NBER WORKING PAPER SERIES

SOME CONVERGENCE PROPERTIES OF BROYDEN'S METHOD

David M. Gay*

Working Paper No. 175

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

July 1977

Preliminary

*NBER Computer Research Center. Research supported in part by National Science Foundation Grant MCS76-00324 to the National Bureau of Economic Research, Inc.

Abstract

In 1965 Broyden introduced a family of algorithms called (rank-one) quasi-Newton methods for iteratively solving systems of nonlinear equations. We show that when any member of this family is applied to an n×n nonsingular system of linear equations and direct-prediction steps are taken every second iteration, then the solution is found in at most 2n steps. Specializing to the particular family member known as Broyden's (good) method, we use this result to show that Broyden's method enjoys local 2n-step Q-quadratic convergence on nonlinear problems.

Contents

1. Introduction

In 1965 Broyden [1965] introduced a family of algorithms called quasi-Newton methods for solving systems of nonlinear equations, i.e., for finding $x^* \in \mathbb{R}^n$ such that $f(x^*) = 0$, where $f: \mathbb{R}^n \to \mathbb{R}^n$ is differentiable. Broyden proposed a modified form of Newton's method in which an approximation H to the inverse of the true Jacobian matrix $f'(x)$ is used and updated after each step. This leads to an iteration of the form $x_{i+1} = x_i - \lambda_i H_i f(x_i)$, where the steplength λ_i is chosen to promote convergence. In what follows we shall usually restrict our attention to direct prediction methods, i.e., $\lambda_i \equiv 1$ as in Newton's method. By analogy with the DFP method [Davidon, 1959; Fletcher & Powell, 1963] for unconstrained minimization, and by considering what is desirable when f is linear, Broyden proposed updating H_i in such a way that the quasi-Newton equation $H_{i+1}\left[f(x_{i+1}) - f(x_i)\right] = x_{i+1} - x_i$ holds. Since new information is picked up in only one direction each step, Broyden suggested obtaining H_{i+1} from H_i by means of a rank 1 update, i.e., by adding a matrix of rank 1 to H_i. This leads to the following iterative procedure.

Choose nonsingular $H_0 \in \mathbb{R}^{n\times n}$ and $x_0 \in \mathbb{R}^n$.

For $k = 0, 1, 2, \ldots$ let

(1.1a) $$s_k = -H_k f(x_k);$$

(1.1b) $$x_{k+1} = x_k + s_k;$$

(1.1c) $$y_k = f(x_{k+1}) - f(x_k);$$

(1.1d) If $y_k = 0$ then $H_{k+1} = H_k$;

(1.1e) else choose $v_k \in \mathbb{R}^n$ such that $v_k^T y_k = 1$

(1.1f) and $v_k^T H_k^{-1} s_k \neq 0$

(1.1g) and let $H_{k+1} = H_k + (s_k - H_k y_k) v_k^T$.

Because of (1.1f) and the Sherman-Morrison [1949] formula, H_{k+1} is nonsingular whenever H_k is, so induction shows that H_k is nonsingular for all k, whence $s_k = 0$ only if $f(x_k) = 0$.

Broyden's [1965] method (sometimes called his first or good method) results from choosing $v_k = H_k^T s_k/(s_k^T H_k y_k)$ in (1.1e) -- and is defined for $y_k \neq 0$ only so long as $s_k^T H_k y_k \neq 0$ and $s_k^T y_k \neq 0$. Broyden's second or bad method results from choosing $v_k = y_k/(y_k^T y_k)$ when $y_k \neq 0$ and is defined so long as $y_k H_k^{-1} s_k \neq 0$.

Broyden has shown that his (first) method converges locally at least linearly on nonlinear problems [1971] and at least R-superlinearly on linear problems [1970]. Later, Broyden, Dennis, and Moré [1973] showed that both Broyden's good and his bad method converge locally at least Q-superlinearly. Moré and Trangenstein [1976] subsequently proved that "locally" could be replaced by "globally" when a modified form of Broyden's method is applied to linear systems of equations. In Section 2 of this paper we show that when any form of (1.1), including Broyden's good and bad methods (so long as they are defined), is applied to a system of linear equations $f(x) = Ax - b$ in which $A \in \mathbb{R}^{n\times n}$ is nonsingular, then the iteration converges in at most 2n steps (i.e., $x_j = x^* \equiv A^{-1}b$

for all $j \geq 2n$). We show further that this result also holds when some nonunit steps are allowed ($s_k = -\lambda_k H_k f(x_k)$, with $\lambda_k \neq 0, 1$). Specializing to Broyden's good method, we show in Section 3 that this method enjoys local 2n-step Q-quadratic convergence on nonlinear problems. Section 4 presents some concluding remarks.

2. Finite Termination on Linear Systems

In this section we show that Algorithm (1.1) converges in at most $2n$ steps when applied to an f representing a nonsingular system of linear equations: $f(x) = Ax - b$, where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, and A is nonsingular. This follows as an easy corollary to the following lemma, which holds even if A is singular. The notation $\lfloor\sigma\rfloor$ used below denotes the greatest integer less than or equal to $\sigma \in \mathbb{R}$, while for non-zero $u, v \in \mathbb{R}^n$, the notation $u \| v$ means that $u = \lambda v$ for some real $\lambda \neq 0$.

Lemma 2.1

If $A \in \mathbb{R}^{n\times n}$ and Algorithm (1.1) is applied to $f(x) \equiv Ax - b$ with the result that $f_k \equiv f(x_k)$ and y_{k-1} are linearly independent, then for $1 \leq j \leq \lfloor(k+1)/2\rfloor$,

(2.1) $(AH_{k-2j+1})^i f_{k-2j+1}$, $0 \leq i \leq j$, are linearly independent.

Proof: Since $y_{k-1} = As_{k-1} = -AH_{k-1}f_{k-1}$, this is easily seen to hold for $j = 1$. Assume it true for $j = m < \lfloor(k+1)/2\rfloor$. Note that $2m \leq k-1$, whence $k-2m-1 \geq 0$. Also note from (1.1a,b,c,d) that $y_i = 0 \implies y_{i+1} = 0$, so $y_{k-1} \neq 0 \implies$

$y_{k-2m-1} \neq 0$. Now

$$s_{k-2m} - H_{k-2m}y_{k-2m} = -H_{k-2m}f_{k-2m} + H_{k-2m}AH_{k-2m}f_{k-2m}$$

$$= -H_{k-2m}(I - AH_{k-2m})f_{k-2m}, \quad \text{so}$$

$AH_{k-2m+1} = AH_{k-2m}(I - [I - AH_{k-2m}]f_{k-2m}v_{k-2m}^T)$. Moreover, $f_{k-2m+1} = (I - AH_{k-2m})f_{k-2m}$. Since $(AH_{k-2m+1})^i f_{k-2m+1}$, $0 \leq i \leq m$, are linearly independent by the induction hypothesis (2.1), we see by the two preceeding equations and induction on i that there exist $\gamma_{i,\ell}$ (dependent on k and m) such that

$$(AH_{k-2m+1})^i f_{k-2m+1} \;\Big\|\; (I-AH_{k-2m})\Big[(AH_{k-2m})^i + \sum_{\ell=1}^{i-1}\gamma_{i,\ell}(AH_{k-2m})^\ell\Big]f_{k-2m}$$

for $0 \leq i \leq m$, whence $(I - AH_{k-2m})(AH_{k-2m})^i f_{k-2m}$, $0 \leq i \leq m$, are linearly independent. But $(I - AH_{k-2m})y_{k-2m-1} = 0$ by (1.1e,g), so $y_{k-2m-1} = -AH_{k-2m-1}f_{k-2m-1}$ and $(AH_{k-2m})^i f_{k-2m}$, $0 \leq i \leq m$, are linearly independent. As before, we see that there exist $\delta_{i,\ell}$ (dependent on k and m) such that

$$(AH_{k-2m})^i f_{k-2m} \;\Big\|\; (I - AH_{k-2m-1})\Big[(AH_{k-2m-1})^i + \sum_{\ell=1}^{i-1}\delta_{i,\ell}(AH_{k-2m-1})^\ell\Big]f_{k-2m-1}$$

for $0 \leq i \leq m$, whence we readily see that $(AH_{k-2m-1})^i f_{k-2m-1}$, $0 \leq i \leq m+1$, are linearly independent. Thus (2.1) holds for $j = m+1$, and the lemma follows by induction. ∎

Theorem 2.2: If $f(x) = Ax - b$ and $A \in \mathbb{R}^{n\times n}$ is nonsingular, then Algorithm (1.1) converges in at most $2n$ steps (i.e., $f_{2n} = 0$).

Proof: As noted above, H_k is nonsingular for all k; since A is nonsingular, we thus see that if f_{2n-2} is nonzero, then the same is true of $s_{2n-2} = -H_{2n-2}f_{2n-2}$ and $y_{2n-2} = As_{2n-2}$. If $f_{2n-1} \neq 0$, then necessarily $f_{2n-2} \neq 0$, so $y_{2n-2} \neq 0$ and Lemma 2.1 implies that $f_{2n-1} \parallel y_{2n-2}$ (since otherwise $\mathbb{R}^n$ would contain $n+1$ linearly independent vectors). Since $s_{2n-2} = H_{2n-1}y_{2n-2} = H_{2n-1}As_{2n-2}$ by (1.1e,g), we have $y_{2n-2} = As_{2n-2} = AH_{2n-1}y_{2n-2}$ and hence $f_{2n-1} = AH_{2n-1}f_{2n-1}$, so

$$f_{2n} = f_{2n-1} + As_{2n-1} = f_{2n-1} - AH_{2n-1}f_{2n-1} = 0. \quad \blacksquare$$

Theorem 2.2 leaves several questions unanswered, such as whether a full $2n$ steps may actually be required. Computer runs suggest for small values of n that Broyden's good and bad methods may both require a full $2n$ steps. As we shall now see, it is possible for arbitrary n and nonsingular A to choose H_o, f_o, and the v_k in (1.1e) so that Algorithm (1.1) requires a full $2n$ steps. This is the content of Theorem 2.4, proof of which requires the following lemma.

Lemma 2.3: In Algorithm (1.1), if $y_k \neq 0$, $v_k^T y_{k-1} \neq 0$, and $\mathrm{Rank}(I - AH_k) = n-1$, then $\mathrm{Rank}(I - AH_{k+1}) = n-1$.

Proof:
$$\begin{aligned} I - AH_{k+1} &= I - AH_k(I - [I - AH_k]f_k v_k^T) \\ &= (I - AH_k)(I + AH_k f_k v_k^T) \\ &= (I - AH_k)(I - y_k v_k^T). \end{aligned}$$

It suffices to show that $(I - AH_{k+1})u \neq 0$ whenever u is linearly independent of y_k. Now $\mathrm{Rank}(I - AH_k) = n-1$ and $(I - AH_k)y_{k-1} = 0$, so $(I - AH_k)w \neq 0$ whenever w is linearly independent of y_{k-1}. But $v_k^T(I - y_k v_k^T) = 0$, so for $w = (I - y_k v_k^T)u$, we have $v_k^T w = 0$, while $v_k^T \mathbf{y}_{k-1} \neq 0$ by assumption. Thus w is linearly independent of y_{k-1}, whence $(I - AH_{k+1})u = (I - AH_k)w \neq 0$. ∎

Theorem 2.4: If $I - AH_o$ is nonsingular, if $(AH_o)^i f_o$, $0 \leq i \leq n-1$, are linearly independent, if $v_k^T f_k \neq 0$ and $v_k^T H_k^{-1} s_k \neq 0$ for $k \geq 0$, and if $v_k^T y_{k-1} \neq 0$ for $k \geq 1$, then Algorithm (1.1) requires a full $2n$ steps to converge.

Proof: As in the proof of Lemma 2.1, we find

(2.2a) $\quad f_{k+1} = (I - AH_k)f_k,$

(2.2b) $\quad AH_{k+1}f_{k+1} = -(v_k^T f_k)(I - AH_k)(AH_k)f_k,$ and

(2.2c) $\quad (AH_{k+1})^i f_{k+1} = -(v_k^T f_k)(I - AH_k)\left[(AH_k)^i + \sum_{\ell=1}^{i-1} \delta_{i,\ell}^{(k)}(AH_k)^\ell\right]f_k$

for $i > 1$. In particular, since $v_o^T f_o \neq 0$, $I - AH_o$ is nonsingular, and $(AH_o)^i f_o$, $0 \leq i \leq n-1$, are linearly independent, we see for $j = 1$ that

(2.3a) $\quad (AH_{2j-1})^i f_{2j-1}$, $0 \leq i \leq n-j$, are linearly independent.

Moreover, since $I - AH_1 = (I - AH_o)(I - y_o v_o^T)$, we see for $j = 1$ that

(2.3b) $\quad \mathrm{Rank}(I - AH_{2j-1}) = n-1.$

Suppose (2.3) holds for $j = k < n$. Since $y_{2k-1} = -AH_{2k-1}f_{2k-1}$, we see from (2.2) and (2.3a) that y_{2k-1} and $(AH_{2k})^i f_{2k}$, $0 \le i \le n-k-1$, are linearly independent. But $(I - AH_{2k})y_{2k-1} = 0$, while $\text{Rank}(I - H_{2k}) = n-1$ by (2.3b) and Lemma 2.3, so $\{y_{2k-1}\}$ spans the null space of $I - AH_{2k}$ and $(I - AH_{2k})(AH_{2k})^i f_{2k}$, $0 \le i \le n-k-1$, must be linearly independent (since otherwise y_{2k-1} were a linear combination of $(AH_{2k})^i f_{2k}$, $0 \le i \le n-k-1$). From (2.2) it follows that (2.3a) holds for $j = k+1$, while (2.3b) holds for $j = k+1$ by Lemma 2.3. Thus (2.3) holds by induction for $1 \le j \le n$. In particular, $f_{2n-1} \ne 0$, whence Algorithm (1.1) runs a full $2n$ steps before converging. ∎

Another question that Theorem 2.2 leaves unanswered is what happens when a step-length parameter is introduced, i.e., when step (1.1a) is replaced by $s_k = -\lambda_k H_k f(x_k)$. For $\lambda_k \ne 0, 1$, (2.2) becomes

$$(2.4) \qquad (AH_{k+1})^i f_{k+1} \,\Big\|\, (I - AH_k)\left[(AH_k)^i + \sum_{\ell=1}^{i-1} \delta_{i,\ell}^{(k)} (AH_k)^\ell\right] f_k + \, + \delta_{i,o}^{(k)} y_k \quad ,$$

i.e., multiples of y_k are added to the right-hand sides of (2.2). Thus the proof of Lemma 2.1 is unaffected if $\lambda_{k-2m-1} \ne 1$, $1 \le m < \lfloor (k+1)/2 \rfloor$; it seems essential only that $\lambda_{k-2m} = 1$. More generally, we see from (2.4) that if $(AH_{k+1})^i f_{k+1}$, $0 \le i \le m$, are linearly independent, then the set $\{(AH_k)^i f_k \mid 0 \le i \le m+1\}$ must contain at least $m+1$ linearly independent vectors, whence $(AH_k)^i f_k$, $0 \le i \le m$, must be linearly independent (since if $(AH_k)^j f_k$ were dependent on $(AH_k)^i f_k$, $0 \le i < j$, for some $j \le m$, then $(AH_k)^\ell f_k$ could be expressed as linear combinations of these same vectors for all $\ell \ge j$). Hence the proof of Lemma 2.1 may be modified to obtain

Theorem 2.5: If Algorithm (1.1) is applied to a linear function $f(x) = Ax - b$ with $A \in \mathbb{R}^{n\times n}$ nonsingular and (1.1a) replaced by $s_k = -\lambda_k H_k f_k$ $(\lambda_k \ne 0)$, and if there are integers k_i, $0 \le i \le n$, such that $k_o = -1$ and $\lambda_{k_i} = 1$ with $k_i \ge k_{i-1} + 2$ for $1 \le i \le n$, then $f_{k_n} = 0$. ∎

Theorem 2.4 is readily generalized to allow $\lambda_{2k} \ne 1$, $0 \le k < n$. Whether a further generalization along the lines of Theorem 2.5 is possible remains an open question.

3. Local 2n-Step Q-Quadratic Convergence of Broyden's Method

We now restrict our attention to the direct-prediction version of Broyden's (good) method. This amounts to Algorithm (1.1) with $v_k = H_k^T s_k/(s_k^T H_k y_k)$. With this v_k, it is well known (and easily seen from the Sherman-Morrison [1949] formula) that if H_k is nonsingular and $s_k^T H_k y_k \neq 0$, then $H_{k+1}^{-1} = H_k^{-1} + (y_k - H_k^{-1} s_k) s_k^T/(s_k^T s_k)$. We shall find it somewhat more convenient to restate the direct-prediction Broyden's method in terms of $B_k = H_k^{-1}$. Thus for $k = 0, 1, 2, \ldots$ we are dealing with the following iteration (in which $f_k \equiv f(x_k)$):

(3.1a) $s_k = -B_k^{-1} f_k$;

(3.1b) $x_{k+1} = x_k + s_k$;

(3.1c) $y_k = f_{k+1} - f_k$;

(3.1d) If $s_k^T B_k^{-1} y_k \neq 0$ then $B_{k+1} = B_k + (y_k - B_k s_k) s_k^T/(s_k^T s_k)$

(3.1e) else $B_{k+1} = B_k$.

In what follows, $\|\cdot\| = \|\cdot\|_2$ denotes the Euclidean vector norm $\|x\| = (x^T x)^{1/2}$ or the corresponding induced matrix norm. We may now state the main result of this section:

Theorem 3.1: Suppose $f: \mathbb{R}^n \to \mathbb{R}^n$ is differentiable with $f(x) - f(y) = \int_0^1 f'(y + \tau(x-y))(x-y)d\tau$, that $f(x^*) = 0$ with $A \equiv f'(x^*)$ nonsingular, and that the Jacobian matrix f' is

Lipschitz continuous at x^*, i.e., for some constant λ and all x sufficiently close to x^*,

$$\|f'(x) - f'(x^*)\| \leq \lambda \|x - x^*\|. \tag{3.2}$$

Then there exist $\gamma, \delta, \varepsilon > 0$ such that if

$$\|x_o - x^*\| \leq \delta \quad \text{and} \quad \|B_o - A\| \leq \varepsilon, \tag{3.3}$$

then the iterates produced by Algorithm (3.1) satisfy

$$\|x_{\ell+2n} - x^*\| \leq \gamma \|x_\ell - x^*\|^2 \tag{3.4}$$

for all $\ell \geq 0$.

Proof: From (3.2) and the proofs of Theorems 3.2 and 4.3 of [Broyden, Dennis, & Moré, 1973], there exist $\delta, \varepsilon > 0$ with $\delta \leq 1/(4\|A^{-1}\|)$ such that if (3.3) holds, then

$$\|x_{k+1} - x^*\| \leq \|x_k - x^*\|, \tag{3.5}$$

$$\|B_k - A\| \leq 2\delta, \tag{3.6}$$

$$\|f(x_{k+1})\| \leq \|f(x_k)\|, \tag{3.7}$$

$$\|B_k^{-1}\| \leq \frac{4}{3}\|A^{-1}\|, \tag{3.8}$$

and $s_k^T B_k^{-1} y_k = 0$ only if $x_k = x^*$. Fix ℓ and let $h = \|x_\ell - x^*\|$: we must show that there is a γ independent of ℓ such that $\|x_{\ell+2n} - x^*\| \leq \gamma h^2$.

Consider the sequences $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{2n}$ of vectors and $\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_{2n}$ of matrices generated from $\hat{x}_o = x_\ell$ and

$\hat{B}_o = B_\ell$ by (3.1) with $f(x)$ replaced by $\hat{f}(x) \equiv A(x - x^*)$, i.e., $\hat{f}_k = \hat{f}(\hat{x}_k)$, $\hat{s}_k = -\hat{B}_k^{-1}\hat{f}_k$, $\hat{x}_{k+1} = \hat{x}_k + \hat{s}_k$, $\hat{y}_k = \hat{f}_{k+1} - \hat{f}_k = A\hat{s}_k$, and

(3.9) $$\hat{B}_{k+1} = \hat{B}_k + (\hat{y}_k - \hat{B}_k\hat{s}_k)\hat{s}_k^T/(\hat{s}_k^T\hat{s}_k)$$

for $\hat{s}_k \neq 0$, with $\hat{B}_{k+1} = \hat{B}_k$ if $\hat{s}_k = 0$. Similarly to (3.6) and (3.8), we have

(3.10) $$\|\hat{B}_k - A\| \leq 4\delta \quad \text{and}$$

(3.11) $$\|\hat{B}_k^{-1}\| \leq 2\|A^{-1}\|.$$

We show by induction that there exist $\gamma_{1,j}$ and $\gamma_{2,j}$ (independent of ℓ) such that

(3.12a) $$\|B_{\ell+j} - \hat{B}_j\| \, \|f_{\ell+j}\| \leq \gamma_{1,j}h^2 \quad \text{and}$$

(3.12b) $$\|x_{\ell+j} - \hat{x}_j\| \leq \gamma_{2,j}h^2$$

for $0 \leq j \leq 2n$.

Since $B_\ell = \hat{B}_o$ and $x_\ell = \hat{x}_o$, (3.12) holds for $j = 0$ with $\gamma_{1,0} = \gamma_{2,0} = 0$. Suppose it holds for $j = k$. To establish (3.12b) for $j = k+1$, we first note that

$$\|s_{\ell+k} - \hat{s}_k\| = \|B_{\ell+k}^{-1}f_{\ell+k} - \hat{B}_k^{-1}\hat{f}_k\|$$

(3.13) $$\leq \|B_{\ell+k}^{-1}\| \, \|\hat{B}_k^{-1}\| \, \|B_{\ell+k} - \hat{B}_k\| \, \|f_{\ell+k}\| + $$

$$+ \|\hat{B}_k^{-1}\| \, \|f_{\ell+k} - \hat{f}_k\|.$$

Now $f_{\ell+k} - \hat{f}_k = \left[f(x_{\ell+k}) - \hat{f}(x_{\ell+k})\right] + \left[\hat{f}(x_{\ell+k}) - \hat{f}(\hat{x}_k)\right]$

and (3.2) and (3.5) imply

$$\left\|f(x_{\ell+k}) - \hat{f}(x_{\ell+k})\right\| = \left\|\int_0^1 \left[f'(x^* + \tau[x_{\ell+k}-x^*]) - f'(x^*)\right](x_{\ell+k}-x^*)d\tau\right\|$$

$$\leq \lambda\|x_{\ell+k} - x^*\| \int_0^1 \|x_{\ell+k} - x^*\|\tau d\tau$$

$$= \frac{\lambda}{2}\|x_{\ell+k} - x^*\|^2 \leq \frac{\lambda}{2}h^2,$$

while for $\gamma_{3,k} = \|A\|\gamma_{2,k}$,

$\|\hat{f}(x_{\ell+k}) - \hat{f}(\hat{x}_k)\| \leq \|A\| \; \|x_{\ell+k} - \hat{x}_k\| \leq \gamma_{3,k}h^2$ by (3.12b), so (3.13), (3.8), (3.11), and (3.12a) imply

$$(3.14) \qquad \|s_{\ell+k} - \hat{s}_k\| \leq \gamma_{4,k}h^2,$$

where $\gamma_{4,k} = \frac{8}{3}\|A^{-1}\|^2\gamma_{1,k} + \|A^{-1}\|(\lambda + 2\gamma_{3,k})$, which, along with (3.12b) for $j = k$, gives (3.12b) for $j = k+1$ with $\gamma_{2,k+1} = \gamma_{2,k} + \gamma_{4,k}$.

It remains to show that (3.14) and (3.12a) for $j = k$ imply (3.12a) for $j = k+1$. If $s_{\ell+k} = 0$ or $\hat{s}_k = 0$, then either $f_{\ell+k+1} = 0$ or $\hat{f}_{k+1} = 0$, whence the reasoning above (3.14) shows $\|f_{\ell+k+1}\| \leq (\frac{\lambda}{2} + \gamma_{3,k+1})h^2$; together with (3.6) and (3.10), this shows for $\gamma_{5,k+1} = 3\delta(\lambda + 2\gamma_{3,k+1})$ that $\|B_{\ell+k+1} - \hat{B}_{k+1}\| \; \|f_{\ell+k+1}\| \leq \gamma_{5,k+1}h^2$. On the other hand, if both $s_{\ell+k} \neq 0$ and $\hat{s}_k \neq 0$ (as we henceforth assume), then in view of (3.1d), (3.9), and (3.7), we need only show that

$$(3.15) \qquad \left\|(y_{\ell+k} - B_{\ell+k}s_{\ell+k})s_{\ell+k}^T/(s_{\ell+k}^T s_{\ell+k}) - (\hat{y}_k - \hat{B}_k\hat{s}_k)\hat{s}_k^T/(\hat{s}_k^T\hat{s}_k)\right\| \; \left\|f_{\ell+k}\right\| \leq \gamma_{6,k}h^2,$$

for then (3.12a) holds for $j = k+1$ with $\gamma_{1,k+1} = \max\{\gamma_{5,k+1}, \gamma_{1,k} + \gamma_{6,k}\}$. Let $A_k = \int_0^1 f'(x_{\ell+k} + \tau s_{\ell+k})d\tau$, so that $y_{\ell+k} = A_k s_{\ell+k}$. Since $\hat{y}_k = A\hat{s}_k$, we have the follow-bound on LHS(3.15), the left-hand side of (3.15):

$$\begin{aligned} \text{LHS}(3.15) &= \left\| (A_k - B_{\ell+k}) s_{\ell+k} s_{\ell+k}^T / (s_{\ell+k}^T s_{\ell+k}) - \right. \\ &\qquad \left. - (A - \hat{B}_k)\hat{s}_k \hat{s}_k^T / (\hat{s}_k^T \hat{s}_k) \right\| \, \| f_{\ell+k} \| \\ &\leq \left\| [(A_k - A) - (B_{\ell+k} - \hat{B}_k)] s_{\ell+k} s_{\ell+k}^T / (s_{\ell+k}^T s_{\ell+k}) \right\| \, \| f_{\ell+k} \| + \\ &\quad + \left\| (A - \hat{B}_k)[s_{\ell+k} s_{\ell+k}^T / (s_{\ell+k}^T s_{\ell+k}) - \hat{s}_k \hat{s}_k^T / (\hat{s}_k^T \hat{s}_k)] \right\| \, \| f_{\ell+k} \| . \end{aligned} \tag{3.16}$$

Now

$$\begin{aligned} \|A_k - A\| &\leq \int_0^1 \| f'(x_{\ell+k} + \tau s_{\ell+k}) - f'(x^*) \| d\tau \\ &\leq \lambda \int_0 \| x_{\ell+k} + \tau s_{\ell+k} - x^* \| d\tau \\ &\leq \lambda \| x_{\ell+k} - x^* \| \end{aligned}$$

by (3.2) and (3.5), while $\|f_\ell\| \leq \gamma_7 h$ by (3.2), (3.5), and the definition of h, where $\gamma_7 = \|A\| + \lambda\delta/2$. Because of (3.5), (3.7), and the fact that $\|s_{\ell+k} s_{\ell+k}^T / (s_{\ell+k}^T s_{\ell+k})\| = 1$, we thus find $\|A_k - A\| \leq \lambda h$ and $\|(A_k - A) s_{\ell+k} s_{\ell+k}^T / (s_{\ell+k}^T s_{\ell+k})\| \, \|f_{\ell+k}\| \leq \lambda\gamma_7 h^2$. From (3.12a) for $j = k$, we thus find that the first term in the right-hand side (RHS) of (3.16) is bounded by $(\lambda\gamma_7 + \gamma_{1,k})h^2$. By (3.10) and (3.6), $\|A - \hat{B}_k\| \, \|f_{\ell+k}\| = \|A - \hat{B}_k\| \, \|B_{\ell+k} s_{\ell+k}\| \leq \gamma_8 \|s_{\ell+k}\|$, where $\gamma_8 = 4\delta(\|A\| + 2\delta)$. Moreover,

$$\left\| \frac{s_{\ell+k} s_{\ell+k}^T}{s_{\ell+k}^T s_{\ell+k}} - \frac{\hat{s}_k \hat{s}_k^T}{\hat{s}_k^T \hat{s}_k} \right\| = \left\| \frac{s_{\ell+k}}{\|s_{\ell+k}\|} \left(\frac{s_{\ell+k}}{\|s_{\ell+k}\|} - \frac{\hat{s}_k}{\|\hat{s}_k\|} \right)^T + \right.$$

$$\left. + \left(\frac{s_{\ell+k}}{\|s_{\ell+k}\|} - \frac{\hat{s}_k}{\|\hat{s}_k\|} \right) \frac{\hat{s}_k^T}{\|\hat{s}_k\|} \right\|$$

$$\leq 2 \left\| (s_{\ell+k}/\|s_{\ell+k}\| - \hat{s}_k/\|\hat{s}_k\|) \right\|$$

$$= [2/\|s_{\ell+k}\|] \left\| (s_{\ell+k} - \hat{s}_k) + [(\|\hat{s}_k\| - \|s_{\ell+k}\|)\hat{s}_k/\|\hat{s}_k\|] \right\|$$

$$\leq 4\|s_{\ell+k} - \hat{s}_k\|/\|s_{\ell+k}\| .$$

Because of (3.14), we therefore conclude that the second term in RHS(3.16) is bounded by $4\gamma_8\gamma_{4,k}h^2$, so (3.15) holds with $\gamma_{6,k} = \lambda\gamma_7 + \gamma_{1,k} + 4\gamma_8\gamma_{4,k}$. Thus (3.12a) holds for $j = k+1$, and by induction we see that (3.12) holds for $j = 2n$. But $\hat{x}_{2n} = x^*$ by Theorem 2.2, so (3.4) holds with $\gamma = \gamma_{1,2n}$. ∎

We could use the same techniques to prove a similar result for Broyden's bad method, i.e., $v_k = y_k/(y_k^T y_k)$ in (1.2e). At the time of this writing, it remains an open question whether a similar result holds for Broyden's method with projected updates [Gay & Schnabel, 1977].

4. Concluding Remarks

Theorem 2.2 came as quite a surprise to a number of us who had confidently shared the belief that Broyden's method did not enjoy finite termination on linear problems. Among other things, this theorem should serve to still the criticism that Broyden [1970, p. 377] had in mind when he wrote, "Thus

Broyden's algorithm will not solve linear systems in a finite number of steps and this has been held to be a disadvantage of the method."

Theorem 3.1 is one interesting consequence of Theorem 2.2: Broyden's method with unit step lengths enjoys local 2n-step Q-quadratic convergence and hence has an R-order of (local) convergence of at least $2^{1/(2n)}$ (see §9.2 of [Ortega & Rheinboldt, 1970]). This result nicely complements that of [Broyden, Dennis, & Moré, 1973], which establishes the local Q-superlinear convergence of Broyden's method. The Q-superlinear convergence assures that eventually more progress is made in the current iteration than in the previous one, while the 2n-step Q-quadratic convergence assures a definite amount of progress at intervals of no more than 2n iterations.

Theorem 2.4 suggests that Broyden's good (or bad) method often converges no faster than 2n-step Q-quadratically and hence has an R-order of exactly $2^{1/(2n)}$. If so, then we may extend the comparison of asymptotic efficiencies in §6 of [Brent, 1973] to include Broyden's method. According to Brent's definition, Broyden's method would have efficiency $E(B) = (\log_2 2)/(2n)$, the lowest of the methods compared. Of course, this says little for practical applications, where the bulk of execution time is consumed in finding the region of fast local convergence and where the simple measure of work (i.e., the number of equivalent function evaluations) that Brent used may not suffice. Moreover, if (as we suspect) a result similar to Theorem 3.1 holds for Broyden's method with projected updates [Gay & Schnabel, 1977], then this version of Broyden's method often enjoys (n+1)-step Q-quadratic convergence, which gives it efficiency $E(P) = (\log_2 2)/(n+1)$, the same as for the finite-difference Newton's method.

Acknowledgment

The author thanks John E. Dennis, Jr. both for helpful comments on a draft of this paper and for suggesting that Theorem 2.2 would probably imply the local 2n-step Q-quadratic convergence of the direct-prediction Broyden's method on nonlinear problems.

5. References

Brent, R.P. (1973), "Some Efficient Algorithms for Solving Systems of Nonlinear Equations", SIAM J. Numer. Anal. 10, pp. 327-344.

Broyden, C.G. (1965), "A Class of Methods for Solving Nonlinear Simultaneous Equations", Math. Comput. 19, pp. 577-593.

Broyden, C.G. (1970), "The Convergence of Single-Rank Quasi-Newton Methods", Math. Comput. 24, pp. 365-382.

Broyden, C.G. (1971), "The Convergence of an Algorithm for Solving Sparse Nonlinear Systems", Math. Comput. 25, pp. 285-294.

Broyden, C.G.; Dennis, J.E.; & Moré, J.J. (1973), "On the Local and Superlinear Convergence of Quasi-Newton Methods", J. Inst. Math. Appl. 12, pp. 223-245.

Davidon, W.C. (1959), "Variable Metric Method for Minimization", Argonne National Labs. report ANL-5990 Rev.

Fletcher, R.; & Powell, M.J.D. (1963), "A Rapidly Convergent Descent Method for Minimization", Comput. J. 6, pp. 163-168.

Gay, D.M.; & Schnabel, R.B. (1977), "Solving Systems of Nonlinear Equations by Broyden's Method with Projected Updates", NBER Working Paper No. 169.

Moré, J.J.; & Trangenstein, J.A. (1976), "On the Global Convergence of Broyden's Method", Math. Comput. 30, pp. 523-540.

Ortega, J.M.; & Rheinboldt, W.C. (1970), Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York.

Sherman, J.; & Morrison, W.J. (1949), "Adjustment of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix", Ann. Math. Statist. 20, p. 621.