IMA Journal of Numerical Analysis (1984) **4**, 233–239

On the Least Q-order of Convergence of Variable Metric Algorithms

Y. YUAN

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW

[Received 12 May 1983 and in revised form 12 October 1983]

It is shown in this paper that the infimum of the Q-order of the convergence of variable metric algorithms is only 1, even though the objective function is twice continuously differentiable and uniformly convex. It is shown by example that the Q-order can be $1+1/N$ for any large N, though the R-order is $(1+N)^{1/2}$.

1. Introduction

THE PROBLEM is about the Q-order of convergence of variable metric algorithms for minimizing a differentiable function of several variables. Let $F(x)$ from $\mathbb{R}^n$ to $\mathbb{R}$ be the objective function to be minimized. Assume $F(x)$ is convex, twice continuously differentiable, and attains its minimum value at a point where $\nabla^2 F$ is positive definite. Variable metric algorithms are iterative. At the kth iteration an estimate of the solution at which $F(x)$ obtains its minimum, x_k say, and a $n \times n$ positive definite matrix B_k are available. Since $d_k = -B_k^{-1}\nabla F(x_k)$ is the solution of the following problem

$$\min F(x_k) + \nabla^T F(x_k)d + \tfrac{1}{2}d^T B_k d, \tag{1.1}$$

and to ensure a reduction in the objective function, we let

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

where the step-length α_k is such that

$$F(x_k + \alpha_k d_k) < F(x_k). \tag{1.3}$$

One special choice of α_k is called the "perfect" (or "exact") line search (Dixon, 1972), that is,

$$F(x_k + \alpha_k d_k) = \min_{\alpha} F(x_k + \alpha d_k). \tag{1.4}$$

Throughout this paper, the perfect line search is used at every iteration. B_{k+1} is updated by any formula in Broyden's family (Broyden, 1970). Due to Dixon (1972), if perfect line searches are used, then the sequence of points $\{x_k\}$ is independent of the choice of formula in Broyden's family, and we have

$$\delta_k^T \nabla F(x_{k+1}) = 0 \tag{1.5}$$

$$\delta_{k+1}^T [\nabla F(x_{k+1}) - \nabla F(x_k)] = 0,$$

0272–4979/84/020233+07 \$03.00/0

where

$$\delta_k = x_{k+1} - x_k = \alpha_k d_k. \tag{1.6}$$

Superlinear convergence of the algorithm, that is,

$$\lim_{n \to \infty} \|x_{k+1} - x^*\| / \|x_k - x^*\| = 0, \tag{1.7}$$

was first proved by Powell (1971, 1972) and Dixon (1972), where $\|.\|$ is any vector norm in $\mathbb{R}^n$ and x^* is the solution at which $F(x)$ attains its minimum value. Burmeister (1973) proved the n-step quadratic convergence condition

$$\|x_{k+n} - x^*\| = O(\|x_k - x^*\|^2), \tag{1.8}$$

and Ritter (1980) improved the result to

$$\|x_{k+n} - x^*\| = o(\|x_k - x^*\|^2). \tag{1.9}$$

Recently, Powell (1983) proved that

$$\|x_{k+n} - x^*\| = O(\|x_{k+1} - x^*\| \, \|x_k - x^*\|), \tag{1.10}$$

and when $n = 2$ his result is optimal. From (1.10), the R-order of convergence is at least the root in $(1, 2)$ of the polynomial equation

$$\theta^n - \theta - 1 = 0, \tag{1.11}$$

see Powell (1983). Powell (1983) also showed that the Q-order of convergence can be less than the R-order, where R-order and Q-order are defined by Ortega & Rheinboldt (1970). We establish that the infimum of the Q-order is only 1 for any fixed $n \geqslant 2$ even if $\nabla^2 F(x^*)$ is positive definite. However, if unit-steps ($\alpha_k = 1$) are used instead of perfect line searches, the R-order of convergence of the BFGS formula may also equal 1 (Powell, 1983).

2. The Result

Since the algorithm attains the solution after the first iteration when $n = 1$, we require $n \geqslant 2$. In this section, the result is stated and an outline of proof is given.

THEOREM 2.1. *For any fixed $n \geqslant 2$, the infimum of the Q-order of convergence of the variable metric algorithms with perfect line searches is only 1, if the class of objective functions is all convex, twice continuously differentiable functions, and the Hessian matrix at the minimum is positive definite.*

Outline of Proof. Obviously it is sufficient to prove that for any fixed $n \geqslant 2$ and any fixed $c > 1$, there exists a $F(x)$ from $\mathbb{R}^n$ to $\mathbb{R}$, which satisfies all the conditions stated in section 1, and there exist a $x_1 \in \mathbb{R}^n$ and a positive definite matrix B_1, which generate $\{x_k\}$, such that the Q-order of convergence of $\{x_k\}$ is less than c. The proof of the theorem is constructive, and the idea of the proof is due to Powell (1983).

It is sufficient to prove the theorem for $n = 2$. We construct a sequence $\{x_k\}$ converging to $x^* = (0, 0)^T$, which has the Q-order of $1 + 1/N$. It is proved that there exist a twice continuously differentiable function $F(.)$, a starting point x_1 and a positive definite matrix B_1 such that the variable metric method generates the sequence $\{x_k\}$. Since variable metric algorithms with exact line search are invariant under nonsingular linear transformations (Powell, 1971), we assume that $F(.)$ satisfies $\nabla^2 F(x^*) = 1$. Thus the angle between $\nabla F(x_{k+1})$ and $x_{k+1}-x^*$ tends to zero. The angle between δ_k and $x_k - x^*$ also tends to zero because $\{x_k\}$ converges superlinearly. Consequently from (1.5) it follows that the angle between $x_k - x^*$ and $x_{k+1}-x^*$ tends to a right angle. This guides us to define $\{x_k\}$ such that the iterates lie on two smooth curves that intersect perpendicularly at x^*:

$$\begin{aligned} x_{2k-1} &= (t_k^N, t_k^{N+1} - t_k^{N(N+1)-1})^T \\ x_{2k} &= (t_k^{N(N+1)} - t_k^{2N+2}, t_k^{N+1})^T \\ t_{k+1} &= t_k^{N+1} \quad k = 1, 2, \dots \end{aligned} \tag{2.1}$$

and $t_1 \in (0, 1)$, N being a very large integer. Since the conditions (1.5) and (1.6) define the sequence $\{x_k\}$ uniquely when $n = 2$ and $F(.)$ is convex, it is sufficient to prove that there exists a twice continuously differentiable function $F(.)$ satisfying (1.5)–(1.6) and (2.1). The Appendix shows the existence of $F(.)$, and a choice of x_1 and B_1 which make (1.5), (1.6) and (2.1) hold for all k.

From (2.1), we have that $\|x_{2k}\| \approx \|x_{2k-1}\|^{1+1/N}$. So for any $c > 1$, we can choose N such that $1+1/N < c$. Hence the Q-order is less than c. Therefore the theorem is true.

3. Discussion

Though we only prove our result when $n = 2$, for any $n > 2$ we can let $F_n(.)$ be

$$F_n(x) = F(t_1, t_2) + \tfrac{1}{2} \sum_{i=3}^{n} t_i^2, \quad x = (t_1, \dots, t_n)^T \in \mathbb{R}^n.$$

Further, we choose the initial point so that its first two components are those of x_1 defined in (2.1), and its other components are zero, and we let the initial matrix be

$$\begin{bmatrix} B_1 & O \\ O & I_{n-2} \end{bmatrix}.$$

Then the sequence generated by the variable metric algorithm is the same as x_k defined by (2.1), if we ignore the zero components of the points. Thus the Q-order is the same as that when $n = 2$, which shows our theorem holds for $n \geqslant 2$.

It is noted that in our example $\|x_{2k+1}\| \approx \|x_{2k-1}\|^{N+1}$, so the R-order of convergence is $(N+1)^{\frac{1}{2}}$. Therefore the R-order can be arbitrarily large, though the Q-order is arbitrarily close to 1.

Let Γ be the set of all twice continuously differentiable functions which solve (1.5)–(1.6), where $\{x_k\}$ is defined by (2.1). It is not known to the author whether Γ contains any many times differentiable functions. If there exists a many times differentiable function $F(.)$ which satisfies (1.5), (1.6) and (2.1), the given theorem is not analogous to the result, pointed out by the referee, that Newton's method without line search has the Q-order of $1+1/N$ if the Hessian is Holder continuous with exponent $1/N$ (Ortega & Rheinboldt, 1970). Otherwise for variable metric methods there might be some relations between the Q-order and the continuity properties of the Hessian.

The author is greatly indebted to Professor M. J. D. Powell for his constant help and encouragement, and for his studying the early drafts of the paper. He made many helpful suggestions which improved the paper greatly. The author also wishes to thank the referee for many valuable comments, including the suggestion of the lemma in the Appendix.

References

Broyden, C. G. 1970 The convergence of a class of double-rank minimization algorithms, Parts I and II. *J. Inst. Maths Applics* **6**, 76–90 and 222–236.

Burmeister, W. 1973 Die Konvergenzordordnung des Fletcher-Powell Algorithmus. *Z. Angew. Math. Mech.* **53**, 693–699.

Dixon, L. C. W. 1972 Variable metric algorithms: necessary and sufficient conditions for identical behaviour of nonquadratic functions. *J. Optim. Theory Appl.* **10**, 34–40.

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

Powell, M. J. D. 1971 On the convergence of the variable metric algorithm. *J. Inst. Maths Applics* **7**, 21–36.

Powell, M. J. D. 1972 Some properties of the variable metric method. In *Numerical Methods for Nonlinear Optimization*, (Lootsma, F. A., Ed.). London: Academic Press.

Powell, M. J. D. 1983 On the rate of convergence of variable metric algorithms for unconstrained optimization. Report DAMTP 1983/NA7, University of Cambridge.

Ritter, K. 1980 On the rates of superlinear convergence of a class of variable metric methods. *Num. Math.* **35**, 293–313.

Yuan, Y. 1983 On the least Q-order of convergence of variable metric algorithms. Report DAMTP 1983/NA10, University of Cambridge.

Appendix

Lemma A.1 *For $s \in \mathbb{R}$ and $t \in \mathbb{R}$, let $x(s)$, $g(s)$, $y(t)$, $h(t)$ be once differentiable paths in $\mathbb{R}^n$ such that at $t = 0 = s$*

$$x(0) = y(0), \qquad g(0) = 0 = h(0), \tag{A.1}$$

$x'(0)$, $y'(0)$ are linearly independent, $g'(0)$, $h'(0)$ are linear independent and

$$g'^T(0)y'(0) = h'^T(0)x'(0). \tag{A.2}$$

Then there exists a differentiable function $F(.)$ defined in a small neighbourhood of $x(0)$ such that

$$\nabla F(x(s)) = g(s), \qquad \nabla F(y(t)) = h(t) \tag{A.3}$$

holds for all small $s \geqslant 0$, $t \geqslant 0$.

Proof. Since $x'(0)$ and $y'(0)$ are linear independent, we can make a differentiable and invertible transformation in a small neighbourhood of $x(0)$ if necessary, in order to assume without loss of generality that $x(s) = (s, 0, 0, \ldots)^T = se_1$, $y(t) = (0, t, 0, \ldots)^T = te_2$. The transformation makes an adjustment to $g(s) = (g_1(s), g_2(s), \ldots)^T$ and $h(t) = (h_1(t), h_2(t), \ldots)^T$, but it is easy to show that (A.2) remains valid, which ensures that $g_2'(0) = h_1'(0)$. Let

$$F(z) = \int_0^{z_1} g_1(s)ds + \sum_{i=2}^{n} z_i g_i(z_1) + \int_0^{z_2} h_2(t)dt + \sum_{i\neq 2} z_i h_i(z_2) - g_2'(0) z_1 z_2 \tag{A.4}$$

$$z = (z_1, z_2, \ldots, z_n)^T \in \mathbb{R}^n.$$

It is straightforward to show that $F(.)$ is continuously differentiable and satisfies (A.3). The lemma is proved.

Our theorem is true if there exists a twice continuously differentiable function $F(.)$, which satisfies (1.5)–(1.6), where $\{x_k\}$ is defined by (2.1). Direct calculations show that

$$\delta_{2k-1} = (-t_k^N + t_k^{N(N+1)} - t_k^{2N+2}, t_k^{N(N+1)-1})^T \tag{A.5}$$

$$\delta_{2k} = (t_k^{2N+2}, -t_k^{N+1} + t_k^{(N+1)^2} - t_k^{(N+1)(N^2+N-1)})^T$$

for all $k = 1, 2, \ldots$. Hence if we let

$$\nabla F(x_{2k-1}) = (a(t_k), a(t_k)t_k/(1 - t_k^N + t_k^{N^2+N-2}))^T \tag{A.6}$$

$$\nabla F(x_{2k}) = (b(t_k)t_k^{N^2-1}/(1 + t_k^{N+2} - t_k^{N^2}), b(t_k))^T$$

then the first equation of (1.5) is satisfied. To ensure that the second equation of (1.5) is satisfied, we require that

$$t_k^{N+1}[b(t_k)t_k^{N^2-1}/(1 + t_k^{N+2} - t_k^{N^2}) - a(t_k)] +$$

$$(-1 + t_k^{N(N+1)} - t_k^{(N+1)(N^2+N-2)})[b(t_k) - a(t_k)t_k/(1 - t_k^N + t_k^{N^2+N-2})] = 0 \tag{A.7}$$

and

$$(-1 - t_k^{(N+1)(N+2)} + t_k^{(N+1)N^2})[a(t_k^{N+1}) - b(t_k)t_k^{N^2-1}/(1 + t_k^{N+2} - t_k^{N^2})] +$$

$$1 - t_k^{(N+1)(N^2-1)}[t_k^{N+1}a(t_k^{N+1})/(1 - t_k^{(N+1)N} + t_k^{(N+1)(N^2+N-2)}) - b(t_k)] = 0, \tag{A.8}$$

for all k. Hence if $a(t)$ and $b(t)$ satisfy

$$b(t) = ta(t)[(1 - t^{(N+1)N} + t^{(N+1)(N^2+N-2)})/(1 - t^N + t^{N^2+N-2}) - t^N]/$$

$$[1 - t^{(N+1)N} + t^{(N+1)(N^2+N-2)} - t^{(N+1)N}/(1 + t^{N+2} - t^{N^2})] \tag{A.9}$$

and

$$a(t^{N+1}) = t^{N^2-1}b(t)[(1 + t^{(N+1)(N+2)} - t^{N^2(N+1)})/(1 + t^{N+2} - t^{N^2}) - t^{(N^2-1)N}]/$$

$$[1 + t^{(N+1)(N+2)} - t^{(N+1)N^2} - t^{(N+1)N^2}/(1 - t^{N(N+1)} + t^{(N+1)(N^2+N-2)})], \tag{A.10}$$

then (A.7) and (A.8) hold. Eliminating $b(t)$ from (A.9) and (A.10) we have the following form of function equation

$$a(t^{N+1}) = t^{NL} \frac{\prod_{j=1}^{m}\left(1 + \sum_{l=1}^{I_j} a_{jl} t^{I_{jl}}\right)}{\prod_{j=1}^{m'}\left(1 + \sum_{l=1}^{J_j} b_{jl} t^{J_{jl}}\right)} a(t) \tag{A.11}$$

where a_{jl}, b_{jl} and I_{jl}, J_{jl} are constants and positive integers, respectively. A special solution of this equation is as follows:

$$a(t) = t^L \prod_{k=0}^{\infty} \frac{\prod_{j=1}^{m'} \left(1 + \sum_{l=1}^{J_j} b_{jl} t^{J_{jl}(N+1)^k}\right)}{\prod_{j=1}^{m} \left(1 + \sum_{j=1}^{I_j} a_{jl} t^{I_{jl}(N+1)^k}\right)} \tag{A.12}$$

for all sufficiently small t. Therefore we define $a(t)$ by the equation

$$a(t) = t^N(1 - t^N + t^{N^2+N-2}) \times \prod_{j=0}^{\infty} \tag{A.13}$$

$$\frac{[(1 + t^{(N+2)(N+1)^{j+1}} - t^{N^2(N+1)^{j+1}})(1 - t^{N(N+1)^{j+1}} + t^{(N^2+N-2)(N+1)^{j+1}}) - t^{N^2(N+1)^{j+1}}]}{[1 + t^{(N+2)(N+1)^{j+1}} - t^{N^2(N+1)^{j+1}} - t^{N(N^2-1)(N+1)^j}(1 + t^{(N+2)(N+1)^j} - t^{N^2(N+1)^j})]} \times$$

$$\frac{[(1 - t^{N(N+1)^{j+1}} + t^{(N^2+N-2)(N+1)^{j+1}})(1 + t^{(N+2)(N+1)^j} - t^{N^2(N+1)^j}) - t^{N(N+1)^{j+1}}]}{[1 - t^{N(N+1)^{j+1}} + t^{(N^2+N-2)(N+1)^{j+1}} - t^{N(N+1)^j}(1 - t^{N(N+1)^j} + t^{(N^2+N-2)(N+1)^j})]}$$

for small t and then we let $b(t)$ have the value (A.9), in order that (A.9) and (A.10) are satisfied. Thus if $F(x)$ satisfies (A.6), then (1.5) holds for all k.

From (A.13) and (A.9), we have that $a(t)$ and $b(t)$ are analytic functions and that

$$\begin{aligned} a(t) &= t^N(1 - t^N + O(|t|^{N^2}))(1 + t^{N+2} + O(|t|^{N^2}))/(1 - t^N + t^{2N} + O(|t|^{N^2})) \\ &= t^N(1 + t^{N+2} - t^{2N} + O(|t|^{3N})), \\ b(t) &= ta(t)[1 + t^N + t^{2N} + O(|t|^{3N}) - t^N]/(1 + O(|t|^{N^2})) \\ &= t^{N+1}(1 + t^{N+2} + O(|t|^{3N})) \qquad |t| < 1. \end{aligned} \tag{A.14}$$

For sufficiently small non-negative t, we define

$$\begin{aligned} \phi(t) &= a(t^{1/N}), \\ \psi(t) &= b(t^{1/(N+1)}). \end{aligned} \tag{A.15}$$

From (A.14), we have

$$\begin{aligned} \phi(t) &= t + \phi^*(t) \\ \psi(t) &= t + \psi^*(t), \end{aligned} \tag{A.16}$$

where ϕ^*, ψ^* are defined for small non-negative t, and

$$\begin{aligned} \phi^*(t) &= t^{2+2/N} + O(|t|^3), \\ \psi^*(t) &= t^{2+1/(N+1)} + O(|t|^3). \end{aligned} \tag{A.17}$$

Further, ϕ^* and ψ^* are twice continuously differentiable for small non-negative t, and

$$\begin{aligned} \frac{d^2}{dt^2} \phi^*(t) &= O(|t|^{2/N}), \\ \frac{d^2}{dt^2} \psi^*(t) &= O(|t|^{1/N+1}). \end{aligned} \tag{A.18}$$

Hence if we define

$$\phi^*(t) = \phi^*(-t) \tag{A.19}$$
$$\psi^*(t) = \psi^*(-t)$$

for sufficiently small negative t, then ϕ^* and ψ^* are well defined and twice continuously differentiable in a small neighbourhood of zero. Thus, by (A.16), ϕ and ψ can be well defined and twice continuously differentiable in a small neighbourhood of zero. Further, for positive t, (A.15) holds.

What we need to show is the existence of $F(t, u)$ such that

$$\nabla F(t, t^{1+1/N} - t^{N+1-1/N}) = [\phi(t), t^{1/N}\phi(t)/(1 - t + t^{N+1-2/N})]^T, \tag{A.20}$$
$$\nabla F(u^N - u^2, u) = [u^{N-1}\psi(u)/(1 + u^{1+1/(1+N)} - u^{N2/(N+1)}), \psi(u)]^T.$$

From Lemma A.1, given at the beginning of the Appendix, there exists a continuously differentiable function $F(.)$ such that (A.20) holds. In our special case, $F(.)$ can be made twice continuously differentiable. For more details see Yuan (1983).

From (2.1), (A.6) and (A.14), it follows that $\nabla^2 F(x^*) = I$, where $x^* = (0, 0)^T$. Therefore $F(.)$ is not only twice continuously differentiable but also uniformly convex in a neighbourhood of x^*. We may modify F if necessary away from the origin, and then choose $t_1 > 0$ small so that the modification is irrelevant to the above analysis. Therefore the condition $\delta_k^T \nabla F(x_{k+1}) = 0$ is sufficient for x_{k+1} to be the point that would be calculated by a perfect line search. Because $\delta_1^T \nabla F(x_1) < 0$ for small $t_1 > 0$, there exists a positive definite matrix B_1 such that $\delta_1 = -B_1^{-1} \nabla F(x_1)$. Thus the variable metric algorithm generates exactly the sequence $\{x_k\}$ as defined in (2.1), whose Q-order of convergence is $1 + 1/N$.