Journal of Computational Mathematics, Vol.12, No.4, 1994, 366–379.

A CLASS OF MODIFIED BROYDEN ALGORITHMS*1)

Pu Ding-guo
(Department of Mathematics, Shanghai Institute of Railway Technology, Shanghai, China)
Tian Wei-wen
(Department of Mathematics, Shanghai University of Science and Technology, China)

Abstract

In this paper we discuss the convergence of the modified Broyden algorithms. We prove that the algorithms are globally convergent for the continuous differentiable function and the rate of convergence of the algorithms is one-step superlinear and n-step second-order for the uniformly convex objective function. From the discussion of this paper, we may get some convergence properties of the Broyden algorithms.

1. Introduction

We know that the variable metric algorithms, such as the Broyden algorithms, are very useful and efficient methods for solving the nonlinear programming problem

$$min\{f(x); x \in R^n\}. \tag{1.1}$$

With exact linear search, Powell(1971) proves that the rate of convergence of these algorithms is one-step superlinear for the uniformly convex objective function, and if the points given by these algorithms are convergent, Pu and Yu(1990) prove that they are globally convergent for the continuous differentiable function. Without exact linear search several results have been obtained. Powell(1976) demonstrates that the convergence rate of the BFGS algorithms without exact linear search is one-step superlinear. Byrd, Nocedal and Yuan(1987) prove that the above result is also true for other Broyden algorithms except the DFP algorithms. Pu(1990, 1992 and 1993) proves that the convergence rate of the prime DFP algorithms without exact linear search is one-step superlinear for the modified Wolfe conditions.

However there are several theoretical problems which have not been solved for the Broyden algorithms today, and some numerical results show that the points given by the

---

* Received July 10, 1993.
1) The Project Supported by National Natural Science Foundation of China.

Broyden algorithms may not converge to the optimal point for the objective function without convexity(Fletcher(1987)). Several modified variable metric algorithms are proposed for solving those problems and increasing the speed of convergence. In this paper we propose a new class of variable metric algorithms called modified Broyden algorithms which is generalized the idea in Pu's (1989a) short paper and prove the algorithms are convergent for the continuous differentiable objective functions, and superlinear and n-step second order convergent for the uniformly convex functions when the linear search is exact.

The modified Broyden algorithms are iterative. Given a starting point x_1, an initial positive definite matrix B_1 and a constant $\phi \in [0,1]$, it generates a sequence of points $\{x_k\}$ and a sequence of matrices of $\{B_k\}$, satisfying (1.2) and (1.3)

$$x_{k+1} = x_k + s_k = x_k - \alpha_k B_k^{-1} g_k \tag{1.2}$$

where $\alpha_k \geq 0$ is the step factor, and g_k is the gradient of f(x) at x_k. If $g_k = 0$, the algorithm terminates, otherwise let

$$B_{k+1} = \tilde{B}_{k+1} - \frac{p_{k+1}\tilde{B}_{k+1}R_{k+1}g_{k+1}g_{k+1}^T R_{k+1}\tilde{B}_{k+1}}{1 + p_{k+1}g_{k+1}^T R_{k+1}\tilde{B}_{k+1}R_{k+1}g_{k+1}} \tag{1.3}$$

where p_{k+1} is a positive real number

$$p_{k+1} = \frac{\|Q_{k+1}\tilde{B}_{k+1}^{-1}g_{k+1}\|}{g_{k+1}^T R_{k+1}g_{k+1}}, \tag{1.4}$$

where $\{Q_{k+1}\}$ and $\{R_{k+1}\}$ are two sequences of positive matrices which are uniformly bounded. All eigenvalues of these matrices are included in [q,r], $0 < q \leq r$, And the $\tilde{B}_{k+1}$ is given by

$$\tilde{B}_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} + \phi(s_k^T B_k s_k) v_k v_k^T, \tag{1.5}$$

where $y_k = g_{k+1} - g_k, v_k = y_k(y_k^T s_k)^{-1} - B_k s_k (s_k^T B_k s_k)^{-1}$. In above programming if B_k are taken $\tilde{B}_k$ for all k, we get the Broyden algorithms. And if $\phi = 0$ we call it modified BFGS algorithms, or abbreviated by MBFGS and if $\phi = 1$ we call it modified DFP algorithms, or MDFP algorithms for short.

The matrix H_k and $\tilde{H}_k$ are denoted the inverses of B_k and $\tilde{B}_k$, we may obtain the Quasi-Newton formula $\tilde{H}_{k+1}y_k = s_k$ by the Broyden algorithms. And Pu(1989a) gave

$$H_{k+1} = \tilde{H}_{k+1} + p_{k+1}R_{k+1}g_{k+1}g_{k+1}^T R_{k+1}. \tag{1.6}$$

From the Broyden algorithms we know that if H_k is positive definitive the $\tilde{H}_{k+1}$ is also positive definitive. H_{k+1} may be implied positive. To use the mathematical induction it is easy to imply the $\tilde{B}_{k+1}, B_{k+1}, \tilde{H}_{k+1}$ and H_{k+1} are positive definitive matrices by the H_1 and B_1 being.

In section 2 the global convergence and several results without the convexity assumption are given; in section 3 some results for the convex objective function are shown; in section 4 and section 5 it is proved that the algorithms are linear convergent for $\phi \in [0,1)$ and $\phi = 1$ respectively; in section 6 one-step superlinear convergence is presented; in section 7 the results of the second order convergence, several numerical examples and a short discussion are stated.

If the linear search is exact we know, for all k, $g_{k+1}^T H_k g_k = 0$ and $y_k^T H_k y_k = g_k^T H_k g_k + g_{k+1}^T H_k g_{k+1}$.

In the case that the ambiguities are not given we may omit the index of character, for example, g, x, R denote g_k, x_k, R_k, and g_*, x_*, R_* denote $g_{k+1}, x_{k+1}, R_{k+1}$ and etc. respectively. For simplicity let

$$U_k = \frac{g_{k+1}^T H_{k+1} g_{k+1}}{g_k^T H_k g_k}; \quad V_k = \frac{y_k^T H_k y_k}{g_k^T H_k g_k}; \quad W_k = \frac{g_{k+1}^T H_k g_{k+1}}{g_k^T H_k g_k}; \quad \eta_k(\phi) = 1 + \phi W_k. \quad (1.7)$$

2. The Global Convergence and Several Results Without Convexity Assumption

In this section we assume:

1. $f(x) \in C^{1,1}$, i.e. there exists $L > 0$ such that, for all $x, y \in R^n$, $\|g(x) - g(y)\| \leq L\|x - y\|$.
2. For all $x_1 \in R^n$, the lever set $S(x_1) = \{x \mid f(x) \leq f(x_1)\}$ is bounded.
3. $f(0) = min f(x) = 0$.

The recurrence formula (1.6) of H and the positive property of R and Q imply

$$(q^{-1} + r\|g\|)\|Q\tilde{H}g\| \geq \|Hg\| \geq (r^{-1} - r\|g\|)\|Q\tilde{H}g\|. \quad (2.1)$$

On the other hand, the definitions of H and p imply

$$g^T Hg = g^T \tilde{H} g + p(g^T Rg)^2 = g^T \tilde{H} g + \|Q\tilde{H}g\|(g^T Rg) \geq \frac{(q\|g\|)^2 \|Hg\|}{1 + qr\|g\|}. \quad (2.2)$$

Assumptions 1, 2, and (2.2) imply

$$L\|s\| \geq \frac{-g^T s}{\|s\|} \geq \frac{(q\|g\|)^2}{1 + qr\|g\|} \geq \frac{\|g\|}{2} min\left\{q^2\|g\|; \ \frac{q}{r}\right\}. \quad (2.3)$$

Theorem 2.1 *The algorithm is globally convergent, i.e.*

$$\lim_{k\to\infty} g_k = 0. \quad (2.4)$$

Proof. Suppose the theorem is not true, there exists a $\varepsilon > 0$ such that $\|g_k\| \geq \varepsilon > 0$ for infinitely large k. The $f(x_k)$ is bounded below because the lever set $S(x_1)$ is bounded. This implies

$$\lim_{k\to\infty} (f(x_k) - f(x_{k+1})) = 0. \tag{2.5}$$

But (2.3) implies that, for such kind of k, $\|g_k\| \geq \varepsilon$, there exists a $C > 0$ such that (Powell 1972)

$$f(x_k) - f(x_{k+1}) \geq \frac{C(-g_k^T s_k)^2}{\|s\|^2} \geq \frac{C\varepsilon^2}{4} min\{q^4\varepsilon^2; \quad \frac{q^2}{r^2}\} > 0. \tag{2.6}$$

The contradiction between (2.5) and (2.6) leads to the theorem.

Several equations and inequalities below are implied without assuming the convexity assumption of objective function. Taking the trace of both sides of (1.3) and (1.5), we get

$$tr(\tilde{B}_*) = tr(\tilde{B}) + \frac{\|y\|^2}{y^Ts} + \frac{\phi\|y\|^2 s^TBs}{(y^Ts)^2} - \frac{2\phi y^TBs}{y^Ts} - (1-\phi)\frac{\|Bs\|^2}{s^TBs} - \frac{p\|\tilde{B}Rg\|^2}{1+pg^TR\tilde{B}Rg}. \tag{2.7}$$

Let

$$\mu = Hg_* + WHg = VHg + Hy. \tag{2.8}$$

Computing directly, we get

$$\tilde{H}_* = (I - \frac{sy^T}{y^Ts})H(I - \frac{ys^T}{y^Ts}) + \frac{ss^T}{y^Ts} - \frac{\phi\mu\mu^T}{g^THg\eta(\phi)}, \tag{2.9}$$

where $\eta(\phi)$ as the definition by (1.7). (2.8) and (2.9) imply

$$\tilde{H}_*g_* = \left(1 - \frac{\phi W}{\eta(\phi)}\right)\mu = \frac{\mu}{\eta(\phi)} \tag{2.10}$$

and

$$Hy = \mu - VHg = \eta(\phi)\tilde{H}_*g_* - VHg. \tag{2.11}$$

Taking the trace of both sides of (2.9), we obtain

$$tr(\tilde{H}_*) = tr(H) - \frac{-2g^THHy}{g^THg} + \frac{V\|Hg\|^2}{g^THg} - \frac{\phi\|\mu\|^2}{g^THg\eta(\phi)} + \frac{\|s\|^2}{y^Ts}. \tag{2.12}$$

Taking the trace of both sides of (1.6), the following inequality is implied by (1.4)

$$tr(H) = tr(\tilde{H}) + p\|Rg\|^2 \leq tr(\tilde{H}) + r^2\|\tilde{H}g\| \tag{2.13}$$

and there exists a $C_1 > 0$ such that, for all k,

$$C_1 g^T Hg \geq \|Hg\|\|g\|^2; \ \ C_1\|\tilde{H}g\|^2 \geq Hgg^T H\|^2 \ and \ C_1\|Hg\| \geq r^2\|\tilde{H}g\|. \qquad (2.14)$$

3. Some Results for the Convex Objective Function

In this section we assume

1. The objective function f(x) is uniformly convex and there exist M and m, $M \geq m > 0$ such that, for all $x, y \in R^n$, $m\|x\|^2 \leq x^T G(y) x \leq M\|x\|^2$ where $G(y)$ is the second-order Hessian of $f(x)$ at y.
2. G(x) satisfies the Lipschitz condition, i.e. there exists a $L > 0$ such that, $\|G(x) - G(y)\| \leq L\|x - y\|$ for all $x, y \in R^n$.
3. $f(0) = min f(x) = 0 \ and \ G(0) = I_{n\times n}$, the identity matrix of order n.

In the assumption 3, $G(0) = I_{n\times n}$ is equal to a linear affine transfer for the objective function and does not affect the results in this paper. By Byrd et al (1987, p1175), there exists $C_2 > 0$ such that, for all k,

$$f(x_{k+1}) \leq (1 - C_2 \cos^2 \beta_k) f(x_k) \qquad (3.1)$$

where β_k denotes the angle between g_k and $H_k g_k$. The sequence of functions $\{f(x_k)\}$ is a monotonically decreasing sequence of k. We get, for all k and $i > 0$,

$$M\|x_k\|^2 \geq 2f(x_k) \geq 2f(x_{k+i}) \geq m\|x_{k+i}\|^2. \qquad (3.2)$$

Let $\tilde{G}_k = \tilde{G}(x_k) = \int_0^1 G(x_k + ts_k)dt$, $C_3 = L\sqrt{\frac{M}{m}}(1 + \frac{1}{m})$ and $(\tilde{G})^{-1}$ denotes the inverse of $\tilde{G}$, then the assumption 2 and (3.2) imply, for all k,

$$\|I - \tilde{G}\| = \|G(0) - \tilde{G}\| \leq C_3\|x\|, \quad \|I - \tilde{G}^{-1}\| \leq \|G^{-1}\|\|I - \tilde{G}\| \leq C_3\|x\|. \qquad (3.3)$$

Because of $y = \tilde{G}s$, $\|y\|^2 - y^T s = s^T(\tilde{G})^{\frac{1}{2}}(\tilde{G} - I)(\tilde{G})^{\frac{1}{2}} s$, we get

$$\max\{m; \ 1 - C_3\|x\|\} \leq \frac{\|y\|^2}{y^T s} \leq min\{M; \ 1 + C_3\|x\|\} \qquad (3.4)$$

and

$$\max\{\frac{1}{M}; \ 1 - C_3\|x\|\} \leq \frac{\|s\|^2}{y^T s} \leq min\{\frac{1}{m}; \ 1 + C_3\|x\|\}. \qquad (3.5)$$

The Quasi-Newton formula $\tilde{H}_* y = s$ and (3.3) imply $|g_*^T \tilde{H}_* Hg| = |\alpha^{-1} g_*^T \tilde{H}_*(\tilde{G}^{-1} - I)y| \leq C_3\|x\|\|\tilde{H}_* g_*\|\|Hg\|$. From the uniform convexity of f(x) and (3.5), we obtain $|g^T y - g^T s| = |g^T(\tilde{G} - I)s| \leq C_3\|x\|\|s\|\|g\|$ and

$$\frac{y^T B s}{y^T s} \geq \alpha - \frac{C_3\|g\|\|s\|\|x\|}{g^T H g}. \tag{3.6}$$

We know that the function $g(x) = \frac{x}{x+a}$, $a > 0$ is monotonically increasing, and $p(g^T Rg)^2 = \|Q\tilde{H}g\|g^T Rg \geq q^2\|\tilde{H}g\|\|g\|^2$, the Cauchy-Schwarz inequality implies

$$\frac{p\|\tilde{B}Rg\|^2}{1+pg^T R\tilde{B}Rg} \geq \frac{p(g^T Rg)^4}{\|Rg\|^2[g^T\tilde{H}g + p(g^T Rg)^2](g^T\tilde{H}g)} \geq \frac{q^3\|g\|^3}{rg^T Hg} \geq \frac{mq^3\|x\|\|g\|^2\|s\|^2}{\alpha r(g^T Hg)^2}. \tag{3.7}$$

Let $C_4 = CC_3^2 rq^{-3}m^{-1}$, for (3.7) and the inequality $a^2 + b^2 \geq 2|ab|$, we have that

$$\frac{p\|\tilde{B}Rg\|^2}{C(1+pg^T R\tilde{B}Rg)} + C_4\alpha\|x\| - \frac{C_3\|g\|\|s\|\|x\|}{g^T Hg} \geq 0 \tag{3.8}$$

holds.

Because of $g_k = \int_0^1 G(tx_k)dtx_k$, for all k and $i > 0$, the following holds

$$m\|x\| \leq \|g\| \leq M\|x\|, \qquad m^3\|g_{k+i}\|^2 \leq M^3\|g_k\|^2. \tag{3.9}$$

4. The linear convergence of algorithms for $\phi \in [0,1)$

In this section we assume the assumptions 1-3 in section 3 hold. The convergence rate for the MDFP algorithms, i.e. $\phi = 1$, is left over until next section because there is some different in proofs of convergence between the MDFP algorithms and other Broyden algorithms. In this section, we only discuss the linear convergence rate of our algorithms as $\phi \in [0,1)$. First we prove the following lemma for $\phi \in [0,1)$.

Lemma 4.1. *For any given $\phi \in [0,1]$ there exists a $C_5 > 0$ such that, for all k,*

$$\sum_{j=1}^{k} \alpha_j^{-1} \leq C_5 k. \tag{4.1}$$

Proof. The Quasi-Newton formula implies $g_*^T\tilde{H}_* y = 0$, (3.3) implies

$$V\|Hg - \phi\tilde{H}_* g_*\|^2 \geq \frac{V(g^T Hg)^2}{\|y\|^2} \geq \frac{Vg^T Hg}{\alpha}(1 - C_3\|x\|). \tag{4.2}$$

From (2.10), (2.11), (4.2) and the definition (1.7) of $\eta(\phi)$, we get, for all $\phi \in [0,1]$,

$$\begin{aligned}
&-2g^T HHy - V\|Hg\|^2 + \phi\eta^{-1}(\phi)\|\mu\|^2 = V\|Hg\|^2 - 2\eta(\phi)g^T H\tilde{H}_* g_* + \phi\eta(\phi)\|\tilde{H}_* g_*\|^2 \\
&= V\|Hg - \phi\tilde{H}_* g_*\|^2 - 2(1-\phi)g^T H\tilde{H}_* g_* + \phi(1-\phi)\|\tilde{H}_* g_*\|^2 \\
&\geq \frac{Vg^T Hg}{\alpha}\left[1 - \frac{C_3(1+m)\|x\|}{m}\right] + [(1-\phi)(\phi - C_3\|x\|)]\|\tilde{H}_* g_*\|^2.
\end{aligned} \tag{4.3}$$

If $\phi = 0$ then $\eta(0) = 1$, we consider the proof of two cases for $\phi = 0$. In the first case, if $\alpha_*\|\tilde{H}_* g_*\| \leq \frac{4r^2\|g_*\|^2}{m}$, then (2.14) implies

$$\|\tilde{H}_* g_*\|\|Hg\| \leq \frac{4r^2\|g_*\|^2\|Hg\|}{\alpha_* m} \leq \frac{4r^2 M^3\|Hg\|\|g\|^2}{\alpha_* m^4} \leq \frac{4C_1 r^2 M^3 g^T Hg}{\alpha_* m^4}. \tag{4.4}$$

In the second case, for $\alpha_*\|\tilde{H}_* g_*\| \geq \frac{4r^2\|g_*\|^2}{m}$, if $r^2\|g\| < 1$, (2.2) implies

$$\alpha_*\|\tilde{H}_* g_*\|^2 \leq 2\alpha_*\|H_* g_*\|^2 \leq \frac{2g_*^T H_* g_*}{m} \leq \frac{2g_*^T \tilde{H}_* g_*}{m} + \frac{\alpha_*\|\tilde{H}_* g_*\|^2}{2}. \tag{4.5}$$

Multiplying both sides of (2.9) by g_*, then taking the inner product for the result previous with g_*, we get $g_*^T \tilde{H}_* g_* = g_*^T H g_*$ as $\phi = 0$ and

$$2\|Hg\|\|\tilde{H}_* g_*\| \leq \frac{\alpha\|Hg\|^2}{\alpha_*} + \frac{\alpha_*\|\tilde{H}_* g_*\|^2}{\alpha} \leq \frac{g^T Hg}{m\alpha_*} + \frac{4y^T Hy}{\alpha}. \tag{4.6}$$

Let $C_6 = 4C_3[2r^2 C_1 M^3 m^{-4} + 1 + m^{-1}]$, combining (4.4) with (4.6) gives for $\phi = 0$

$$\frac{2C_3\|Hg\|\|\tilde{H}_* g_*\|}{g^T Hg} \leq C_6\left(\frac{1}{\alpha_*} + \frac{V}{\alpha}\right). \tag{4.7}$$

Given any $\phi \in (0, 1]$, $\phi - C_3\|x_k\| \geq 0$ holds for sufficient large k. Substituting (4.3) into (2.12) for $\phi \in (0, 1]$, and substituting the second equation of (4.3), (4.2) and (4.7) into (2.12) for $\phi = 0$, then using (2.13) and (2.14) the following (4.8) holds, for sufficiently large k and any given $\phi \in [0, 1]$

$$\begin{aligned} tr(\tilde{H}_*) + &\frac{V[1 - (C_3 + C_3 m^{-1} + C_6)\|x\|]}{\alpha} - \frac{C_6\|x\|}{\alpha_*} \\ \leq &tr(H) + \frac{\|s\|^2}{y^T s} \leq tr(H) + 1 + C_3\|x\| + C_1\|Hg\|. \end{aligned} \tag{4.8}$$

$\alpha\|d\| \leq C_3\|x\|$ is obtained by the definition of C_3. Let $C_7 = C_3(1 + m^{-1} + C_1) + 2MC_6$, We may assume $1 \geq 2C_7\|x\|$ for all k. For any given $\phi \in [0, 1]$, $\alpha M g^T H g\, tr(H) \geq \alpha M g^T H H g \geq \alpha M\|d\|^2 \geq g^T Hg$ and (4.8) imply

$$\begin{aligned} \sum_{j=1}^{k} \frac{V_j}{2\alpha_j} \leq &\frac{tr(H_{k+1})}{2} + \sum_{j=1}^{k} \frac{V_j[1 - C_7\|x_j\|]}{\alpha_j} \\ \leq &tr(H_{k+1}) + \sum_{j=1}^{k} \left\{ \frac{V_j[1 - (C_3 + C_3 m^{-1} + C_6)]\|x_j\|]}{\alpha_j} - \frac{C_6\|x_j\|}{\alpha_{j+1}} - C_1\|d_j\| \right\} \\ \leq &tr(H_1) + k + \sum_{j+1}^{k} C_7\|x_j\|. \end{aligned} \tag{4.9}$$

$V > 1$, (4.9) and $\|x_j\|$ is bounded imply Lemma 4.1 is true.

Because the discussion which give the recurrence formula of $\tilde{B}_{k+1}$ from B_k is same as the Broyden update formula, we may use the proof of the Lemma 4.2 in Byrd et al(1987) word for word almost to prove the following (4.10), (4.11) and Theorem 4.1 which corresponding to (4.5), the final line of page 1182 and Lemma 4.2 in the Byrd's paper respectively. That is , given a $\varepsilon \in (0,1)$, there exist $C_9 > 0, C_8 > 0$ and integer number K such that, for all $k > K$,

$$0 < tr(B_*) \leq tr(\tilde{B}_*) \leq tr(B) + M + \frac{\alpha}{cos^2\beta}[\frac{2\phi\varepsilon cos\beta}{mC_8} - \frac{1-\phi}{C_9}] \tag{4.10}$$

$$0 < tr(B_{k+1}) < tr(B_1) + Mk + C_8 - C_9\sum_{j=1}^{k}\frac{\alpha_j}{cos^2\beta_j} \tag{4.11}$$

and

Theorem 4.1. *For given $\phi \in [0,1)$, there exists C_{10}, $0 \leq C_{11} < 1$ such that, $f(x_{k+1}) \leq C_{10}C_{11}^k f(x_1)$ for all k.*

5. The Linear Convergence of the MDFP Algorithms

In this section we also assume the assumptions 1-3 in section 3 hold. We discuss the linear convergence for MDFP algorithm, i.e. $\phi = 1$. Because $\|x_k\| \to 0$ and $M\|x_k\|^2 \geq m\|x_{k+j}\|^2$ for all k and j, the proof of Lemma 5.1 may be same as that in Pu (1987, 1992).

Lemma 5.1. *Let $\{D_k\}$ be a sequence of positive numbers, $\{E_k\}$ be a sequence of positive numbers except finite numbers, and d_1, d_2, d_3, and d_4 be four positive numbers which satisfy following inequality, for all k,*

$$E_k + \sum_{j=1}^{k}[D_j(1 - d_1\|x_j\|)] \leq d_3 + d_2k + \sum_{j=1}^{k} d_4\|x_j\| \tag{5.1}$$

then there exists a $d_5 > 0$ such that, for all k,

$$E_k + \sum_{j=1}^{k} D_j \leq d_2k + \sum_{j=1}^{k} d_5\|x_j\|. \tag{5.2}$$

From the definition (1.7) of $\eta(\phi)$, we derive the relation $\eta(1) = V$.

Lemma 5.2. *There exists a $C_{12} > 0$ such that, for all k,*

$$tr(\tilde{B}_{k+1}) + \sum_{j=1}^{k}\left[\alpha + \frac{p_j\|\tilde{B}_jR_jg_j\|^2}{2(1 + p_jg_j^TR_j\tilde{B}_jR_jg_j)}\right] \leq k + \sum_{j=1}^{k} C_{12}\|x_j\|. \tag{5.3}$$

Proof. As $\phi = 1$, let C=2 and substituting (3.6) and (3.8) into (2.7), we get

$$tr(\tilde{B}_*) + \alpha(1 - C_3\|x\|) + \frac{p_j\|\tilde{B}_j R_j g_j\|^2}{2(1 + p_j g_j^T R_j \tilde{B}_j R_j g_j)} - C_4\alpha\|x\| \tag{5.4}$$
$$\leq tr(\tilde{B}) + \frac{\|y\|^2}{y^T s} \leq tr(\tilde{B}) + 1 + C_3\|x\|.$$

Summing the two sides of (5.4) over $j = 1, 2, \cdots, k$ implies there exist a $C > 0$ such that, for all k,

$$tr(\tilde{B}_{k+1}) + \sum_{j=2}^{k}\left[\alpha_j + \frac{p_j\|\tilde{B}_j R_j g_j\|^2}{2(1 + p_j g_j^T R_j \tilde{B}_j R_j g_j)}\right][1 - (C_3 + C_4)\|x_j\|] \tag{5.5}$$
$$\leq tr(\tilde{B}_{k+1}) + \sum_{j=1}^{k}\alpha_j[1 - (C_3 + C_4)\|x_j\|] + \sum_{j=2}^{k}\frac{p_j\|\tilde{B}_j R_j g_j\|^2}{2(1 + p_j g_j^T R_j \tilde{B}_j R_j g_j)} + (C_3 + C_4)\alpha_1\|x_1\|$$
$$\leq tr(H_1) + (C_3 + C_4)\alpha_1\|x_1\| + k + \sum_{j=1}^{k} C_3\|x_j\|.$$

Because $\|x_j\|$ is bounded the lemma is implied by (5.5) and Lemma 5.1 easily.

Lemma 5.3. *There exists a $C_{13} > 0$ such that, for all k,*

$$\frac{\|Hy\|^2}{V^2} \geq [\frac{g^T H g}{\alpha} + \frac{g_*^T H_* g_*}{\alpha_*}](1 - C_{13}\|x\|). \tag{5.6}$$

Proof. $\eta(1) = V$, (2.11), (3.5), And the description before (3.6) imply

$$\frac{\|Hy\|^2}{V^2} = \|\tilde{H}_* g_* - Hg\|^2 = \|\tilde{H}_* g_*\|^2 + \|Hg\|^2 - 2g_*^T \tilde{H}_* Hg \tag{5.7}$$
$$\geq (\|Hg\|^2 + \|\tilde{H}_* g_*\|^2)(1 - C_3\|x\|)$$
$$\geq \|\tilde{H}_* g_*\|^2(1 - C_3\|x\|) + \frac{g^T H g(1 - 2C_3\|x\|)}{\alpha}$$

and (2.14) implies

$$\|\tilde{H}_* g_*\|^2 \geq C_1^{-1}\|H_* g_*\|^2 \geq \alpha_*^{-1}(g_*^T H_* g_*)(1 - C_{14}\|x\|) \tag{5.8}$$

where $C_{14} = C_3 + C_1$. We, thus, have

$$\frac{\|Hy\|^2}{V^2} \geq \frac{g^T H g}{\alpha}(1 - C_3\|x\|) + \frac{g_*^T H_* g_*}{\alpha_*}(1 - C_3\|x\| - C_{14}\|x\|). \tag{5.9}$$

Let $C_{13} = 2C_3 + C_{14}$, it completes the proof of this lemma.

From the recurrence formula of the trace of the DFP algorithms (or taking $\phi = 1$) in (2.12), (5.6) and the definition (1.7) imply

$$tr(H) + 1 + C_3\|x\| \geq tr(\tilde{H}_*) + \frac{\|Hy\|^2}{V^2 g^T H g} \geq tr(\tilde{H}_*) + (\frac{U}{\alpha_*} + \frac{1}{\alpha})[1 - C_{13}\|x\|]. \tag{5.10}$$

Summing the two sides of (5.10) over j=1,2,...,k, then summing the two sides of (2.13) over j=2,3,...,k, and adding the two sides of these summing forms respectively and omitting the same forms $\{tr(\tilde{H}_{j+1})\}$, we get by (2.14)

$$tr(H_{k+1})+\sum_{j=1}^{k}\frac{U_j}{\alpha_{j+1}}(1-C_{13}\|x\|)+\frac{1}{\alpha_j}[1-C_{13}\|x_j\|-C_7\|x_j\|] \tag{5.11}$$
$$\leq tr(H_1)+k+\sum_{j=1}^{k}C_1C_3\|x_j\|.$$

Theorem 5.1. *There exist $C_{15}>0$ and $0<\delta<1$ such that, for all k,*

$$\|g_k\|\leq C_{15}\delta^k \tag{5.12}$$

. *Proof.* Adding the two sides of (5.3) and (5.11), $tr(H)g^THg\geq g^THHg\geq \alpha m g^THg$ gives

$$tr(H_{k+1}+\tilde{B}_{k+1})+\sum_{j=1}^{k}\{\alpha_j(1-C_{13}\|x_j\|)+[\frac{1}{\alpha_j}+\frac{U_j}{\alpha_{j+1}}] \tag{5.13}$$
$$[1-(C_{13}+C_1C_3)\|x_j\|]\}\leq tr(H_1)+2k+\sum_{j=1}^{k}(C_3+C_{12})\|x_j\|.$$

Lemma 5.1 implies that there exists a $C_{16}>0$ such that, for all k,

$$tr(H_{k+1})+tr(\tilde{B}_{k+1})+\sum_{j=1}^{k}\left[\frac{(\alpha_j-1)^2}{\alpha_j}+\frac{U_j}{\alpha_{j+1}}\right]\leq\sum_{j=1}^{k}C_{16}\|x_j\|. \tag{5.14}$$

We get $\sum_{j=1}^{k}\frac{U_j}{\alpha_{j+1}}\leq\sum_{j=1}^{k}C_{16}\|x\|$, let $\Delta_k=\left[\frac{k+\sum_{j=1}^{k}C_{12}\|x_j\|}{k}\right]^k$, because $(\Delta)^{\frac{1}{k}}\to 1$ as $\|x_k\|\to 0$, (5.3) implies

$$\prod_{j=1}^{k}\alpha_j\leq\left[k^{-1}\sum_{j=1}^{k}\alpha_j\right]^k\leq\Delta_k \tag{5.15}$$

$tr(\tilde{B}_{k+1})\leq tr(B_1)+k+\sum_{j=1}^{k}C_{11}\|x_j\|$. Let $C_{17}=tr(B_1)+C_{12}sup\{\|x_j\|\}+1$ then $tr(B_{k+1})\leq C_{17}k$ holds for all k, we get

$$\frac{\alpha_1 m\|g_{k+1}\|^2}{(g_1^TH_1g_1)\Delta_k(C_{17}k)^2}\leq\frac{\alpha_1(g_{k+1}^TH_{k+1}g_{k+1})}{\alpha_{k+1}(g_1^TH_1g_1)}\prod_{j=1}^{k}\frac{1}{\alpha_j}=\prod_{j=1}^{k}\frac{U_j}{\alpha_{j+1}}\leq\left(k^{-1}\sum_{j=1}^{k}C_{16}\|x_j\|\right)^k. \tag{5.16}$$

Because of $(\Delta_k)^{\frac{1}{k}}\to 1$, and $\|x\|\to 0$ it is clear that the lemma is true.

6. The One-step Superlinear Convergence Rate

In this section we also assume the assumptions 1-3 in section 3 hold. The algorithm presented in this paper has been proved having the linear convergence. First we know$\sum_{j=1}^{\infty} \|x\| < +\infty$ by sections 4-5. Similar to Powell (1976) and Byrd et al(1987), this may imply that our algorithms have one-step superlinear convergence rate for $\|\tilde{B} - B\| = O(\|x\|)$ and $\|\tilde{H} - H\| = O(\|x\|)$. But we rather use the different method with other papers to get some more strong convergence properties and relations.

Theorem 6.1. *The algorithms presented in this paper are one-step superlinear convergent for uniformly objective function, i.e.* $\lim_{k\to\infty} \frac{\|g_{k+1}\|}{\|g_k\|} = 0$.

Proof. (1.5) implies, for all k and $\phi \in [0,1]$,

$$\tilde{B}_{k+1} = B_{DFP} - (1-\phi)(s_k^T B_k s_k) v_k v_k^T \le B_{DFP} \tag{6.1}$$

where B_{DFP} denotes the DFP update formula. Therefore all relations in the proof of Lemma 5.2 are also true. Specifically (5.3) is true, i.e. the following inequality holds for all $\phi \in [0,1]$

$$tr(\tilde{B}_{k+1}) + \sum_{j=1}^{k} \left[\alpha_j + \frac{p_j \|\tilde{B}_j R_j g_j\|}{2 + (1 + p_j g_j^T R_j \tilde{B}_j R_j g_j)} \right] \le tr(B_1) + k + \sum_{j=1}^{k} C_{12}\|x_j\| \le k + C_{18} \tag{6.2}$$

where C_{18} is independent of k. On the other hand, by (4.8), we obtain

$$\frac{tr(H_{k+1})}{2} + \sum_{j=1}^{k} \frac{V_j}{\alpha_j}[1 - C_7\|x_j\|] \le tr(H_1) + k + \sum_{j=1}^{k} C_7\|x_j\|. \tag{6.3}$$

Therefore Lemma 5.1 implies there exist $C_{19} > 0$ and $C_{20} > 0$ such that, for sufficient large k,

$$\frac{tr(H_{k+1})}{2} + \sum_{j=1}^{k} \frac{V_j}{\alpha_j} \le tr(H_1) + k + \sum_{j=1}^{k} C_{19}\|x_j\| \le k + C_{20}. \tag{6.4}$$

The definitions of V and W in (1.7) imply

$$1 + W = V. \tag{6.5}$$

Adding both sides of (6.2) and (6.4) respectively, we get,

$$\frac{tr(H_{k+1})}{2} + tr(\tilde{B}_{k+1}) + \sum_{j=1}^{k} \left[\frac{W_j}{\alpha_j} + \frac{(1-\alpha_j)^2}{\alpha_j} \right] \le C_{18} + C_{20}. \tag{6.6}$$

So we may obtain $tr(H_{k+1} + B_{k+1}) \le C_{18} + C_{20}$, $\sum_{j=1}^{\infty} W_j < +\infty$ and $\alpha_j \to 1$, and

$$\sum_{j=1}^{\infty} \frac{\|g_{j+1}\|^2}{\|g_j\|^2} \le \sum_{j=1}^{\infty} W_j tr(H_j) tr(B_j) < +\infty, \tag{6.7}$$

this completes the proof of theorem 6.1.

From the proof of Theorem 6.1, we may obtain there exist B and H such that $\lim_{k\to\infty} B_k = B$, and $\lim_{k\to\infty} H_k = H$.

7. The Two-order Convergence Rate Examples and the Discussion

In this section we list the following lemma and theorem without proof.

Lemma 7.1. *(7.1)-(7.3) below are true for all k and p, $0 < k < p \le k+n$,*

$$|x_k^T x_p - \|x_p\|^2| \le 5^{p-k} L \|x_k\|^2 \|x_{k+1}\| \tag{7.1}$$

$$|g_p^T s_k| = \alpha_k |g_p^T H_k g_k| \le 2 \times 5^{p-k} L \|x_k\|^2 \|x_{k+1}\| \tag{7.2}$$

$$|g_k^T s_p - g_p^T s_p| \le 3 \times 5^{p-k} L \|x_k\|^2 \|x_{k+1}\| \tag{7.3}$$

where L is the Lipschitz number and let $L > 1$ for simplicity.

Theorem 7.1. *The algorithm is n-step second order convergent.*

We have done a lot of numerical tests for the Broyden algorithms, modified Broyden algorithms and other variable metric algorithms with or without exact linear search respectively. The results of these tests show the modified Broyden algorithms have the least iterations to reach the optimal point on the whole. Comparing with the Broyden algorithm, there are about 25-40 per cent iterations cut down for the modified Broyden algorithm in general and has better stability. Parts of these examples as follows:

Function 1. $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$;

Function 2. $f(x) = x_1^4 + x_1 x_2 + (1 + x_2)^2$;

Function 3. $f(x) = (x_1 - 1)^2 + (x_2 - x_1)^2 + (x_3 - x_2)^4$;

Function 4. $f(x) = (x_1 - 1)^2 + (x_2 - x_1)^2 + (x_3 - x_4)^4 + (x_4 - x_5)^4$;

Function 5 .$f(x) = f_1 + x_4 f_2 + \frac{C f_2^2}{2} + (\frac{\partial L}{\partial x_1})^2 + (\frac{\partial L}{\partial x_2})^2 + (\frac{\partial L}{\partial x_3})^2$;

where f_1 is equal to Function 3 above, $f_2 = x_1(1 + x_2^2) + x_3^4 - 4 - 3\sqrt{2}$, $L = f_1 + x_4 f_2$. When we take $S = I_{n\times n}$ and $R = 0.001\|g_1\|^{-1} n^{-2} I_{n\times n}$, the results for the iteration numbers of each class of the algorithm as follows

Function	Starting Points	Broyden(BFGS)	MBFGS	MDFP
1	0;0	17	12	14
1	2;-2	31	17	17
2	5;5	12	6	7
2	5;-5	10	6	6
2	-5;5	10	7	6
3	0;0;0	99	12	12
3	2;-2;2	19	9	10
3	-2;2;-2	17	8	8
4	0;0;0;0;0	50	35	40
4	2;2;2;2;2	51	34	32
4	2;0;2;0;2	120	35	37
4	0;2;0;2;0	98	31	37
5(C=1)	1;1;1;1	20	16	16
5(C=4)	0;0;0;0	21	17	16
5(C=4)	1;1;1;1	20	14	16

References

[1] R.H. Byrd, J. Nocedal and Y. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, *SIAM. J. Numer. Anal.*, 24 (1987), 1171-1190.

[2] R.H. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, *SIAM.J. Numer. Anal.*, 26 (1989), 727-739.

[3] N.I. Conn, M. Gould and PH.L.Toint, Convergence of quasi-Newton matrices generated by the symmetric rank one update, *Math.Prog.*, 52 (1991), 177-195.

[4] R. Fletcher, Practical Methods of Optimization, Unconstrained Optimization, John Wiley and Sons, New York, 1987.

[5] H.Y. Huang, Unified approach to quadratically convergent algorithms for function minimization, *JOTA*, 5 (1070), 405-423.

[6] M.J.D. Powell, On the convergence of the variable metric algorithm, *Math. Appl.*, 7 (1971), 21-36.

[7] M.J.D. Powell, Some properties of the variable metric algorithm, in: "Numerical Methods for Nonlinear Optimization" Academic Press, London, 1972, 1-17.

[8] M.J.D. Powell, Some global convergence properties of the variable metric algorithm for minimization without exact line searchers, *Nonlinear Programming, SIAM-AMS.Proceedings 6*, R.W. Cottle and C.E. Lemke, eds., *SIAM*, Publications, 1976, 53–72.

[9] D. Pu, Another proof of the superlinear convergence of the DFP variable metric algorithm, *J. of Shanghai Institute of Railway Technology*, 8 (1987), 19-32.

[10] D. Pu, The modified DFP algorithms, *L. of Acta Mathematicae (China)*, 13 (1989), 118-121. (1989a).

[11] D. Pu, A class of the DFP algorithm with linear constrained conditions, *J. of Oper. Res. (China)*, 8 (1989), 53-54. (1989b)

[12] D. Pu and W. Yu, On the convergence property of the DFP algorithm, *J. of Ann. of Oper. Res.*, 24 (1990), 175-184.

[13] D. Pu, A class of the DFP variable metric algorithm with modified wolfe conditions, *J. of Oper. Res. (China)*, 9 (1990) 49-50.

[14] D. Pu, A class of DFP algorithm without exact linear search, *Asia-Pacific J. of Oper. Res.*, **9** : 2 (1992), 207-220.

[15] D. Pu and W. Tian, A class of the DFP variable metric algorithm without exact linear search and convexity assumption, *J. of Oper. Res. (China)*, **22** : 2 (1993) .

[16] W. Tian, A class of the DFP algorithm with constraint conditions, *J. of Appl. Math and Comp. Math.*, **6** : 1 (1992), 42-45.

[17] W. Tian and D. Pu, A class of constrained variable metric algorithm with modified search directions, *J. of Appl. Math and Comp. Math.*, **7** : 1 (1993), 50-57.

[18] F. Wu and X. Gui, A class of the variable metric algorithm with n+1 parameters, *J. of Acta Math. Sinica (China)*, 24 (1981), 921-930.

[19] F. Wu, Some Advances in Linear and Nonlinear Programming, *Proc. 1991 Conference APOR Soc.*, Peking University Press, Beijing, 35-45.

[20] Y. Yuan, A modified BFGS algorithm for unconstrained optimization, *IMA J. Numer. Anal.*, 11 (1991), 325–332.