Q-Superlinear Convergence of Biggs' Method and Related Methods for Unconstrained Optimization

Robert. B. Schnabel

CU-CS-133-78 August 1978

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

Abstract:

The most successful quasi-Newton methods for solving unconstrained optimization problems when second derivatives are unavailable or expensive have used the BFGS update. In recent tests, Brodlie reports that an update introduced by Biggs performs equally well. This update differs from the BFGS in that it alters the secant equation to incorporate information from a cubic model. In this paper we show that Biggs' method retains the Q-superlinear convergence properties of the BFGS exhibited by Broyden, Dennis, and Moré and by Powell. Our proofs show that near the solution of most problems, Biggs' method and the BFGS are essentially the same. We also establish necessary and sufficient conditions for the Q-superlinear convergence of a general class of quasi-Newton methods which modify the secant equation similarly to Biggs.

1. Introduction

This paper is concerned with the convergence of a class of quasi-Newton methods for solving the unconstrained minimization problem

$$\min_{x \in R^n} f: \quad R^n \to R \tag{1.1}$$

where f is assumed twice continuously differentiable. We assume that the reader is familiar with quasi-Newton methods. Recent references include Brodlie [3] and Dennis and Moré [9].

Quasi-Newton methods generate a sequence of points $x_i \in R^n$ which hopefully converge to the solution x_* of (1.1). They are related to Newton's method, in which this sequence is produced by the iteration

$$x_{i+1} = x_i - \nabla^2 f(x_i)^{-1} \nabla f(x_i). \tag{1.2}$$

In the class of quasi-Newton methods with which we are concerned, due to the unavailability or cost of obtaining $\nabla^2 f(x)$, (1.2) is modified by replacing $\nabla^2 f(x_i)^{-1}$ with an approximation H_i which is modified following each iteration. In addition, a line search parameter λ_i may be included, so that the iteration becomes

$$x_{i+1} = x_i - \lambda_i H_i \nabla f(x_i), \; \lambda_i > 0. \tag{1.3}$$

If H_i is positive definite, λ_i can be chosen so that $f(x_{i+1}) < f(x_i)$. The value $\lambda_i = 1$ is often attempted first, and is called the direct prediction value of λ_i.

The most successful quasi-Newton methods have chosen H_{i+1} by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update, [5,10,11,15],

$$H_{i+1} = H_i + \frac{(s_i - H_i y_i) s_i^T + s_i (s_i - H_i y_i)^T}{s_i^T y_i} - \frac{< s_i - H_i y_i, y_i > s_i s_i^T}{(s_i^T y_i)^2} \tag{1.4}$$

$$s_i \triangleq x_{i+1} - x_i$$
$$y_i \triangleq \nabla f(x_i) - \nabla f(x_i).$$

This is one of many updates obeying the secant equation,

$$H_{i+1} y_i = s_i \tag{1.5}$$

which is satisfied by $H_{i+1} = \nabla^2 f(x_{i+1})^{-1}$ if f is a quadratic, and otherwise is approximately true (of $\nabla^2 f(x_{i+1})^{-1}$) since

$$\left[\int_{\tau=0}^{1} \nabla^2 f(x_i + \tau s_i) \, d\tau \right] s_i = y_i .$$

Subject to (1.5), (1.4) is the update which solves

$$\min_{H_{i+1} \in R^{n \times n}} \| H_{i+1} - H_i \|_{F, \overline{H}^{-1/2}}$$
$$\text{subject to } H_{i+1} \text{ symmetric} \tag{1.6}$$

where $\overline{H}$ is any fixed matrix obeying (1.5). Here $\|A\|_{F,M}$, $A, M \in R^{n \times n}$ denotes the Frobenius norm of A weighted by M, $\|MAM\|_F$, where the Frobenius norm $\|B\|_F$ of any $B \in R^{n \times n}$ is the square root of the sum of the squares of the elements of B.

In this paper we are concerned with methods which modify the above by changing the secant equation to

$$H_{i+1} y_i = \rho_i s_i \tag{1.7}$$

for some free parameter ρ_i. This degree of freedom was first suggested by Huang [12]. Biggs [1,2] then introduced strategies for choosing ρ_i in order to incorporate information from models of higher than quadratic order. The main reason we are interested in analyzing such methods is that in recent computational tests by Brodlie [4], Biggs' [2] method performed as well as the BFGS. Therefore it is of continuing interest.

The update used by Biggs in [2] is the one which solves (1.6) subject to H_{i+1} obeying the modified secant equation (1.7). This is the variant of the BFGS,

$$H_{i+1}(\rho_i) = H_i + \frac{(\rho_i s_i - H_i y_i)s_i^T + s_i(\rho_i s_i - H_i y_i)^T}{s_i^T y_i} - \frac{\langle \rho_i s_i - H_i y_i, y_i \rangle \, s_i s_i^T}{(s_i^T y_i)^2} \quad (1.8)$$

Parameter ρ_i is chosen to incorporate information from a cubic model of f. Its choice is explained in the beginning of section 4.

Our task in this paper is to analyze the behavior of Biggs' method, and the general class of methods which modify the secant equation to (1.7). In particular, we wish to compare their theoretical behavior with that of the BFGS. The BFGS has strong theoretical properties which complement its computational performance. Broyden, Dennis and Moré [6] have shown that if λ_i is always set to 1, then the points generated by (1.3) using the BFGS update are Q-superlinearly convergent to the solution x_* of (1.1) if x_0 and H_0 are sufficiently close to x_* and $\nabla^2 f(x_*)^{-1}$ respectively, $\nabla^2 f(x_*)$ is positive definite, and some lesser conditions on f are met. (Q-superlinear convergence of $\{x_i\}$ to x_* is defined as

$$\lim_{i \to \infty} \frac{\| x_{i+1} - x_* \|}{\| x_i - x_* \|} = 0$$

for some vector norm $\| \cdot \|$). Powell [13] has shown that if λ_i is chosen by a certain realistic line search strategy, then the points generated by (1.3) using the BFGS are globally convergent to the solution of any convex function f which also satisfies some less restrictive conditions, and Q-superlinearly convergent if $\nabla^2 f(x_*)$ is also positive definite and $\lambda_i = 1$ is used whenever it is permissible.

Our main result in this paper is that Biggs' [2] method has exactly the same properties. First we establish some general necessary and sufficient conditions for the Q-superlinear convergence of quasi-Newton methods which modify the secant equation to (1.7). In section 2 we show that for any such method of form (1.3) with $\lambda_i \equiv 1$ to converge locally Q-superlinearly to x_* on the entire set of "well-behaved" functions, it is necessary that $\lim_{i \to \infty} \rho_i = 1$. Next, in section 3 we show that a slightly stronger condition,

$$\sum_{i=0}^{\infty} | \rho_i - 1 | \leq \sigma \tag{1.9}$$

is sufficient to extend the results of Broyden, Dennis and Moré [6] and Powell [13] to update (1.8). For the Broyden-Dennis-Moré result σ is some fixed constant; for the Powell result any $\sigma < +\infty$ is sufficient. In section 4 we extend the theorems of Broyden-Dennis-Moré and Powell to Biggs' method. As a consequence of the preceeding results, it is clear that in doing so we establish that Biggs' choice of ρ_i converges to 1; indeed, it satisfies (1.9).

Of course, one consequence of this analysis is that Biggs' method, and any successful method using update (1.8), is eventually virtually the same as the BFGS. This could be said to argue against consideration of this entire class of methods. However, firstly this relationship between Biggs' method and the BFGS has not previously been exhibited, and secondly, away from the solution the methods may differ substantially. We comment briefly on this issue in section 5.

2. A necessary condition for Q-superlinear convergence of modified secant methods

In this section we show that for any quasi-Newton method of the form

$$x_{i+1} = x_i - H_i \nabla f(x_i) \qquad (2.1)$$

$$H_{i+1} \text{ obeys } H_{i+1} y_i = \rho_i s_i$$

$$(s_i, y_i \text{ defined as in (1.4)})$$

to converge Q-superlinearly to the minima of the entire "well-behaved" class of functions f (from an arbitrarily good starting guess x_0), it is necessary that $\lim_{i \to \infty} \rho_i = 1$. This is established through use of a trivial example. Note that in one dimension ($f: R \to R$), (2.1) and the values of x_0 and H_0 completely specify the sequence $\{x_i\}$, subject only to the choices of ρ_i. So consider the nicest one dimensional function of them all, $f(x) = 1/2\, x^2$, whose minimum is $x_* = 0$. Then $\nabla f(x) = x$, so that at each iteration $s_i = y_i$, $H_{i+1} = \rho_i$, and

$$x_{i+2} = x_{i+1} - H_{i+1} \nabla f(x_{i+1})$$
$$= x_{i+1} - \rho_i x_{i+1} = (1-\rho_i) x_{i+1}.$$

Therefore

$$\frac{|x_{i+2} - x_*|}{|x_{i+1} - x_*|} = |1 - \rho_i|$$

and so if $x_0 \neq 0$ and $H_0 \neq 1$, x_i converges Q-superlinearly to x_* only if $\lim_{i \to \infty} \rho_i = 1$. This result is stated formally below. Note that by an expansion of the above example, the condition $\lim_{i \to \infty} \rho_i = 1$ can easily be shown to be necessary for the Q-superlinear convergence of (2.1) on any twice continuously differentiable one dimensional function f with $f''(x_*)$ nonzero and f'' Lipschitz continuous at x_*.

For the remainder of the paper, $\| \cdot \|$ (without subscript) will denote the ℓ_2 vector norm or its induced matrix norm.

Theorem 2.1 Let $f: R^n \to R$ be twice continuously differentiable in the open convex set D, and assume for some $x_* \in D$ and $\ell \geq 0$,

$$\| \nabla^2 f(x) - \nabla^2 f(x_*) \| \leq \ell \| x - x_* \| \qquad (2.2)$$

for all $x \in D$, where $\nabla f(x_*) = 0$ and $\nabla^2 f(x_*)$ is positive definite. Let the sequence $x_i \in R^n$, $i = 1,2,\ldots$ be generated from $x_0 \in R^n$, $H_0 \in R^{n\times n}$ by an iterative method of form (2.1). Then there exists some such f such that if $x_0 \neq x_*$ and $H_0 \neq \nabla^2 f(x_0)^{-1}$, $\{x_i\}$ converges Q-superlinearly to x_* only if $\lim_{i \to \infty} \rho_i = 1$.

3. Sufficient conditions for the Q-superlinear convergence of modified secant methods

In this section we show that the convergence results of Broyden, Dennis and Moré [6] and Powell [13] for quasi-Newton methods using the BFGS update extend to the modified secant form of the BFGS, (1.8), if $\sum_{i=0}^{\infty} |\rho_i - 1|$ is bounded. Our proofs are just extensions of the original results. The main component of all of them is to show that the update under consideration, like the BFGS, is of bounded deterioration if our condition on ρ_i is met. (Bounded deterioration, expressed in equation (3.2) below, is some condition which means essentially that

$$\sum_{i=0}^{\infty} (\| H_{i+1} - \nabla^2 f(x_*)^{-1} \| - \| H_i - \nabla^2 f(x_*)^{-1} \|) \leq c \, \| H_0 - \nabla^2 f(x_*)^{-1} \|$$

for some constant c, which implies that

$$\| H_i - \nabla^2 f(x_*)^{-1} \| / \| H_0 - \nabla^2 f(x_*)^{-1} \| \leq 1 + c$$

for all $i \geq 0$.) This is the main step in applying the techniques of Broyden, Dennis and Moré [6] to any new update, and while it is only one component of Powell's [13] proof, it is the only part requiring non-trivial modification.

The results of Broyden, Dennis and Moré are concerned with functions f which obey $\nabla^2 f(x_*)$ positive definite and several lesser conditions. They analyze the behavior of the direct prediction quasi-Newton iteration

$$x_{i+1} = x_i - H_i \nabla f(x_i) \tag{3.1}$$

in the case when x_0 and H_0 are close to x_* and $\nabla^2 f(x_*)^{-1}$, respectively. Their proofs are based on their Theorem 3.4, which establishes a general condition on the update of H_i under which the update is of bounded deterioration and linear convergence is achieved by (3.1). In Theorem 3.1 we extend this condition slightly to allow for our class of updates. The only change is the addition of the term $w(x,H)$ in (3.2). The proof, which is a straightforward extension of the techniques of Broyden, Dennis and Moré [6], is omitted; it is contained in Schnabel [14], Theorems 9.2.2 and 9.2.3.

Theorem 3.1 Let $f: R^n \to R$ be twice continuously differentiable in the open convex set D, and assume for some $x_* \varepsilon D$ and $\ell \geq 0$, (2.2) holds, where $\nabla f(x_*) = 0$ and $\nabla^2 f(x_*)$ is positive definite. Let $U: R^n \times R^{n\times n} \to 2^{R^{n\times n}} \sim \{\}$ be defined in a neighborhood $N = N_1 \times N_2$ of $(x_*, \nabla^2 f(x_*)^{-1})$ where $N_1 \varepsilon D$. Suppose there exist nonnegative constants α_1, α_2 and a nonsingular symmetric $M \varepsilon R^{n\times n}$ such that for any $(x,H) \varepsilon N$ and $x_+ = x - H \nabla f(x)$, the function U satisfies

$$\| H_+ - \nabla^2 f(x_*)^{-1} \|_{F,M} \le (1+\alpha_1 m) \| H - \nabla^2 f(x_*)^{-1} \|_{F,M} + \alpha_2 m + w(x,H) \tag{3.2}$$

for each $H_+ \in U(x,H)$, where $m = \max \{ \| x_+ - x_* \|, \| x - x_* \| \}$ and $w\colon R^n \times R^{n\times n} \to R$. Consider the sequences $x_i \in R^n$, $H_i \in R^{n\times n}$ generated from $(x_0, H_0) \in N$ by (2.1) with $H_{i+1} \in U(x_i, H_i)$. Then for each $r \in (0,1)$ and any nonnegative constant ω, there exist positive constants $\varepsilon(r)$, $\delta(r)$ such that if $\| x_0 - x_* \| \le \varepsilon(r)$, $\| H_0 - \nabla^2 f(x_*)^{-1} \|_{F,M} \le \delta(r)$ and $\sum_{i=0}^{j} w(x_i, H_i) \le \omega\delta(r)$ for all $j \le 0$, then the sequence $\{H_i\}$ is well defined and $\{x_i\}$ converges to x_*. Furthermore,

$$\| x_{i+1} - x_* \| \le r \, \| x_i - x_* \|$$

for each $i \ge 0$, and $\{ \| H_i \| \}$, $\{ \| H_i^{-1} \| \}$ are uniformly bounded.

From Theorem 3.1 and the techniques of Broyden, Dennis, and Moré [6] it is easy to establish that the direct prediction method using update (1.8) is Q-superlinearly convergent if x_0 and H_0 are sufficiently close to x_* and $\nabla^2 f(x_*)^{-1}$ and $\sum_{i=0}^{\infty} |\rho_i - 1| \le \sigma$ for σ sufficiently small. This is done in Theorem 3.2. The main portion of this proof consists of noticing that the matrices H_{i+1} and $H_{i+1}(\rho_i)$ generated by the BFGS update (1.4) and its modified secant form (1.8) differ only by the term $(\rho_i - 1)\, s_i s_i^T / s_i^T y_i$. From this it is established that when x_i and H_i are near x_* and $\nabla^2 f(x_*)^{-1}$, the difference between the two updates is basically determined by the magnitude of $(\rho_i - 1)$. It follows that $\{H_{i+1}(\rho_i)\}$ is of bounded deterioration if $\sum_{i=0}^{\infty} |\rho_i - 1|$ is bounded.

Theorem 3.2 Let $f:R^n \to R$, $x_* \in R^n$ satisfy the conditions of Theorem 3.1, and define $M = \nabla^2 f(x_*)^{1/2}$. Consider the sequences $x_i \in R^n$, $H_i \in R^{n \times n}$, $i = 1,2,\ldots$ generated from $x_0 \in R^n$ and a symmetric $H_0 \in R^{n \times n}$ by (3.1) with $H_{i+1} = H_{i+1}(\rho_i)$ given by (1.8), s_i, y_i defined as in (1.4) and ρ_i a nonzero real number. Then there exist positive constants ε, δ, σ such that if

$$\|x_0 - x_*\| \le \varepsilon,\ \|H_0 - \nabla^2 f(x_*)^{-1}\|_{F,M} \le \delta \text{ and } \sum_{i=0}^{j} |\rho_i - 1| \le \sigma \text{ for all}$$

$j \ge 0$, then the sequence $\{H_i\}$ is well defined and $\{x_i\}$ converges Q-superlinearly to x_*.

Proof: The proof is a small extension of the work of Broyden, Dennis Moré [6], who have proven this theorem in the case $\rho_i \equiv 1$. With a slight expansion of their techniques we show that update (1.8) satisfies (3.2) with $w(x_i,H_i) = \frac{3}{2}|\rho_i - 1|$. This proves the linear convergence of $\{x_i\}$ by Theorem 3.1, and then the Q-superlinear convergence follows directly from the work of Broyden, Dennis and Moré.

Broyden, Dennis and Moré show that under the assumptions of this theorem, there exist open neighborhoods N around x_* and N_2 around $\nabla^2 f(x_*)^{-1}$ such that if $x_i \in N_1$ and $H_i \in N_2$ is symmetric, then there exist nonnegative constants α_1, α_2, for which

$$\| H_{i+1}(1) - H \|_{F,M} \le (1+\alpha_1 m_i) \| H_i - H \|_{F,M} + \alpha_2 m_i \tag{3.3}$$

where $m_i \triangleq \max\{\|x_{i+1} - x_*\|, \|x_i - x_*\|\}$, $H \triangleq \nabla^2 f(x_*)^{-1}$. ($H_{i+1}(1)$, (1.8) with $\rho_i = 1$, is simply the BFGS.) They show in addition that under these conditions

$$s_i^T y_i \ge (2/3)\|Ms_i\|^2. \tag{3.4}$$

Now update (1.8) can be arranged as

$$H_{i+1}(\rho_i) = H_{i+1}(1) + (\rho_i - 1)s_i s_i^T / s_i^T y_i$$

so that

$$\|H_{i+1}(\rho_i) - H\|_{F,M} \le \|H_{i+1}(1) - H\|_{F,M} + |\rho_i - 1| \; \|s_i s_i^T\|_{F,M} / s_i^T y_i$$

$$= \|H_{i+1}(1) - H\|_{F,M} + |\rho_i - 1| \; \|Ms_i\|^2 / s_i^T y_i .$$

Hence by (3.3) and (3.4),

$$\|H_{i+1}(\rho_i) - H\|_{F,M} \le (1+\alpha_1 m_1) \|H_i - H\|_{F,M} + \alpha_2 m_i + \frac{3}{2} |\rho_i - 1| .$$

Therefore by Theorem 3.1 there exist positive constants ε, δ, and σ such that if $\|x_0 - x_*\| \le \varepsilon$, $\|H_0 - H\|_{F,M} \le \delta$ and $\sum_{i=0}^{j} |\rho_i - 1| \le \sigma$ for all $j \ge 0$, the sequence $\{H_i\}$ is well defined and $\{x_i\}$ converges Q-linearly to x_*. The proof of Q-superlinear convergence is then identical to that of Broyden, Dennis and Moré [6] for the case $\rho_i \equiv 1$.

Powell's [13] convergence result for the BFGS concerns a slightly different method, where

$$x_{i+1} = x_i - \lambda_i H_i \nabla f(x_i), \tag{3.5}$$

λ_i chosen to satisfy two line search conditions given in (3.6). These conditions, which are sometimes used in practice, basically assure that each iteration results in an adequate reduction in f(x) and a step which is not too small. They were shown by Wolfe [16] to guarantee convergence to the solution in many cases.

Powell's [13] theorem has two parts. First, he shows that the points generated by (3.5) using this line search and the BFGS update converge to a minimum of a convex function f which obeys other less restrictive conditions. No assumption about x_0 or H_0 is made (though the assumption of convexity is a strong one). Second he shows that if in addition $\nabla^2 f(x_0)$ is positive definite and $\lambda_i = 1$ is selected whenever it meets the line search conditions, then the convergence is Q-superlinear. In Theorem 3.3 we exhibit conditions under which these results extend to update (1.8). For the first part of the result, establishing convergence, we require only that ρ_i be bounded above and below by positive constants. This fact is noted by Brodlie [4] and takes virtually no new work to prove. For the portion of the proof establishing Q-superlinear convergence, we require $\sum_{i=0}^{\infty} |\rho_i - 1| < +\infty$. The extension of this portion of Powell's proof needs only our bounded deterioration technique from Theorem 3.2.

Theorem 3.3 Let $f:R^n \to R$ be convex, $x_0 \in R^n$, and assume that $D = \{x | f(x) \le f(x_0)\}$ is bounded and f is twice continuously differentiable in D. Let the sequence $x_i \in R^n$, $i = 1,2,\ldots$ be generated from x_0 and a symmetric $H_0 \in R^{n\times n}$ by (3.5), where $H_{i+1} = H_{i+1}(\rho_i)$ given by (1.8), ρ_i is a nonzero real number, and $\lambda_i > 0$ is chosen so that

$$f(x_{i+1}) \le f(x_i) + \beta_1 \nabla f(x_i)^T (x_{i+1}-x_i)$$
$$\nabla f(x_{i+1})^T (x_{i+1}-x_i) \ge \beta_2 \nabla f(x_2)^T (x_{i+1}-x_i) \tag{3.6}$$

β_1, β_2 constants obeying $0 < \beta_1 < \beta_2 < 1$, $\beta_1 < 1/2$. If there exist constants $0 < \rho\min \le 1 \le \rho\max$ such that each $\rho_i \in [\rho\min, \rho\max]$, then the sequence $\{x_i\}$ is well defined and converges to a minimum x_* of f. If in addition $\nabla^2 f(x_*)$ is positive definite, (2.2) holds for all x in some open neighborhood around x_*, and there exists some $\nu > 0$ such that $\sum_{i=0}^{j} |\rho_i - 1| < \nu$ for all $j \ge 0$, then the rate of convergence is Q-superlinear.

Proof: The plain convergence to a minimum follows almost without change from Powell [13]. Redefining his γ_i as $(\nabla f(x_{i+1}) - \nabla f(x_i))/\rho_i$, the only place ρ_i then enters is in his Lemma 1 and equation (3.16), where the extension given $\rho_i \in [\rho min, \rho max]$ is trivial. To extend his proof of Q-superlinear convergence, the only place where the update enters is in the use of Theorem 8.7 of Dennis and Moré [9]. This theorem is readily extended to update (1.8) using the techniques of the proof of Theorem 3.2. Since $\{x_i\} \to x_*$, these show that there exists $i_0 \geq 0$ such that for all $i \geq i_0$,

$$\|H_{i+1}(\rho_i) - \nabla^2 f(x_*)^{-1}\|_{F,M} \leq \|H_{i+1}(1) - \nabla^2 f(x_*)^{-1}\|_{F,M} + (3/2)|\rho_i - 1| . \tag{3.7}$$

(The condition $\{x_i\} \to x_*$ replaces the need for $\|H_0 - \nabla^2 f(x_*)^{-1}\| \leq \delta$ in order for (3.4), and hence (3.7) to hold.) Consulting the original proof of Theorem 8.7 of Dennis and Moré [9] in Theorems 3.1 - 3.4 of Dennis and Moré [8], it is seen that (3.7), $\sum_{i=0}^{\infty} |\rho_i - 1| < +\infty$ and Powell's result that $\sum_{i=0}^{\infty} \|x_i - x_*\| < +\infty$ (which carries over without change) are all that is needed to extend Theorem 8.7 to update (1.8). The remainder of Powell's proof of Q-superlinear convergence carries over without change.

4. The Q-superlinear convergence of Biggs' method

In this section we show that Biggs' [2] modification of the BFGS has the same convergence properties as Broyden-Dennis-Moré [6] and Powell [13] prove for the BFGS, under the same assumptions. Essentially, this amounts to showing that Biggs' choice of the secant parameter ρ_i (in the equation $H_{i+1} y_i = \rho_i s_i$) converges sufficiently quickly to 1. First we describe Biggs' choice of ρ_i.

Biggs' choice can be motivated as follows. He retains the general form of the secant equation $H_{i+1}y_i = \rho_i s_i$ because $\nabla^2 f(x_{i+1})s_i \simeq y_i$ when f is nearly quadratic or s_i is small (and H_{i+1} approximates $\nabla^2 f(x_{i+1})^{-1}$). However, he uses the free parameter ρ_i to introduce some third order information about f. In particular, from the step just taken we have four pieces of information about f: $f(x_i)$, $f(x_{i+1})$, $\nabla f(x_i)$ and $\nabla f(x_{i+1})$. From these, it is possible to model the restriction of f to the line connecting x_i and x_{i+1}, $\hat{f}(\omega) \triangleq f(x_i + \omega s_i)$, by the unique cubic $q(\omega)$ which interpolates $\hat{f}$ and $\hat{f}'$ at x_i and x_{i+1} $(\omega = 0,1)$. Note that $\hat{f}'(\omega) = \nabla f(x_i + \omega s_i)^T s_i$, $\hat{f}''(\omega) = s_i^T \nabla^2 f(x_i + \omega s_i) s_i$. Using standard interpolation techniques, it is found that the second derivative of this cubic model at x_{i+1} $(\omega = 1)$ is

$$q''(1) = 4\nabla f(x_{i+1})^T s_i + 2\nabla f(x_i)^T s_i - 6(f(x_{i+1}) - f(x_i)).$$

Since $q''(1) \simeq \hat{f}''(1) = s_i^T \nabla^2 f(x_{i+1}) s_i$ and we are trying to get $H_{i+1} \simeq \nabla^2 f(x_{i+1})^{-1}$, it would therefore be nice if

$$q''(1) = s_i H_{i+1}^{-1} s_i . \tag{4.1}$$

But we have $H_{i+1}y_i = \rho_i s_i$, and so (4.1) is achieved by setting $q''(1) = s_i^T y_i / \rho_i$, or

$$\rho_i = \frac{s_i^T y_i}{q''(1)} = \frac{s_i^T y_i}{4\nabla f(x_{i+1})^T s_i + 2\nabla f(x_i)^T s_i - 6(f(x_{i+1}) - f(x_i))} . \tag{4.2}$$

This is the value used by Biggs [2].

In Theorem 3.1 we show that the direct prediction quasi-Newton method (3.1) using Biggs' [2] update is locally Q-superlinearly convergent under the assumptions of Broyden, Dennis and Moré [6]. The proof is an extension of their result for the BFGS. We do not use our general result from Theorem 3.2. Instead, it is easier to show directly that when x_i and H_i are sufficiently close to x_* and $\nabla^2 f(x_*)^{-1}$, this update satisfies the original bounded deterioration condition of Broyden, Dennis and Moré [6] (our equation (3.2) without the $w(x,H)$ term). This is done by showing that under these conditions, ρ_i given by (4.2) satisfies

$$|\rho_i - 1| \le k \max\{\|x_{i+1} - x_*\|, \|x_i - x_*\|\}$$

for some constant k. The theorem then follows from Theorem 3.4 of Broyden, Dennis and Moré, the original version of our Theorem 3.1.

Theorem 4.1 Let $f: R^n \to R$, $x_* \in R^n$ satisfy the conditions of Theorem 3.1, and let the sequences $x_i \in R^n$, $H_i \in R^{n\times n}$, $i = 1,2,\ldots$ be generated from $x_0 \in R^n$ and a symmetric $H_0 \in R^{n\times n}$ as in Theorem 3.2, using ρ_i given by (4.2). Define $M = \nabla^2 f(x_*)^{1/2}$. Then there exist positive constants ε, δ such that if $\|x_0 - x_*\| \le \varepsilon$ and $\|H_0 - \nabla^2 f(x_*)^{-1}\|_{F,M} \le \delta$, then the sequence $\{H_i\}$ is well defined and $\{x_i\}$ converges Q-superlinearly to x_*.

Proof: The proof, like that of Theorem 3.2, is an extension of the proof of Broyden, Dennis and Moré [6] for the case $\rho_i \equiv 1$. We show that there exist postive ε and δ such that update (1.8) obeys (3.2) with $w(x_i, H_i) \equiv 0$. The theorem then follows from Theorem 3.1.

As in the proof of Theorem 3.2, we write the update as

$$H_{i+1}(\rho_i) = H_{i+1}(1) + (\rho_i - 1)s_i s_i^T / s_i^T y_i. \tag{4.3}$$

Broyden, Dennis and Moré [6] show that under the assumptions of this theorem, there exist open neighborhoods $\hat{N}_1$ around x_* and $\hat{N}_2$ around $\nabla^2 f(x_*)^{-1}$ such that if $x_i \in \hat{N}_1$ and $H_i \in \hat{N}_2$ is symmetric, then there exist non-negative constants $\hat{\alpha}_1$, $\hat{\alpha}_2$, such that

$$\|H_{i+1}(1) - H\|_{F,M} \leq (1 + \hat{\alpha}_1 m_i)\|H_i - H\|_{F,M} + \hat{\alpha}_2 m_i \tag{4.4}$$

where $m_i \triangleq \max\{\|x_{i+1} - x_*\|, \|x_i - x_*\|\}$, $H \triangleq \nabla^2 f(x_i)^{-1}$.

We simply show that in addition there exist possibly smaller neighborhoods N_1, N_2 such that also

$$|\rho_i - 1| \leq (9/2)\,\ell\,\|H\| m_i. \tag{4.5}$$

Then, since Broyden, Dennis and Moré have also shown that under these assumptions

$$s_i^T y_i \geq (2/3)\,\|M s_i\|^2 \tag{4.6}$$

(4.5) and (4.6) show that

$$\|(\rho_i - 1)s_i s_i^T / s_i^T y_i\|_{F,M} \leq (27/4)\,\ell\,\|H\| m_i. \tag{4.7}$$

Thus from (4.3), (4.4), (4.7) and the triangle inequality, update (1.8) satisfies (3.2) with $\alpha_1 = \hat{\alpha}_1$, $\alpha_2 = \hat{\alpha}_2 + (27/4)\,\ell\,\|H\|$. Linear convergence now follows from Theorem 3.1, and Q-superlinear convergence without change from the proof of Broyden, Dennis and Moré for the BFGS.

To show (4.5), it is easier to first show that

$$|\rho_i^{-1} - 1| \leq (9/4)\,\ell\,\|H\| m_i.$$

For m_i sufficiently small this clearly implies (4.5). For the remainder of the proof we write x, x_+, f, f_+, g, g_+, m, s and y for x_i, x_{i+1}, $f(x_i)$, $f(x_{i+1})$, $\nabla f(x_i)$, $\nabla f(x_{i+1})$, m_i, s_i and y_i

respectively. Then

$$\rho_i^{-1} - 1 = \frac{4\, g_+^T s + 2 g^T s - 6(f_+ - f)}{s^T y} - 1$$

$$= 3 \left[\frac{g_+^T s + g^T s - 2(f_+ - f)}{s^T y} \right] \tag{4.8}$$

since $y = g_+ - g$. From elementary calculus

$$g_+ = g + \left[\int_{\tau=0}^{1} \nabla^2 f(x + \tau s)\, d\tau \right] s,$$

$$f_+ = f + g^T s + s^T \left[\int_{\tau=0}^{1} (1 - \tau)\, \nabla^2 f(x + \tau s)\, d\tau \right] s.$$

Taking the inner product of the first equation with s and subtracting from it twice the second,

$$g_+^T s + g^T s - 2(f_+ - f) = s^T \left[\int_{\tau=0}^{1} (2\tau - 1)\, \nabla^2 f(x + \tau s)\, d\tau \right] s$$

and using the transform $\phi = 2\tau - 1$,

$$g_+^T s + g^T s - 2(f_+ - f) =$$

$$s^T \left[\int_{\phi=0}^{1} \phi \{ \nabla^2 f(x + \tfrac{1+\phi}{2} s) - \nabla^2 f(x + \tfrac{1-\phi}{2} s \} \, d\phi/2 \right] s. \tag{4.9}$$

The proof is essentially completed by showing that the term in curly brackets has norm $\leq 2\,\ell\, m$ for any $\phi \in [0,1]$. By the triangle inequality,

$$\| \nabla^2 f(x + \tfrac{1+\phi}{2} s) - \nabla^2 f(x + \tfrac{1-\phi}{2} s) \| \leq$$

$$\| \nabla^2 f(x + \tfrac{1+\phi}{2} s) - \nabla^2 f(x_*) \| + \| \nabla^2 f(x + \tfrac{1-\phi}{2} s) - \nabla^2 f(x_*) \|$$

and so using Lipschitz condition (2.2) and the triangle inequality again,

$$\| \nabla^2 f(x+\tfrac{1+\phi}{2}s) - \nabla^2 f(x+\tfrac{1-\phi}{2}s) \| \le$$
$$\ell \left[\| \tfrac{1-\phi}{2}x + \tfrac{1+\phi}{2}x_+ - x_* \| + \| \tfrac{1+\phi}{2}x + \tfrac{1-\phi}{2}x_+ - x_* \| \right] \le$$
$$\ell \left[\| x - x_* \| + \| x_+ - x_* \| \right] \le 2\ell m . \qquad (4.10)$$

Using (4.10) we get from (4.9)

$$|g_+^T s + g^T s - 2(f_+ - f)| \le \| s \|^2 \int_{\phi=0}^{1} \phi\, \ell m \, d\phi = \| s \|^2 \ell m/2 . \qquad (4.11)$$

Thus from (4.6), (4.8) and (4.11),

$$|\rho_i^{-1} - 1| \le \frac{3}{2} \frac{\| s \|^2 \ell m}{s^T y} \le \frac{9}{4} \frac{\| s \|^2 \ell m}{\| Ms \|^2}$$

$$\le (9/4)\, \ell \, \| M^{-2} \| m = (9/4)\, \ell \, \| H \| m .$$

For N_1 and N_2 sufficiently small, m_i is sufficiently small for this to imply (4.5), which completes the proof.

Theorem 4.2 contains the extension of Powell's [13] convergence theorem to Biggs' [2] method. As in Theorem 3.3, the proof is almost immediate given the bounded deterioration result for Biggs' method established above.

Theorem 4.2 Let $f: R^n \to R$, $x_0 \in R^n$ satisfy the conditions of Theorem 3.3, and let the sequences $x_i \in R^n$, $H_i \in R^{n \times n}$, $i = 1,2,\ldots$ be generated from $x_0 \in R$ and a symmetric $H_0 \in R^{n\times n}$ as in Theorem 3.3, using ρ_i given by (4.2). If there exist constants $0 < \rho\text{min} \le 1 \le \rho\text{max}$ such that ρ_i is also constrained to be in $[\rho\text{min}, \rho\text{max}]$, then the sequence $\{x_i\}$ is well-defined and converges to a mimimum x_* of f. If in addition $\nabla^2 f(x_*)$ is positive definite and (2.2) holds for all x in some open neighborhood around x_*, then the rate of convergence is Q-superlinear.

Proof: The proof of convergence to a minimum is the same as in Theorem 3.3. The proof of the Q-superlinear rate again requires only the extension of Theorem 8.7 of Dennis and Moré [9] to Biggs' update. It is seen by consulting the original proof of Theorem 8.7 in Theorems 3.1 - 3.4 of Dennis and Moré [8] that Theorem 8.7 can now be applied directly to Biggs' update, since this update has been shown in the proof of Theorem 4.1 to obey the standard bounded deterioration condition, (3.2) with $w(x,H) \equiv 0$. The proof of Q-superlinear convergence is thus immediate by the techniques of Powell [13].

5. Summary and concluding remarks

We have shown in this paper that if the BFGS method is modified to obey the secant equation $H_{i+1}y_i = \rho_i s_i$, then its theoretical convergence properties are preserved as long as ρ_i converges quickly enough to 1. We also show that at least convergence to 1 is necessary to assure Q-superlinear convergence under the normal assumptions. Furthermore, we have shown that in the instance of such a method which has been successful in practice, Biggs' [2] method, ρ_i has precisely this property and the Q-superlinear convergence properties of the BFGS are retained.

One interesting consequence of our results is that they show that for most problems, Biggs' method will be virtually the same as the BFGS close to the solution. This raises the question as to whether there is any point in using Biggs' method. Indeed, Brodlie [4] finds the two methods to be about the same in practice, and so suggests using the BFGS as it is somewhat simpler. Our theoretical results provide no firm basis for disagreement. However, it should be considered that Bigg's method will differ from the BFGS further from the solution where the function is less quadratic, and in this situation the use of a cubic model may be helpful. Therefore, since our results show that Biggs' method is essentially equivalent to the BFGS near the solution, and since it may be attractive further away, its continued consideration may be justified.

It should finally be noted that our techniques in Theorems 3.2 and 4.1 of extending the BFGS proof of Broyden, Dennis and Moré [6] to its modified secant form can equivalently be used to prove Q-superlinear convergence of the modified secant form of any other update which has been proven Q-superlinearly convergent by the techniques of Broyden, Dennis and Moré. These include the PSB, DFP (see Broyden, Dennis and Moré [6]) and versions of the method of Davidon [7] considered by Schnabel [14]. However, since none of these updates have been superior to the BFGS in practice, there is no reason to expect that their modified secant form would be superior to Biggs' update.

References

[1] M. C. Biggs, "Minimization algorithms making use of non-quadratic properties of the objective function", Journal of the Institute of Mathematics and its Applications 8 (1971) 315-327.

[2] M. C. Biggs, "A note on minimization algorithms which make use of non-quadratic properties of the objective function", Journal of the Institute of Mathematics and its Applications 12 (1973) 337-8.

[3] K. W. Brodlie, "Unconstrained minimization", in: D. Jacobs, ed., The State of the Art in Numerical Analysis (Academic Press, London, 1977).

[4] K. W. Brodlie, "An assessment of two approaches to variable metric methods", Mathematical Programming 12 (1977) 344-355.

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

[6] C. G. Broyden, J. E. Dennis and J. J. Moré, "On the local and superlinear convergence of quasi-Newton methods", Journal of the Institute of Mathematics and its Applications 12 (1973) 223-246.

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

[8] J. E. Dennis and J. J. Moré, "A characterization of superlinear convergence and its application to quasi-Newton methods", Mathematics of Computation 28 (1974) 549-560.

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

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

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

[12] H. Y. Huang, "Unified approach to quadratically convergent algorithms for function minimization, Journal of Optimization Theory and its Applications 5 (1970) 405-423.

[13] M. J. D. Powell, "Some global convergence properties of a variable metric algorithm for minimization without exact line searches", in: R. W. Cottle and C. E. Lemke, eds., Nonlinear Programming (SIAM-AMS proceeding, vol. IX, 1976).

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

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

[16] P. Wolfe, "Convergence conditions for ascent methods", SIAM REVIEW 11 (1969) 226-235.