Commentationes Mathematicae Universitatis Carolinae

J. M. Martínez
On the order of convergence of Broyden-Gay-Schnabel's method

Commentationes Mathematicae Universitatis Carolinae, Vol. 19 (1978), No. 1, 107--118

Persistent URL: http://dml.cz/dmlcz/105837

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE

19,1 (1978)

ON THE ORDER OF CONVERGENCE OF BROYDEN-GAY-SCHNABEL'S METHOD

J.M. MARTINEZ, Rio de Janeiro

Abstract: In a recent paper ([4]) Gay conjectured that Broyden's method with projected updates ([5]) has (n + 1) - - Q -quadratic convergence and R-order at least $2^{1/(n+1)}$. In this paper we prove that under certain conditions which are weaker than uniform linear independence, the method has R-order at least $t > 0$, with $t^{2n} - t^{2n-1} - 1 = 0$.

Key words: Nonlinear systems, Broyden's method, sequential secant method, order of convergence.

AMS: 65H10 Ref. Ž.: 8.314

---

1. Introduction. Broyden's method for solving nonlinear systems has been preferred by many authors for several years. This method solves systems without using derivatives (see [2]). Diverse modifications of his method resulted in robust, fast and efficient algorithm (see [7] and [9]).

The Q-superlinear convergence of this algorithm has been known for many years (see, for instance, [3]) but only in 1977 Gay ([4]) proved that it terminates in a finite number of steps: 2n, when dealing with linear systems, that it has 2n-Q-quadratic convergence and that its R-order of convergence is $2^{1/2n}$.

Also in 1977, Gay and Schnabel proposed a modification of the traditional updating used and proved for this method the (n + 1) steps - termination property for linear systems and Q-superlinear convergence, using the same techniques as Dennis-Moré ([3]).

The new update formula uses the projection of the last increment over the orthogonal subspace to the space generated by some of the previous increments. Hence, it is strongly related with the sequential secant method (see [1],[6], [8] and [10]), and it could be considered as a modification of it.

Gay conjectures in [4] that this method has (n + 1) - Q -quadratic convergence and thus that its R-order is $2^{1/(n+1)}$.

In the following sections we prove that the R-order of the method is at least the positive root of $t^{2n} - t^{2n-1} - 1 = 0$ (which is greater than $2^{1/(n+1)}$). In order to obtain this result we exploit the existing relationship between this method and the sequential secant method, and make use of a weaker assumption than the one generally used for proving its convergence.

2. Preliminary results. From now on $\|\cdot\|$ will always denote the 2-norm.

Let

(2.1.a) $$F: A \subset R^n \longrightarrow R^n,$$

A an open set,

(2.1.b) $$F \in C^1(A),$$

(2.1.c) $$F = (f_1, \ldots, f_n)^t$$

(2.2) $x^* \in A$, where $F(x^*) = 0$,

(2.3) $J(x) = (\partial f_i / \partial x_j)(x)$,

(2.4) $J(x^*)$ non-singular

and for all $x, y \in A$

(2.5) $\| J(x) - J(y) \| \leqq M \| x - y \|$, $M > 0$.

Suppose that we generate a sequence in R^n by the formulae:

(2.6.a) $x^0 \in A$, $B_0 \in R^{nxn}$,

(2.6.b) $x^{k+1} = x^k - B_k^{-1} F(x^k)$

with

(2.7) $B_{k+1} = B_k + \Delta B_k$

and

$$\Delta B_k = (\Delta F_k - B_k \Delta x_k) z_k^t / z_k^t \Delta x_k,$$

where

(2.9) $\Delta x_k = x^{k+1} - x^k$,

(2.10) $\Delta F_k = F(x^{k+1}) - F(x^k)$,

and

(2.11) $| z_k^t \Delta x_k | / \| z_k \| \, \| \Delta x_k \| \geq \sigma > 0$

for all $k = 0,1,2,\ldots$

In order to avoid extraneous hypotheses we assume

(2.12) $F(x^k) \neq 0$ and B_k invertible for all $k = 0,1,2,\ldots$,

thus, the above formulae are always well-defined.

The formulae (2.6) - (2.11) describe a "Broyden´s method with variable updating". When $z_k = \Delta x_k$ we obtain the original Broyden's method.

<u>Lemma 2.1</u>. With the assumptions (2.1),(2.3) and (2.5), we have that for every $x,\ y \in A$,

$$\| F(y) - F(x) - J(x)(y - x) \| \leq (M/2) \| y - x \|^{2}. \tag{2.13}$$

<u>Proof</u>. See [8].

<u>Lemma 2.2</u>. If (2.1),(2.3),(2.5) and (2.6) - (2.12) are assumed and $x^{k},\ x^{k+1} \in A$, then,

$$\| B_{k+1} - J(x^{k+1}) \| \leq (1 + 1/\sigma) \| B_k - J(x^{k}) \| + \\ + M(1 + 1/(2\sigma)) \| \Delta x_k \| . \tag{2.14}$$

<u>Proof</u>.

$$\| B_{k+1} - J(x^{k+1}) \| \leq \| B_{k+1} - B_k \| + \| B_k - J(x^{k}) \| + \\ + \| J(x^{k}) - J(x^{k+1}) \| , \tag{2.15}$$

and, by (2.13),

$$\| \Delta F_k - B_k \Delta x_k \| \leq \| \Delta F_k - J(x^{k}) \Delta x_k \| + \| J(x^{k}) \Delta x_k - \\ - B_k \Delta x_k \| \leq (M/2) \| \Delta x_k \|^{2} + \| J(x^{k}) - B_k \| \; \| \Delta x_k \| .$$

Then, by (2.7) - (2.11),

$$\| B_{k+1} - B_k \| = \| \Delta B_k \| \leq \| \Delta F_k - B_k \Delta x_k \| \; \| z_k \| / | z_k^{t} \Delta x_k | \\ (M \| \Delta x_k \|^{2}/2 + \| J(x^{k}) - B_k \| \; \| \Delta x_k \|)/(\sigma \| \Delta x_k \|).$$

Then,

$$\| B_{k+1} - B_k \| \leq (M/(2\sigma)) \| \Delta x_k \| + \| J(x^{k}) - B_k \| / \sigma . \tag{2.16}$$

But, by (2.5),

$$\| J(x^{k+1}) - J(x^{k}) \| \leq M \| \Delta x_k \| . \tag{2.17}$$

Finally, the desired result is obtained substituting (2.16) and (2.17) in (2.15).

The following lemmas show the relationship between the classical criteria of uniform linear independence and the tests used to detect singularities in the orthogonalization processes.

Lemma 2.3. Let $A = (v_1,\dots,v_n) \in R^{n\times n}$, $S_i = [v_1,\dots,v_i]$ the subspace generated by $v_1,\dots,v_i$; α_j the angle between S_j and v_{j+1}, $i = 1,\dots,n$, $j = 1,\dots,n-1$. Then,

$$|\det A| = |\sin \alpha_1|\dots|\sin \alpha_{n-1}| \, \|v_1\| \dots \|v_n\|. \tag{2.18}$$

Proof. If A is singular the proof is trivial. Otherwise, let Q and R be such that Q is orthogonal ($QQ^t = I$), R upper-triangular and $A = QR$. Let $R = (r_1,\dots,r_n) = (r_{ij})$, $i = 1,\dots$ $\dots,n$, $j = 1,\dots,n$; $r_i \in R^n$ for all $i = 1,\dots,n$. Then $|\det A| =$ $= |\det R|$ and $\|v_j\|^2 = \|r_j\|^2 = r_{1j}^2 + \dots + r_{jj}^2$, $j = 1,\dots,n$. Also,

$$S_i = [Qr_1,\dots,Qr_i] \text{ and}$$

$$v_{i+1} = QR_{i+1}.$$

Then, if β_i is the angle between $[r_1,\dots,r_i]$ and r_{i+1}, the orthogonality of Q implies $|\alpha_i| = |\beta_i|$, and, because R is triangular:

$$[r_1,\dots,r_i] = [e_1,\dots,e_i],$$

where e_k has a one in the k-th coordinate and zeros in the others.

Then

$$|\sin \alpha_i| = |\sin \beta_i| = |r_{i+1,i+1}| / \|r_{i+1}\| \text{ for all } i = 1,\dots,n-1,$$

and thus

$|\sin \alpha_1| \dots |\sin \alpha_{n-1}| = |r_{22} \dots r_{nn}| / (\|r_2\| \dots \|r_n\|) =$
$= |r_{11} \dots r_{nn}| / (\|r_1\| \dots \|r_n\|) = |\det R| / (\|r_1\| \dots \|r_n\|) =$
$= |\det A| / (\|v_1\| \dots \|v_n\|)$, and the thesis is proved.

Lemma 2.4. Let $\{A_\lambda\}_{\lambda \in L}$ be a family of non-singular $n \times n$ matrices, $A_\lambda = (v_1(\lambda), \dots, v_n(\lambda))$, and denote by $\alpha_i(\lambda)$ the angle between $[v_1(\lambda), \dots, v_i(\lambda)]$ and $v_{i+1}(\lambda)$, $i = 1, \dots, n-1$, $\lambda \in L$; then,

a) If $|\det A_\lambda| / (\|v_1(\lambda)\| \dots \|v_n(\lambda)\|) \geqq \varepsilon > 0$ for all $\lambda \in L$, then $|\sin \alpha_i(\lambda)| \geqq \varepsilon > 0$ for all $i = 1, \dots, n-1$, $\lambda \in L$.

b) If $|\sin \alpha_i(\lambda)| \geqq \delta > 0$ for all $i = 1, \dots, n-1$, $\lambda \in L$, then $|\det A_\lambda| / (\|v_1(\lambda)\| \dots \|v_n(\lambda)\|) \geqq \delta^{n-1}$ for all $\lambda \in L$.

Proof. Follows trivially from Lemma 2.3.

Now we state a generalization of Theorem 11.3.3 of [8]. We omit its proof which is entirely analogous to the one given in [8].

Lemma 2.5. Assuming (2.1),(2.3),(2.5) and $x_1, \dots, x_n \in A$; $p_1, \dots, p_n \in R^n$; $x \in A$; $x_i' = x_i + p_i \in A$ for all $i = 1, \dots, n$; $q_i = F(x_i') - F(x_i)$, $i = 1, \dots, n$;

$$|\det (p_1, \dots, p_n)| / (\|p_1\| \dots \|p_n\|) \geqq \varepsilon > 0;$$

$$B = (q_1, \dots, q_n)(p_1, \dots, p_n)^{-1};$$

then, there exists $K > 0$, which depends only on n and ε, such that

$$\|J(x) - B\| \leq K \max \{\|p_j\|/2 + \|x_j - x\|\}_{j=1}^{n}.$$

3. Gay-Schnabel's updating. Next we describe Gay-Schnabel's choice of the vector z_k in (2.8). We shall assume (2.1), (2.3),(2.6) - (2.10) and (2.12); and also let $m_k, k = 1,2,\ldots$ be a sequence of positive integers; $\hat{z}_k$, $k = 0,1,2,\ldots$ a sequence of vectors in R^n generated as follows:

(3.1.a) $$\hat{z}_0 = 0.$$

(3.1.b) If $|\hat{z}_k^t \Delta x_k| \geq \sigma \|\hat{z}_k\| \, \|\Delta x_k\|$, then $z_k = \hat{z}_k$ and $m_{k+1} = m_k + 1$. If $|\hat{z}_k^t \Delta x_k| < \sigma \|\hat{z}_k\| \, \|\Delta x_k\|$, then $z_k = \Delta x_k$ and $m_{k+1} = 1$.

(3.1.c) For each $k = 1,2,\ldots$; $\hat{z}_k$ is the orthogonal projection of Δx_k on $[\Delta x_{k-1},\ldots, \Delta x_{k-m_k}]^{\perp}$.

The estimate $|\hat{z}_k^t \Delta x_k| \gtrless \sigma \|\hat{z}_k\| \, \|\Delta x_k\|$ is used to determine if the new increment is dependent of the former m_k, with tolerance σ. The quotient $|\hat{z}_k^t \Delta x_k| / (\|z_k\| \, \|\Delta x_k\|)$ is the absolute value of the sine of the angle between Δx_k and $[\Delta x_{k-1},\ldots, \Delta x_{k-m_k}]$. Naturally, when m_k reaches n, this sine is 0, Δx_k is declared dependent and $z_k = \Delta x_k$, thus rederiving Broyden's classical formula. This is what distinguishes Gay-Schnabel's method from the sequential secant method.

Proposition 3.1. Using the definition (3.1) and the hypotheses which make it valid, we have $B_{k+1}\Delta x_{k-j} = \Delta F_{k-j}$ for all $j = 0,1,\ldots,m_{k+1} - 1$.

Proof. See [5].

Proposition 3.2. Assume (3.1) with the hypotheses which make it valid. If $x^k, x^{k+1},\dots,x^{k+n} \in A$ and

$$(3.2) \quad |\hat{z}_j^t \Delta x_j| \geqq \sigma \|\hat{z}_j\| \, \|\Delta x_j\| \text{ for all } j = k+1,\dots \\ \dots,k+n-1,$$

then there exists $K > 0$, independent of k, such that

$$\|B_{k+n} - J(x^{k+n})\| \leqq K(\|\Delta x_k\| + \dots + \|\Delta x_{k+n-1}\|).$$

Proof. By Proposition 3.1,

$$B_{k+n}\Delta x_{k+j} = \Delta F_{k+j}, \; j = 0,1,\dots,n-1.$$

By Lemma 2.4 and (3.2), $(\Delta x_k,\dots, \Delta x_{k+n-1})$ is non-singular,

$$B_{k+n} = (\Delta F_k,\dots, \Delta F_{k+n-1})(\Delta x_k,\dots, \Delta x_{k+n-1})^{-1}$$

and

$$|\det(\Delta x_k,\dots, \Delta x_{k+n-1})| / (\|\Delta x_k\| \dots \|\Delta x_{k+n-1}\|) \geqq \\ \geqq \sigma^{n-1},$$

then the thesis follows from Lemma 2.5.

Proposition 3.3. Assuming (2.1),(2.3),(2.5) - (2.12), (3.1),(3.2) and $x^{k+n+1},\dots,x^{k+n+s} \in A$; there exists $K_s > 0$ independent of k such that

$$\|B_{k+n+s} - J(x^{k+n+s})\| \leqq K_s(\|\Delta x_k\| + \dots + \|\Delta x_{k+n+s-1}\|).$$

Proof. For $s = 0$, we have Proposition 3.2. Assume the proposition to be true for $s - 1$. Then, by Lemma 2.2,

$$\|B_{k+n+s} - J(x^{k+n+s})\| \leqq (1 + 1/\sigma)\|B_{k+n+s-1} -$$

$$- J(x^{k+n+s-1})\| + M(1 + 1/(2\sigma))\|\Delta x_{k+n+s-1}\| \leqq$$

$$\leqq (1 + 1/\sigma)K_{s-1}(\|\Delta x_k\| + \dots + \|\Delta x_{k+n+s-2}\|) +$$

$$+ M(1 + 1/(2\sigma))\|\Delta x_{k+n+s-1}\| \leqq$$

$\leqq K_s(\| \Delta x_k \| + \ldots + \| \Delta x_{k+n+s-1} \|)$,

with $K_s = \max\{(1 + 1/\sigma)K_{s-1}, M(1 + 1/(2\sigma))\}$.

Proposition 3.4. Assume the hypotheses of Proposition 3.3 and that (2.19) holds for every k of the form i.n + p, with i = 1,2,...; and p is a fixed positive integer (that is, it is always possible to complete n successive linearly independent increments). We also assume that $x^k \in A$ for all k = 0,1,2,... . Then there exists $K > 0$ (independent of k) such that

$$\| B_k - J(x^k) \| \leqq K(\| \Delta x_{k-1} \| + \ldots + \| \Delta x_{k-2n+1} \|)$$

for all k = 2n - 1, 2n, 2n + 1,... .

Proof. Let $K = \max\{K_s\}_{s=1}^{n-1}$. In the set of integers k - 2n + 1,...,k - n there is necessarily one (j) of the form i.n + p. In the worst possible case it is k - 2n + 1. Applying Proposition 3.3 with j instead of k, we have:

$$\| B_{j+n+s} - J(x^{j+n+s}) \| \leqq K_s(\| \Delta x_j \| + \ldots + \| \Delta x_{j+n+s-1} \|).$$

Now, if j + n + s = k (then s = k - j - n),

$$\| B_k - J(x^k) \| \leqq K_{k-j-n}(\| \Delta x_j \| + \ldots + \| \Delta x_{k-1} \|).$$

But $j \geqq k - 2n + 1$, then $k - j - n \leqq n - 1$, and we obtain the desired result.

Theorem 3.1. Assume (2.1) - (2.12) and the hypotheses of the proposition 3.4. Let $E_k = x^k - x^*$. Then:

a) There exists $K > 0$, independent of k, such that

$$\| B_k - J(x^k) \| \leqq K(\| E_k \| + \ldots + \| E_{k-2n+1} \|).$$

b) If $\lim x^k = x^*$ (see [5] for conditions quaranteeing this hypotheses) then there exists $K > 0$ (independent of k) such that

$$\| E_{k+1} \| \leqq K \| E_k \| \; (\| E_k \| + \ldots + \| E_{k-2n+1} \|) \text{ for all } k \geqq 2n - 1.$$

c) With the hypotheses of b), x^k converges to x^* with R-order greater than or equal to the positive root of

$$t^{2n} - t^{2n-1} - 1 = 0.$$

Proof. Part a) follows immediately from Proposition 3.4. The derivation of b) is classical: If $x^{k+1} = x^k - B_k^{-1}F(x^k)$, then

$$\| F(x^{k+1}) - F(x^k) + J(x^k)B_k^{-1}F(x^k) \| \leqq \| B_k^{-1}F(x^k) \|^2, \text{ by Lemma 2.1.}$$

But $\lim E_k = 0$ implies that $\lim B_k = J(x^*)$ because of a), then $\lim B_k^{-1} = J(x^*)^{-1}$. Then, there exists $K_1 > 0$ such that

$$\| F(x^{k+1}) - F(x^k) + J(x^k)B_k^{-1}F(x^k)) \| \leqq K_1 \| F(x^k) \|^2.$$

Now, by Lemma 2.1, there exists $K_2 > 0$ such that

$$\| F(x^k) \| \leqq K_2 \| E_k \| \quad \text{for all } k = 0,1,2,\ldots$$

and part a) of the theorem implies:

$$\| F(x^{k+1}) \| \leqq K_3(\| E_k \| + \ldots + \| E_{k-2n+1} \|) \; \| E_k \| ,$$

with $K_3 > 0$.

As $J(x^*)$ is non-singular, then Lemma 2.1 guarantees the existence of $K_4 > 0$ satisfying

$$\| F(x^{k+1}) \| \geqq K_4 \| E_{k+1} \| \quad \text{for all } k = 0,1,2,\ldots,$$

then b) follows with $K = K_3/K_4$.

Finally c) is a consequence of b) because of Theorem 9.2.9 of [8].

4. Final remarks. Broyden's method with projected updates has an R-order of convergence apparently smaller than that of the sequential secant method. (The R-order of this is the positive root of $t^{n+1} - t^n - 1 = 0$.) Even so, other facts justify strongly the use of Broyden-Gay-Schnabel's method. Particularly, the modification proposed for the B_k is the one with the least norm among all which verify $B_{k+1}\Delta x_k = \Delta F_k, \dots, B_{k+1}\Delta x_{k-m_k} = \Delta F_{k-m_k}$ (see [5]). Thus, if B_k is periodically reset to a discretization of the jacobian matrix (or even if only B_0 is such a discretization and x^0 is close to x^*), the matrices $B_{k+1}, B_{k+2}, \dots$ are better approximations of the jacobian than the ones obtained by the sequential secant method. Generally speaking, we could say that the method is an adequate combination of the virtues of Broyden's traditional method and the sequential secant method.

References

[1] J.G.P. BARNES (1965), An algorithm for solving nonlinear equations based on the secant method, Computer Journal, 8, 66-72.

[2] C.G. BROYDEN (1965), A class of methods for solving nonlinear simultaneous equations, Mathematics of Computation, 19, 577-593.

[3] J.E. DENNIS and J.J. MORE (1974), A characterization of superlinear convergence and its application to quasi-Newton methods, Mathematics of Computation, 28, 549-560.

[4] D.M. GAY (1977), Some convergence properties of Broyden's method. Working Paper No 175, National Bureau

of Economic Research, USA.

[5] D.M. GAY and R.B. SCHNABEL (1977), Solving systems of non-linear equations by Broyden's method with projected updates, Working Paper No 169, National Bureau of Economic Research, USA.

[6] W.B. GRAGG and G.W. STEWART (1976), A stable variant of the secant method for solving nonlinear equations, SIAM J. of Numerical Analysis, 13, 127-140.

[7] J.J. MORE and J. TRAGENSTEIN (1976), On the global convergence of Broyden's method, Mathematics of Computation, 30, 523-540.

[8] J.M. ORTEGA and W.C. RHEINBOLDT (1970), Iterative solution of nonlinear equations in several variables, Academic Press, New York.

[9] M.J.D. POWELL (1970), A hybrid method for nonlinear equations, en Rabinovitz P. (editor), Numerical methods for nonlinear algebraic equations, Gordon and Breach, London.

[10] P. WOLFE (1959), The secant method for solving nonlinear equations, Communications ACM, 12, 12-13.

Universidade Candido Mendes

Rua Visconde de Piraja No 351, 4^o andar.

Ipanema - 20000 Rio de Janeiro, RJ

Brasil

(Oblatum 26.10.1977)