [1965], "A class of methods for solving nonlinear simultaneous equations", *Mathematics of Computation* 19, pp. 577-593.
- C. G. Broyden [1970], "The convergence of a class of double-rank minimization algorithms", Parts I and II, *Journal of the Institute of Mathematics and its Applications* 6, pp. 76-90, 222-236.
- C. G. Broyden, J. E. Dennis Jr., and J. J. Moré [1973], "On the local and superlinear convergence of quasi-Newton methods", *Journal of the Institute of Mathematics and its Applications* 12, pp. 223-246.
- W. C. Davidon [1959], "Variable metric methods for minimization", Argonne National Laboratory Report ANL-5990 (rev.).
- W. C. Davidon [1975], "Optimally conditioned optimization algorithms without line searches", *Mathematical Programming* 9, pp. 1-30.
- J. E. Dennis Jr. and J. J. Moré [1974], "A characterization of superlinear convergence and its application to quasi-Newton methods", *Mathematics of Computation* 28, pp. 549-560.
- J. E. Dennis Jr. and J. J. Moré [1977], "Quasi-Newton methods, motivation and theory", *SIAM Review* 19, pp. 46-89.
- J. E. Dennis Jr. and R. B. Schnabel [1979], "Least change secant updates for quasi-Newton methods", *SIAM Review* 21, pp. 443-459.
- J. E. Dennis Jr. and R. B. Schnabel [1981], "A new derivation of symmetric positive definite secant updates", in *Nonlinear Programming 4*, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, New York, pp. 167-199.
- J. E. Dennis Jr. and R. B. Schnabel [1983], *Numerical Methods for Nonlinear Equations and Unconstrained Optimization*, Prentice-Hall, Englewood Cliffs, New Jersey.
- R. Fletcher [1970], "A new approach to variable metric methods", *Computer Journal* 13, pp. 317-322.
- R. Fletcher [1980], *Practical Method of Optimization, Vol 1, Unconstrained Optimization*, John Wiley and Sons, New York.
- R. Fletcher and M. J. D. Powell [1963], "A rapidly convergent descent method for minimization", *Computer Journal* 6, pp. 163-168.
- D. M. Gay and R. B. Schnabel [1978], "Solving systems of nonlinear equations by Broyden's method with projected updates", in *Nonlinear Programming 3*, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, New York, pp. 245-281.

P. E. Gill, W. Murray, and M. H. Wright [1981], *Practical Optimization*, Academic Press, London.

D. Goldfarb [1970], "A family of variable metric methods derived by variational means", *Mathematics of Computation* 24, pp. 23-26.

L. Nazareth [1976], "On Davidon's optimally conditioned algorithms for unconstrained optimization", Technical Memorandum 283, Applied Mathematics Division, Argonne National Laboratory.

J. M. Ortega and W. C. Rheinboldt [1970], *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.

M. J. D. Powell [1970], "A new algorithm for unconstrained optimization", in *Nonlinear Programming* J. B. Rosen, O. L. Mangasarian, and K. Ritter, eds., Academic Press, New York, pp. 31-65.

R. B. Schnabel [1977], "Analyzing and improving quasi-Newton methods for unconstrained optimization", Ph.D. thesis, Department of Computer Science, Cornell University.

R. B. Schnabel [1978], "Minimum norm symmetric quasi-Newton updates restricted to subspaces", *Mathematics of Computation* 32, pp. 829-837.

R. B. Schnabel and P. Frank [1983], "Tensor methods for nonlinear equations", Technical Report CU-CS-243-83, Department of Computer Science, University of Colorado at Boulder.

D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization", *Mathematics of Computation* 24, pp. 647-657.