The Ninth International Symposium on Operations Research and Its Applications (ISORA'10)
Chengdu-Jiuzhaigou, China, August 19–23, 2010

Global Convergence of the Non-Quasi-Newton Method with Non-Monotone Line Search for Unconstrained Optimization Problem

Hong-Wei Liu[1,2,*]

[1]School of Information, Beijing Wuzi University, Beijing 101149, China
[2]Institute of Applied Mathematics,
Academy of Mathematics and Systems Science, CAS, Beijing 100190, China

Abstract In this paper, non-monotone line search procedure is studied, which is combined with the non-quasi-Newton family. Under the uniformly convexity assumption on objective function, the global and superlinear convergence of the non-quasi-Newton family with the proposed non-monotone line search is proved under suitable conditions.

Keywords Quasi-Newton method; Broyden class; non-quasi-Newton; non-monotone line search; global convergence; unconstrained optimization

1 Introduction

Consider the following nonlinear programming problem

$$\min f(x), \tag{1}$$

where $f: R^n \to R^1, f \in C^2$. General line search methods for solving (1) have the following form

$$x_{k+1} = x_k + \lambda_k d_k, \quad k = 0, 1, 2, \cdots.$$

where x_0 is any given starting point, λ_k is a stepsize, d_k is a search direction. It is known that the quasi-Newton methods are efficient iterative methods. Many papers were devoted to investigating the properties of the Broyden class algorithms[1,2,3,4,5]. Meanwhile, the study on non-quasi-Newton method, a method including function information which does not satisfy the quasi-Newton equation and has merits comparing to Broyden's class in some fields, has also made good progress. In 1991, Yuan Yaxiang[6] proposed a modified BFGS algorithm. In 1995, Yuan Yaxiang and Byrd[7] gave a non-quasi-Newton class. In 1997 and 2000, Chen Lanping and Jiao Baocong[8,9] extended a new non-quasi-Newton family, they just gave global convergence with Wolfe-type line search. In 2006, Liu Hongwei[10] inroduce a new update formula for non-quasi-Newton's family and prove that the algorithm with the update formula by Wolfe-type and Armijo-type line search

*E-mail address: ryuhowell@163.com.

converges globally and Q-superlinearly if the function to be minimized has Lipschitz continuous gradient. The purpose of this paper is to study this problem further. The search direction of the non-quasi-Newton methods is determined as follows:

$$d_k = -H_k g_k \ , \ \ g_k = \nabla f(x_k),$$

where H_0 is any given $n \times n$ symmetric positive definite matrix, $H_k = B_k^{-1}$. The Hessian approximation B_k is updating by[9]:

$$B_{k+1}(t, \varphi_k) = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{Q_k(t)}{(y_k^T s_k)^2} y_k y_k^T + \varphi_k V_k V_k^T , \tag{2}$$

where φ_k is a scalar, $f_k = f(x_k)$, $y_k = g_{k+1} - g_k$, $s_k = x_{k+1} - x_k = \lambda_k d_k$ and

$$\begin{aligned} Q_k(t) &= t y_k^T s_k + 2(1-t) R_k, t \in [0,1], \\ R_k &\overset{\triangle}{=} f_{k+1} - f_k - g_k^T s_k, \\ V_k &= (s_k^T B_k s_k)^{\frac{1}{2}} \left(\frac{y_k}{y_k^T s_k} - \frac{B_k s_k}{s_k^T B_k s_k} \right). \end{aligned}$$

The choice of the parameter t is important, since it can greatly effect the performance of the methods. When t=1 or 0 from (2), we can obtain the Broyden algorithm or the quasi-Newton-B algorithm[11].

It is well known that if the initial Hessian approximation B_0 is symmetric and positive definite, together with $y_k^T s_k > 0$ for all k and

$$\varphi_k > \varphi_k^* \equiv \frac{1}{1-\mu_k}, \ \ \mu_k = \frac{s_k^T B_k s_k y_k^T H_k y_k}{(y_k^T s_k)^2}, \tag{3}$$

then all the matrices B_k remain symmetric and positive definite[12].

Powell[13] showed that the BFGS method is globally convergent for convex functions and Byrd, Nocedal and Yuan[14] extended his result to $\varphi_k \in [0,1)$. For convex functions, Zhang and Tewarson[15] proved the global convergence of Broyden's class with $\varphi_k \in [(1-\nu)\varphi_k^*, 0]$, where ν is a number in (0,1). For uniformly convex functions, Byrd, Liu and Nocedal[16] proved the global convergence of Broyden's class with

$$\varphi_k \in [(1-\nu)\varphi_k^*, 1-\delta], \ \ \delta, \nu \in (0,1) \tag{4}$$

and this work is also done about non-quasi-Newton family[8,9].

It is well known that the objective functions sequences generated by the above algorithms are monotonically decreasing; i.e., $f(x_{k+1}) \geq f(x_k)$, $k = 1, 2, \cdots$. In 1986, Grippo et al.[17] proposed a non-monotone line search technique for Newton's method. Since then, the non-monotone technique has been studied by many authors[18,19,20]. Theoretic analysis and numerical results show that the algorithms with non-monotone properties are more efficient than the algorithms with monotone properties. In this paper, under the condition (4), we combine with non-monotone technique to propose a non-monotonical non-quasi-Newton method based on [8] and study its convergence properties.

Algorithm 1

Step1. Initially $0 < \varepsilon_1 \le \varepsilon_2 < 1,\ p < 1,\ \lambda_k = 1$ is given. M_0 is a nonnegative integer. $f_1 := f(x_k),\ f_1^{'} := g_k^T d_k < 0$. Compute the largest index $m(k)$ such that

$$f(x_{m(k)}) = \max_{\max\{k-M_0,1\} \le j \le k} f(x_j).$$

Step2. Calculate $f := f(x_k + \lambda_k d_k)$ and the ratios

$$\rho_{1,k} = \begin{cases} \dfrac{f(x_{m(k)}) - f(x_k + \lambda_k d_k)}{\sum\limits_{j=m(k)}^{k} -\lambda_k g_k^T d_k} & ,k > 1, \\ 0 & ,otherwise \end{cases} \tag{5}$$

and

$$\rho_{2,k} = \frac{f(x_k) - f(x_k + \lambda_k d_k)}{-\lambda_k g_k^T d_k} \tag{6}$$

and set

$$\rho_k = \min\{\rho_{1,k}, \rho_{2,k}\} \tag{7}$$

If $\rho_k \ge \varepsilon_1$, then go to Step 4;

Step3. Evaluate $\widehat{\lambda}$ by restricted quadratic interpolation using $f_1,\ f_1^{'}$ and f. Set $\lambda_k := \widehat{\lambda}$, go to Step 2;

Step4. Calculate $g := g(x_k + \lambda_k d_k)$ and $f^{'} := g^T d_k$. If

$$g(x_k + \lambda_k d_k)^T d_k \ge \max\{\varepsilon_2, 1 - (\lambda_k \|d_k\|^p)\} g_k^T d_k, \tag{8}$$

then stop. Otherwise, evaluate $\widehat{\lambda}$ by restricted quadratic extrapolation using $f_1,\ f_1^{'}$ and $f^{'}$. Set $f_1 := f,\ f_1^{'} := f^{'}$ and $\lambda_k := \widehat{\lambda}$, go to Step 2;

We define

$$h(k) = \begin{cases} m(k), & if\ \ \rho_k = \rho_{1,\,k}, \\ k, & if\ \ \rho_k = \rho_{2,\,k}. \end{cases}$$

We will call iteration $h(k)$ the reference iteration associated with iteration k. The above non-monotone line search sketch is motivated by [17] and [20]. Obviously, the Wolfe line search, which is often used in theory and application, is a special case of the above line search with $M_0 = 0,\ p = 0$.

2 Preliminary assumption and lemma

We give the following Assumptions:

Assumption 1. The level set $D = \{x | f(x) \le f(x_0)\}$ is bounded and there exists positive constants m and M such that

$$m\|z\|^2 \le z^T G(x) z \le M\|z\|^2. \tag{9}$$

for all $z \in R^n$ and all $x \in D$, and $G(x)$ denotes the Hessian matrix of f.

Assumption 2. $f \in C^2$.

From Assumption 1 we can easily induce that there exist two positive numbers m and M such that

$$m\|s_k\|^2 \le y_k^T s_k \le M\|s_k\|^2.$$

Lemma 2.1.[11] Assume that Assumption 1 and Assumption 2 hold, then there exists a positive number M such that

$$\frac{\|y_k\|^2}{y_k^T s_k} \le M, \quad k = 1, 2, \cdots.$$

Lemma 2.2.[9] Assume that Assumption 1 and Assumption 2 hold, the sequence $\{x_k\}$ is generated by the algorithm belonging to non-quasi-Newton family with φ_k satisfies (4), then there exists a positive number M_1 such that

$$\frac{Q_k(t)\|y_k\|^2}{(y_k^T s_k)^2} \le M_1 \quad k = 1, 2, \cdots.$$

Lemma 2.3. $det(B_{k+1}) \ge \nu det(B_k)\dfrac{Q_k(t)}{s_k^T B_k s_k}, \nu \in (0,1)$, where $det(B_k)$ denotes the determinant of B_k.

Proof. If $0 \le \varphi_k \le 1-\delta$, from chen[8], we easily have

$$det(B_{k+1}(\varphi_k)) \le det(B_k(\varphi_k))\frac{Q_k(t)}{s_k^T B_k s_k}. \tag{10}$$

When $\varphi_k = 0$, (10) turns to

$$det(B_{k+1}(0)) = det(B_k(0))\frac{Q_k(t)}{s_k^T B_k s_k}.$$

Then we now see the case of $\varphi_k \in [(1-\nu)\varphi_k^*, 0],\ \nu \in (0,1)$. From (2) we have

$$B_{k+1}(\varphi_k) = B_{k+1}(0) + \varphi_k V_k V_k^T,$$

so we have

$$\begin{aligned} det(B_{k+1}(\varphi_k)) &= det(B_{k+1}(0) + \varphi_k V_k V_k^T) \\ &= det[B_{k+1}(0)(I + \varphi_k H_{k+1}(0) V_k V_k^T)] \\ &= det(B_{k+1}(0)) det(I + \varphi_k H_{k+1}(0) V_k V_k^T). \end{aligned} \tag{11}$$

Where

$$H_{k+1}(0) = H_k(0) - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{Q_k(t)}. \tag{12}$$

From (4), (10), (11), (12) and notes that $det(I+xy^T)=1+x^Ty$, we have

$$det(B_{k+1}(\varphi_k)) = (1+\varphi_k(\mu_k-1))det(B_{k+1}(0))$$

$$\begin{aligned}&= (1+\varphi_k(\mu_k-1))det(B_k(0))\frac{Q_k(t)}{s_k^TB_ks_k}\\&\geq (1+\varphi_k(\mu_k-1))det(B_k)\frac{Q_k(t)}{s_k^TB_ks_k}\\&\geq \nu det(B_k)\frac{Q_k(t)}{s_k^TB_ks_k}.\end{aligned}$$

This completes the proof.

Lemma 2.4. If $f(x_{k+1}) \leq f(x_{h(k)}), k=0,1,\cdots$, then the sequence $\{f(x_{h(k)})\}$ monotonically decreases, and $x_k \in D$ for all $k \geq 0$.

Proof. By $f(x_k) \leq f(x_{h(k-1)})$, we have

$$\begin{aligned}f(x_{h(k)}) &= \max_{\max\{k-M_0,1\}\leq j\leq k} f(x_j)\\&\leq \max\{\max_{\max\{k-M_0,1\}\leq j\leq k} f(x_{j-1}), f(x_k)\}\\&= \max\{f(x_{h(k-1)}), f(x_k)\}\\&= f(x_{h(k-1)}),\ \ k=1,2,\cdots,\end{aligned}$$

i.e., the sequence $\{f(x_{h(k)})\}$ monotonically decreases. Since $f(x_{h(0)})=f(x_0)$, we deduce

$$f(x_k) \leq f(x_{h_{(k-1)}}) \leq \cdots \leq f(x_{h(0)}) = f(x_0)\ \ x_k \in D.$$

Lemma 2.5. Assume that the stepsize λ_k is determined by Algorithm 1. Then

$$\sum_{k=1}^{\infty}(-g_j^Ts_j) < +\infty. \tag{13}$$

Proof. From the definition of $h(k)$, (5), (6), (7) and $\rho_k \geq \varepsilon_1$, we can easily have

$$f(x_{h(k)}) - f(x_{k+1}) \geq \varepsilon_1 \sum_{j=h(k)}^{k}(-g_j^Ts_j).$$

Consider the kth iteration. We see that iteration has an associated reference iteration $h(k)$, in turn, the $h(k)$th iteration has an associated reference iteration $h(k-1)$, $\cdots$, up to the point where x_0 is reached by this backwards reference process.

$$x_1 = x_{h_1},\ \ x_{h_{(j-1)}+1} = x_{h(h_j)},\ \ j=2,\cdots,q,\ \ x_{h_q+1} = x_{h(k)}.$$

Notes that

$$\begin{aligned}&f(x_1)-f(x_{k+1}) = f(x_1)-f(x_{h_1+1})\\&+\sum_{j=2}^{q}[f(x_{h_{(j-1)}+1})-f(x_{h_{(j)}+1})]+f(x_{h(k)})-f(x_{k+1}),\end{aligned}$$

apply Lemma 2.4 2 to each term in the right-hand side of this equation, we have

$$f(x_1) - f(x_{k+1}) \geq \sum_{k=1}^{\infty} (-g_j^T s_j). \tag{14}$$

By induction, we have that $\{x_k\} \subset D$. It follow (H) that $\{f(x_k)\}$ is bounded below on D, together with (14), which imply that (13) is true.

Lemma 2.6. Assume that the sequence $\{x_k\}$ is generated by the algorithm belonging to non-quasi-Newton family with φ_k is satisfies (3), in which the stepsize λ_k is determined by Algorithm 1. Then

$$\lim_{k\to\infty} \frac{(g_k^T s_k)^2}{y_k^T s_k} = 0.$$

Proof. From (8), we have

$$\begin{aligned} y_k^T s_k &\geq -(1 - \max\{\varepsilon_2, 1 - (\lambda_k \|d_k\|)^p\}) g_k^T s_k \\ &= -\min\{1 - \varepsilon_2, (\|s_k\|)^p\} g_k^T s_k. \end{aligned} \tag{15}$$

Lemma 2.5 2 implies that

$$\lim_{k\to\infty} (-g_k^T s_k) = 0. \tag{16}$$

Assumption 1 and Assumption 2 indicate that

$$\|g_k\| \leq c_0, \;\; k = 1, 2\cdots, \tag{17}$$

where $c_0 > 0$ is a constant. From (15), (16), (17) we have

$$\begin{aligned} 0 \;\; &\leq \frac{(g_k^T s_k)^2}{y_k^T s_k} \leq \frac{-g_k^T s_k}{\min\{1-\varepsilon_2, (\|s_k\|)^p\}} = \max\{\frac{-g_k^T s_k}{1-\varepsilon_2}, \frac{-g_k^T s_k}{(\|s_k\|)^p}\} \\ &\leq \max\{\frac{-g_k^T s_k}{1-\varepsilon_2}, (-g_k^T s_k)^{1-p} (c_0)^p\} \to 0. \end{aligned}$$

This completes the proof.

Lemma 2.6 2 is an important property of our non-monotone algorithm, it plays a vital role in the later proof of the global convergence of the non-monotone algorithm.

3 Global convergence

In this section, we give our main result, which establishes superlinearly the global convergence of our non-monotone algorithm belonging to non-quasi-Newton family with φ_k is satisfies (3).

Theorem 3.1. Suppose that Assumption 1 and Assumption 2 hold. Assume that x_0 is any starting point, B_0 is any symmetric positive define matrix, and that the sequence $\{x_k\}$ is generated by the algorithm belonging to non-quasi-Newton family with φ_k is satisfies (3), in which the stepsize λ_k is determined by Algorithm 1. Then

$$\lim_{k\to\infty} \|g_k\| = 0.$$

Proof. We proceed to prove by contradiction. We may assume that there exists a constant $c > 0$ such that

$$\|g_k\| \geq c. \tag{18}$$

From (2) and $tr(xy^T) = x^T y$, we have

$$\begin{aligned} tr[B_{k+1}(t,\varphi_k)] &= tr(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{Q_k(t)\|y_k\|^2}{(y_k^T s_k)^2} + \varphi_k \|V_k\|^2 \\ &= tr(B_k) - (1-\varphi_k)\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{Q_k(t)\|y_k\|^2}{(y_k^T s_k)^2} + \varphi_k \frac{s_k^T B_k s_k}{y_k^T s_k}\frac{\|y_k\|^2}{y_k^T s_k} - 2\varphi_k \frac{y_k^T B_k s_k}{y_k^T s_k}, \end{aligned} \tag{19}$$

where $tr(B_k)$ denotes the trace of B_k.

Denote $K_1 = \{k | 0 \leq \varphi_k \leq 1-\delta, k \in N\}$, and $K_2 = \{k | (1-\nu)\varphi_k^* \leq \varphi_k < 0, k \in N\}$. Now we consider the following two cases.

(1) $k \in K_1$.

Lemma 2.12 indicates that

$$\frac{\|y_k\|^2}{y_k^T s_k} \cdot \frac{s_k^T B_k s_k}{y_k^T s_k} / \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \leq M \frac{(s_k^T B_k s_k)^2}{y_k^T s_k \|B_k s_k\|^2} = M \frac{(g_k^T s_k)^2}{y_k^T s_k \|g_k\|^2}, \tag{20}$$

and

$$\begin{aligned} &\frac{|y_k^T B_k s_k|}{y_k^T s_k} / \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \leq \frac{\|y_k\| s_k^T B_k s_k}{y_k^T s_k \|B_k s_k\|} \\ &\leq \sqrt{M} \frac{s_k^T B_k s_k}{\sqrt{y_k^T s_k}\|B_k s_k\|} = -\sqrt{M} \frac{g_k^T s_k}{\sqrt{y_k^T s_k}\|g_k\|}. \end{aligned} \tag{21}$$

From Lemma 2.1 2, Lemma 2.2 2, Lemma 2.5 2, (4), (18), (19)-(21), we have that

$$tr(B_{k+1}) \leq tr(B_k) - \delta \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + M_1, \tag{22}$$

holds for all sufficiently large $k \in K_1$. Without loss of generality, we can assume that (22) holds for all $k \in K_1$.

(2) $k \in K_2$.

From the first equality of (19) and Lemma 2.2 2, we have

$$tr(B_{k+1}) \leq tr(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + M_1,$$

which implies that (22) also holds in this case.

Therefore, the relation (22) holds for both cases. It follows that

$$tr(B_{k+1}) \leq tr(B_1) - \delta \sum_{j=1}^{k} \frac{\|B_j s_j\|^2}{s_j^T B_j s_j} + kM_1 \leq kM_2 - \delta \sum_{j=1}^{k} \frac{\|B_j s_j\|^2}{s_j^T B_j s_j},$$

where $M_2 = M_1 + tr(B_1)$. (22) implies that

$$det(B_{k+1}) \leq [\frac{tr(B_{k+1})}{n}]^n \leq [\frac{kM_2}{n}]^n, \tag{23}$$

and

$$\sum_{j=1}^{k} \frac{\|B_j s_j\|^2}{s_j^T B_j s_j} \leq \frac{kM_2}{\delta} \equiv kM_3. \tag{24}$$

It follows from the geometric-arithmetic mean value formula, we have

$$\prod_{j=1}^{k} \frac{\|B_j s_j\|^2}{s_j^T B_j s_j} \leq M_3^k. \tag{25}$$

Lemma 2.3 2 indicates that

$$\prod_{j=1}^{k} \nu \frac{Q_j(t)}{s_j^T B_j s_j} \leq \frac{det(B_{k+1})}{det(B_1)}. \tag{26}$$

By (18), (23)-(26), we have

$$\prod_{j=1}^{k} \frac{y_j^T s_j}{(g_j^T s_j)^2} \leq \prod_{j=1}^{k} \frac{\|g_j\|^2}{c^2} \cdot \frac{y_j^T s_j}{(g_j^T s_j)^2} = \prod_{j=1}^{k} \frac{\|B_j s_j\|^2 y_j^T s_j}{c^2 (s_j^T B_j s_j)^2}$$
$$= \prod_{j=1}^{k} \frac{\|B_j s_j\|^2}{c^2 (s_j^T B_j s_j)} \cdot \frac{Q_j(t)}{s_j^T B_j s_j} \cdot \frac{y_j^T s_j}{Q_j(t)} \leq \frac{[kM_3/n]^n}{det(B_1)} \cdot \left(\frac{M_3 M}{c^2 m \nu}\right)^k \leq M_4,$$

where $M_4 > 0$ is a constant. The above relation is in contradiction with Lemma 2.6.

The Q-superlinear convergence of Algorithm 1 then follows from the related assumption in addition if the line search algorithms sets $\lambda_k = 1$ for all sufficiency large k, which will satisfy this conditin if the unit stepsize is always tried first.

Assumption 3. The Hessian matrix $G(x)$ of $f(x)$ is Lipschitz continuous at x^*(a stationary point), i.e., there exists a constant L^* such that

$$\|G(x) - G(x^*)\| \leq L^* \|x - x^*\|$$

for all x in some neighborhood of x^*.

Theorem 3.2. Suppose that Assumptions 1-3 hold, the sequence $\{x_k\}$ is generated by Algorithm 1, then $\{x_k\}$ converges to x^* Q-superlinearly.

Proof. We begin by showing that there exists k_0 such that for all $k \geq k_0$, $\lambda_k = 1$. From the definition of $f(x_{h(k)})$, step 2 of the algorithm 1 and Lemma 2.4 2, we can easily get

$$f(x_k + d_k) - f(x_{h(k)}) - \varepsilon_1 g(x_k)^T d_k \leq f(x_k + d_k) - f(x_k) - \varepsilon_1 g(x_k)^T d_k$$

we find for some $a_k \in [0,1]$ that

$$\begin{aligned} &f(x_k + d_k) - f(x_k) - \varepsilon_1 g(x_k)^T d_k = (1-\varepsilon_1) g(x_k)^T d_k + \frac{1}{2} d_k^T G(x_k + a_k d_k) d_k = \\ &-(\frac{1}{2} - \varepsilon_1) d_k^T B_k d_k + \frac{1}{2} d_k^T [G(x_k + a_k d_k) - B_k] d_k \leq \\ &-\|d_k\|^2 [\sigma_* (\frac{1}{2} - \varepsilon_1) - \|G(x_k + a_k d_k) - G^*\| - \frac{\|(B_k - G^*) d_k\|}{\|d_k\|}], \end{aligned} \tag{27}$$

where we have used the lower bound σ_* on the eigenvalue of B_k. Now $d_k \to 0$ as $k \to \infty$, because $g(x_k) \to 0$ as $k \to \infty$ and the eigenvalue of B_k are bounded. Hence, by continuity,

$$\|G(x_k + a_k d_k) - G^*\| \to 0$$

as $k \to \infty$. Next, since $\lambda_k d_k = s_k$, it follows $\frac{\|(B_k - G^*)d_k\|}{\|d_k\|} \to 0$ as $k \to \infty$ from [2]. Hence there must exist a k_0 such that the right-hand side of (27) is negative, which implies that $\lambda_k = 1$ for $k \geq k_0$. since the remainder proof is the same as the superlinear convergence proof in [10], we omit the following proof.

Acknowledgements

This work is partly supported by Chinese National Science Foundations (grant No. 10701080), Beijing Natural Science Foundations (grant No. 1092011), Foundation of Beijing Education Commission (grant No. SM200910037005), Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (grant No. PHR200906210), Funding Project for Base Construction of Scientific Research of Beijing Municipal Commission of Education (grant No. WYJD200902), Beijing Philosophy and Social Science Planning Project (grant No. 09BaJG258) and Funding Project for Science and Technology Program of Beijing Municipal Commission of Education (grant No. KM200910037002).

References

[1] Byrd R H, Nocedal J. A tool for analysis of quasi-Newton methods with application to unconstrained minimization, SIAM J Number Anal, 1989, 26: 727-739.

[2] Dennis J E and Moré J J. Quasi-Newton methods, motivation and theory, SIAM Rev, 1977, 19: 46-89.

[3] Ke X W. Convergence of the preconvex part of Broyden's Family, Journal of Beijing Normal University, 1995, 31(1): 6-10.

[4] Pearson J D. Variable metric methods of minimization, The Computer Journal, 1969, 12: 171-178.

[5] Yuan Y X. Numerical methods for nonlinear programming. Shanghai Scientific and Technical Publishers, Shanghai, 1993.

[6] Yuan Y X. A modified BFGS algorithm for unconstrained optimization, IMA Joural of Numerical Analyses, 1991, 11: 325-332.

[7] Yuan Y X and Byrd R H. Non-quasi-Newton Updates for Unconstrained Optimization, J of Computational Mathematics, 1995, 13(2): 95-107.

[8] Chen L P, Jiao B C. A Class of Non-Quasi-Newton Methods and Its Convergence, Comm on appl and Comput, 1997, 11(2): 9-17.

[9] Chen L P, Jiao B C. Convergence Properties of the Preconvex Part of Non-Quasi-Newton's family, Mathematica Numerica Sinica, 2000, 22(3): 369-378.

[10] Liu H W , Wang M J , Li J S , Zhang X S. Global Convergence of the Non-quasi-Newton Method for Unstrained Optimization Problems, Appl Math J Chinese Univ, 2006, 21(3): 276-288.

[11] Zhao Y B, Duan Y R. The Property and The Derivation of Quasi-Newton-B Family, Comm on appl and Comput, 2000, 11(2): 82-90.

[12] Fletcher R. Practical methods of optimization, 2nd ed, John Wiley & Sons, Chichseter, 1987.

[13] Powell M J D. Some global convergence properties of variable metric algorithm for minimization without exact line searches in nonlinear programming. In: Cottle R W and Lemke C E eds, SIAM-AMS Proc IX AMS Providence RI, 1976, 53-72.

[14] Byrd R H, Nocedal J and Yuan Y X. Global convergence of a class of quasi-Newton methods on convex problems, SIAM J Number Anal, 2000, 24: 1171-1190.

[15] Zhang Y, Tewarson R P. Quasi-Newton algorithms with updates from the preconvex part of Broyden's family, IMA Journal on Numerical Analysis, 1988, 8: 487-509.

[16] Byrd R, Liu D C, Nocedal J. On the behavior of Broyden's class of quasi-Newton methods, Technical Report NAM 01, Department of Electrical Engineering and Computer Science, North-Western University, 1990.

[17] Grippo L, Lampariello F, Lucidi S. A non-monotone line search technique for Newton's method, SIAM Journal on Numerical Analysis, 1986, 23: 707-716.

[18] Han J Y, Liu G H. Global convergence of the BFGS algorithm with non-monotone line search, Optimization, 1995, 34: 147-159.

[19] Han J Y, Liu G H. Global convergence analysis of a new non-monotone BFGS algorithm on conex objective functions, Computational Optimization and Applications, 1997, 7: 277-289.

[20] Xu D C. Global Convergence of the Broyden's Class of Quasi-Newton Methods with Nonmonotone Line search, Acta Mathematicae Sinica(English Series), 2003, 19: 19-24.

[21] Toint PH L. Global convergence of the patitioned BFGS algorithms for convex partially separable optimization, Mathematical Programming,1 997, 77: 69-94.

[22] Moré J J, Garbow B S and Hillstrome K E. Testing unconstrained optimization software, ACM Trans Math, Software, 1981, 7: 17-41.