Available online at www.sciencedirect.com

ScienceDirect

Journal of Computational and Applied Mathematics 223 (2009) 40–47

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

# Global convergence of the nonmonotone MBFGS method for nonconvex unconstrained minimization[☆]

Weijun Zhou[*], Li Zhang

*College of Mathematics and Computational Science, Changsha University of Science and Technology, Changsha 410076, China*

Received 12 March 2006; received in revised form 29 August 2007

**Abstract**

In this paper, we propose a new nonmonotone Armijo type line search and prove that the MBFGS method proposed by Li and Fukushima with this new line search converges globally for nonconvex minimization. Some numerical experiments show that this nonmonotone MBFGS method is efficient for the given test problems.


*MSC:* 90C30; 65K05

*Keywords:* Nonmonotone line search; MBFGS method; Nonconvex minimization

## 1. Introduction

In this paper, we consider the following unconstrained optimization problem:

$$\min f(x), \quad x \in R^n, \tag{1.1}$$

where $f : R^n \to R$ is a continuously differentiable function whose gradient will be denoted by $g$.

Monotone methods for solving (1.1) require that $f(x_k) \le f(x_{k-1})$ hold at each iteration. But this does not necessarily hold at some iterations for nonmonotone methods. The nonmonotone line search technique was first introduced by Grippo et al. in 1986 [5]. They considered the following general nonmonotone line search sketch : Given constants $a > 0$, $\delta, \rho \in (0, 1)$ and nonnegative integer $M$, select stepsize $\alpha_k = \max\{a\rho^0, a\rho^1, \ldots\}$ satisfying

$$f(x_k + \rho^m a d_k) \le \max_{0 \le j \le M} f(x_{k-j}) + \delta \rho^m a g_k^T d_k. \tag{1.2}$$

When $M = 0$, it becomes standard Armijo line search. An advantage of the nonmonotone line search is that the stepsize $\alpha_k$ can be selected as loosely as possible.

Extensive numerical experiments have showed that the nonmonotone line search technique is very efficient [6,12, 15–19]. This technique was originally applied to Newton's methods [5] and has been applied to conjugate gradient

☆ Supported by the NSF (10701018) of China. Part work of the first author was done while he was visiting Hirosaki University.
* Tel.: +86 07312309200; fax: +86 07312309200.
*E-mail addresses:* weijunzhou@126.com (W. Zhou), zl606@tom.com (L. Zhang).

0377-0427/$ - see front matter 
doi:10.1016/j.cam.2007.12.011

methods and quasi-Newton methods [2,4,7,9–11]. In particular, Han and Liu [7] proposed a nonmonotone Wolfe type line search for BFGS method: Compute stepsize $\alpha_k$ such that

$$f(x_k + \alpha_k d_k) \le \max_{0\le j\le M} f(x_{k-j}) + \epsilon_1 \alpha_k g_k^T d_k,$$

and

$$d_k^T g(x_k + \alpha_k d_k) \ge \max\{\epsilon_2, 1 - (\alpha_k \|d_k\|)^p\} g_k^T d_k,$$

where $\epsilon_1 \in (0, 1)$, $\epsilon_2 \in (0, \frac{1}{2})$, $p \in (-\infty, 1)$. Under suitable assumptions, Han and Liu [7] proved that the BFGS method with this nonmonotone Wolfe line search converges globally for convex objective functions.

Dai [3] constructed an example to show that the BFGS method with Wolfe line search may diverge for nonconvex unconstrained optimization problems. Moreover, Mascarenhas [13] also showed that the BFGS method even with exact line search does not convergence for nonconvex functions. So it is impossible for us to prove the standard BFGS method with nonmonotone line search converges globally for nonconvex minimization.

Fortunately, Li and Fukushima [8] proposed a modified BFGS method (MBFGS) and proved that the MBFGS methods with Armijo or Wolfe line search converges globally even for nonconvex minimization. But it is not known whether the MBFGS method with nonmonotone line search such as (1.2) converges for nonconvex objective functions. The purpose of the paper is to study this problem.

In this paper, we first propose a new nonmonotone Armijo type line search and then apply it to the MBFGS method. In the next section, we present this concrete algorithm. In Section 3, we prove the global convergence of the proposed method for nonconvex minimization. In Section 4, we report some numerical results.

## 2. Algorithm

It is noted that global convergence of one algorithm with the nonmonotone line search (1.2) is often required some strong assumptions. For example, it needs the sufficient descent condition $g_k^T d_k \le -c_1 \|g_k\|^2$, where $c_1$ is a positive constant. Moreover the relation formula $\lim_{k\to\infty} \|s_k\| = 0$ plays an important role in the global convergence analysis.

But the sufficient descent condition is difficult to be satisfied by the BFGS method. In order to ensure that $\lim_{k\to\infty} \|s_k\| = 0$ for the BFGS method, we construct a new nonmonotone Armijo type line search. Now we state our method which we call the MBFGS method as follows:

**Algorithm 2.1** (*MBFGS Method with Nonmonotone Line Search*).

Step 1 : Given $x_0 \in R^n$, $B_0 = I$, $\delta_1$, $\rho \in (0, 1)$, $\delta_2 > 0$ and nonnegative integer $M$. Let $k := 0$. In this paper, for simplicity, we always set $M := \min(k, M)$.

Step 2 : Compute $d_k$ by the following linear equations

$$B_k d + g_k = 0. \tag{2.1}$$

Step 3 : Compute stepsize $\alpha_k = \max\{\rho^0, \rho^1, \ldots\}$ satisfying

$$f(x_k + \rho^m d_k) \le \max_{0\le j\le M} f(x_{k-j}) + \delta_1 \rho^m g_k^T d_k - \delta_2 \|\rho^m d_k\|^2. \tag{2.2}$$

Step 4 : Let $x_{k+1} = x_k + \alpha_k d_k$.

Step 5 : Update $B_k$ to get $B_{k+1}$ by the following MBFGS formula proposed by Li and Fukushima [8]:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{z_k z_k^T}{z_k^T s_k}, \tag{2.3}$$

where $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$,

$$z_k = y_k + C\|g(x_k)\|^r s_k + \max\left\{0, -\frac{y_k^T s_k}{\|s_k\|^2}\right\} s_k, \tag{2.4}$$

where $C,\ r > 0$ are given constants.

Step 6 : Let $k := k + 1$ and go to Step 1.

**Remark.** The MBFGS update formula (2.3) has an attractive property that for each $k$, it always holds that

$$z_k^T s_k \geq C\|g(x_k)\|^r \|s_k\|^2 > 0, \tag{2.5}$$

which ensures that $B_{k+1}$ inherits the positive definiteness of $B_k$. This property is independent of the convexity of $f$ as well as the line search used. Thus the search direction defined by (2.1) is always a descent direction of the objective function, namely, $g_k^T d_k < 0$. The following result shows that the nonmonotone Armijo line search (2.2) is well defined.

**Proposition 2.1.** *Algorithm 2.1 is well defined.*

**Proof.** In fact, we only need to prove that steplength $\alpha_k$ can be obtained in finite steps. If it is not true, then for all sufficiently large positive integer $m$, we have

$$f(x_k + \rho^m d_k) \geq \max_{0\leq j\leq M} f(x_{k-j}) + \delta_1 \rho^m g_k^T d_k - \delta_2 \|\rho^m d_k\|^2 \geq f(x_k) + \delta_1 \rho^m g_k^T d_k - \delta_2 \|\rho^m d_k\|^2. \tag{2.6}$$

Let $m \to \infty$ in (2.6), then

$$g_k^T d_k \geq \delta_1 g_k^T d_k,$$

which implies that $g_k^T d_k \geq 0$ since $\delta_1 \in (0, 1)$. This yields a contradiction. So Algorithm 2.1 is well defined. □

## 3. Global Convergence

In this section, we prove the global convergence of Algorithm 2.1 under the following assumptions.

**Assumption A.** (1) The level set $\Omega = \{x \in R^n | f(x) \leq f(x_0)\}$ is bounded.

(2) In some neighborhood $N$ of $\Omega$, $f$ is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in N. \tag{3.1}$$

It is clear that the sequence $\{x_k\}$ generated by Algorithm 2.1 is contained in $\Omega$. In addition, we get from Assumption A that there is a constant $\gamma_1 > 0$, such that

$$\|g(x)\| \leq \gamma_1, \quad \forall\, x \in \Omega. \tag{3.2}$$

From now on, we always suppose that the conditions in Assumption A hold. Without specification, we let $\{x_k\}$ and $\{d_k\}$ be the iterative sequence and the direction sequence generated by Algorithm 2.1, respectively.

The following lemma comes from Theorem 2.1 of [1] and is very useful to prove the global convergence of the BFGS method.

**Lemma 3.1.** *Let $B_k$ be updated by the BFGS formula (2.3). Suppose that $B_0$ is symmetric and positive definite. If there are positive constants $m \leq M_1$ such that for all $k \geq 0$, $z_k$ and $s_k$ satisfy*

$$\frac{z_k^T s_k}{\|s_k\|^2} \geq m, \qquad \frac{\|z_k\|^2}{z_k^T s_k} \leq M_1, \tag{3.3}$$

*then for any $\kappa \in (0, 1)$, there exist constants $\beta_1, \beta_2, \beta_3 > 0$ such that for any $k \geq 1$, inequalities*

$$\|B_j s_j\| \leq \beta_1 \|s_j\|, \qquad \beta_2 \|s_j\|^2 \leq s_j^T B_j s_j \leq \beta_3 \|s_j\|^2 \tag{3.4}$$

*hold for at least $\lceil \kappa k \rceil$ values of $j \in [1, k]$.*

For every $k$, we define index sets $K_k$ and $K$ as follows:

$$K_k = \{i \leq k |\ (3.4) \text{ hold}\}, \qquad K = \bigcup_{k=0}^{\infty} K_k. \tag{3.5}$$

Since $s_k = \alpha_k d_k$, if we replace $s_k$ by $d_k$, then inequality (3.4) still holds. In addition, we have from (2.1) and the inequality (3.4) that

$$\|g_k\| = \|B_k d_k\| \le \beta_1 \|d_k\|,$$

and

$$\beta_2 \|d_k\|^2 \le \|d_k^T B_k d_k\| = \|d_k^T g_k\| \le \|d_k\| \|g_k\|.$$

Then it follows from the above inequalities that

$$\beta_2 \|d_k\| \le \|g_k\| \le \beta_1 \|d_k\|, \quad \forall k \in K, \tag{3.6}$$

where $\beta_1, \beta_2$ are same as that of (3.4).

The following lemma shows that $\lim_{k\to\infty} \|s_k\| = 0$. Its proof is similar to Theroem 3.1 of [5], for completeness, we give the proof.

**Lemma 3.2.** *Let Assumption A hold, if steplength $\alpha_k > 0$ is computed by the nonmonotone line search (2.2), then we have*

$$\lim_{k\to\infty} \alpha_k d_k = 0, \qquad \lim_{k\to\infty} \alpha_k g_k^T d_k = 0. \tag{3.7}$$

**Proof.** Let $h(k)$ be an integer satisfying

$$k - M \le h(k) \le k, \qquad f(x_{h(k)}) = \max_{0\le j\le M} f(x_{k-j}). \tag{3.8}$$

It follows from (2.2) that the sequence $\{f(x_{h(k)})\}$ is decreasing. In fact, note that $g_k^T d_k < 0$, we have from (2.2) that

$$f(x_{k+1}) \le \max_{0\le j\le M} f(x_{k-j}) = f(x_{h(k)}).$$

Then we get from the above inequality that

$$\begin{aligned} f(x_{h(k+1)}) &= \max_{0\le j\le M} f(x_{k+1-j}) \\ &= \max(\max_{0\le j\le M-1} f(x_{k-j}), f(x_{k+1})) \\ &\le \max(\max_{0\le j\le M-1} f(x_{k-j}), f(x_{k-M}), f(x_{k+1})) \\ &= \max(\max_{0\le j\le M} f(x_{k-j}), f(x_{k+1})) \\ &= \max(f(x_{h(k)}), f(x_{k+1})) \\ &= f(x_{h(k)}). \end{aligned}$$

We have from the last inequality, (2.2) and (3.8) that

$$\begin{aligned} f(x_{h(k)}) &= f(x_{h(k)-1} + \alpha_{h(k)-1} d_{h(k)-1}) \\ &\le \max_{0\le j\le M} f(x_{h(k)-1-j}) + \delta_1 \alpha_{h(k)-1} g_{h(k)-1}^T d_{h(k)-1} - \delta_2 \|\alpha_{h(k)-1} d_{h(k)-1}\|^2 \\ &= f(x_{h(h(k)-1)}) + \delta_1 \alpha_{h(k)-1} g_{h(k)-1}^T d_{h(k)-1} - \delta_2 \|\alpha_{h(k)-1} d_{h(k)-1}\|^2. \end{aligned}$$

Since $\{f(x_{h(k)})\}$ is decreasing and bounded from below from Assumption A, let $k \to \infty$ in the above inequality, we have

$$\lim_{k\to\infty} \alpha_{h(k)-1} d_{h(k)-1} = 0. \tag{3.9}$$

Denote $\hat{h}(k) = h(k + M + 2)$. Now by induction, we prove that for any $j \ge 1$, the following two formulae hold:

$$\lim_{k\to\infty} \alpha_{\hat{h}(k)-j} d_{\hat{h}(k)-j} = 0, \tag{3.10}$$

$$\lim_{k\to\infty} f(x_{\hat{h}(k)-j}) = \lim_{k\to\infty} f(x_{h(k)}). \tag{3.11}$$

For $j = 1$, since $\{\hat{h}(k)\} \subset \{h(k)\}$, it follows from (3.9) that (3.10) holds, which shows that $\|x_{\hat{h}(k)} - x_{\hat{h}(k)-1}\| \to 0$. As $f(x)$ is uniformly continuous in the level set, (3.11) holds for $j = 1$.

Now we suppose that (3.10) and (3.11) hold for given $j$. It follows from (2.2) that

$$f(x_{\hat{h}(k)-j}) \le f(x_{h(\hat{h}(k)-j-1)}) + \delta_1 \alpha_{\hat{h}(k)-j-1} g^T_{\hat{h}(k)-j-1} d_{\hat{h}(k)-j-1} - \delta_2 \|\alpha_{\hat{h}(k)-j-1} d_{\hat{h}(k)-j-1}\|^2.$$

Let $k \to \infty$, we get from (3.11) that

$$\lim_{k\to\infty} \alpha_{\hat{h}(k)-(j+1)} d_{\hat{h}(k)-(j+1)} = 0,$$

which implies that $\|x_{\hat{h}(k)-j} - x_{\hat{h}(k)-(j+1)}\| \to 0$. Since $f(x)$ is uniformly continuous in the level set,

$$\lim_{k\to\infty} f(x_{\hat{h}(k)-(j+1)}) = \lim_{k\to\infty} f(x_{\hat{h}(k)-j}) = \lim_{k\to\infty} f(x_{h(k)}).$$

Thus (3.10) and (3.11) hold for any $j \ge 1$.

Now for any $k$, it holds that

$$x_{k+1} = x_{\hat{h}(k)} - \sum_{j=1}^{\hat{h}(k)-k-1} \alpha_{\hat{h}(k)-j} d_{\hat{h}(k)-j}. \tag{3.12}$$

Since $\hat{h}(k) - k - 1 = h(k + M + 2) - k - 1 \le k + M + 2 - k - 1 = M + 1$, we have from (3.10) and (3.12) that

$$\lim_{k\to\infty} \|x_{k+1} - x_{\hat{h}(k)}\| = 0.$$

We get from the uniform continuity of $f(x)$ that

$$\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} f(x_{\hat{h}(k)}).$$

It follows from (2.2) that

$$f(x_{k+1}) \le f(x_{h(k)}) + \delta_1 \alpha_k g_k^T d_k - \delta_2 \|\alpha_k d_k\|^2.$$

Let $k \to \infty$, we have

$$\lim_{k\to\infty} \alpha_k d_k = 0, \qquad \lim_{k\to\infty} \alpha_k g_k^T d_k = 0.$$

This finishes the proof. □

By the use of Lemmas 3.1 and 3.2, we can prove the following global convergence theorem for the MBFGS method.

**Theorem 3.3.** *We have*

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

**Proof.** We suppose that the conclusion is not true. Then there exists a constant $\varepsilon > 0$ such that for any $k \ge 0$, it holds that

$$\|g_k\| \ge \varepsilon. \tag{3.13}$$

From the above inequality, (2.4), (2.5), (3.1) and (3.2), there exists a constant $C_2 > 0$ such that

$$z_k^T s_k \ge C\|g_k\|^r s_k^T s_k \ge C\varepsilon^r s_k^T s_k \triangleq C_2 s_k^T s_k, \tag{3.14}$$

$$\|z_k\| \le \|y_k\| + C\|g_k\|^r \|s_k\| + \frac{\|y_k\|\|s_k\|^2}{\|s_k\|^2} \le (2L + C\gamma_1^r)\|s_k\|. \tag{3.15}$$

By (3.14) and (3.15), we have

$$z_k^T z_k \le (2L + C\gamma_1^r)^2 \|s_k\|^2 \le \frac{(2L + C\gamma_1^r)^2}{C_2} z_k^T s_k \triangleq C_3 z_k^T s_k.$$

Therefore the conditions (3.4) and (3.6) in Lemma 3.1 are satisfied, so the conclusions of Lemma 3.1 hold.

If $\liminf_{k\in K,\ k\to\infty} \alpha_k > 0$, it follows from (3.6) and (3.7) that

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

This yields a contradiction.

If $\liminf_{k\in K,\ k\to\infty} \alpha_k = 0$, then for sufficiently large $k \in K$, it holds that

$$f\left(x_k + \frac{\alpha_k d_k}{\rho}\right) \geq \max_{0\leq j\leq M} f(x_{k-j}) + \delta_1 \frac{\alpha_k}{\rho} g_k^T d_k - \delta_2 \left(\frac{\alpha_k}{\rho}\right)^2 \|d_k\|^2,$$

which implies that

$$f\left(x_k + \frac{\alpha_k d_k}{\rho}\right) \geq f(x_k) + \delta_1 \frac{\alpha_k}{\rho} g_k^T d_k - \delta_2 \left(\frac{\alpha_k}{\rho}\right)^2 \|d_k\|^2.$$

By the mean-value theorem, there exists $t_k \in (0, 1)$ such that

$$(g(u_k))^T d_k \geq \delta_1 g_k^T d_k - \delta_2 \left(\frac{\alpha_k}{\rho}\right) \|d_k\|^2, \tag{3.16}$$

where $u_k \in [x_k, x_k + \frac{\alpha_k d_k}{\rho}]$. Let $\{x_k\}_K$ is a subsequence which converges to $x^*$ such that

$$\lim_{k\in K,\ k\to\infty} x_k = x^*, \qquad \lim_{k\in K,\ k\to\infty} \frac{d_k}{\|d_k\|} = d^*.$$

Then it also holds that

$$\lim_{k\in K,\ k\to\infty} u_k = x^*.$$

Let $k \in K$ and $k \to \infty$ in (3.16), we have

$$g^T(x^*)d^* \geq \delta_1 g^T(x^*)d^*.$$

Since $1 - \delta_1 > 0$, we have

$$g^T(x^*)d^* \geq 0. \tag{3.17}$$

But (2.1), (3.4), (3.6) and (3.13) imply

$$g_k^T d_k = -d_k^T B_k d_k \leq -\beta_2 \|d_k\|^2 \leq -\frac{\beta_2}{\beta_1} \|g_k\| \|d_k\| \leq -\frac{\beta_2}{\beta_1} \varepsilon \|d_k\|. \tag{3.18}$$

Let $k \in K$ and $k \to \infty$ in (3.18), we have

$$(g(x^*))^T d^* < 0,$$

which contradicts (3.17). The proof is then completed. □

## 4. Numerical experiments

In this section, we report some numerical experiments. We test Algorithm 2.1 with different values of $M$ and the standard BFGS method on some problems in [14].

All codes were written in Matlab code. We stop the iteration if the total number of iterations exceeds $4 \times 10^3$ or $\|g(x_k)\| \leq 10^{-5}$. We set parameters in Algorithm 2.1 as follows: $\delta_1 = \delta_2 = 0.1$, $C = 10^{-6}$, $r = 2$, $\rho = 0.4$.

The test problems include some nonconvex functions such as the following "band" and "jensam" functions.

- Band function ($n = 10$):

$$f(x) = \sum_{i=1}^{n} \left[ x_i(2 + 15x_i^2) + 1 - \sum_{j\in J_i} x_j(1 + x_j) \right]^2,$$

Table 1
Test results on the MBFGS method with different $M$

| Problem | $n$ | BFGS($M = 0$) iter / fn / gn | MBFGS($M = 0$) iter / fn / gn | MBFGS($M = 3$) iter / fn / gn | MBFGS($M = 5$) iter / fn / gn |
|---|---|---|---|---|---|
| bard | 3 | 224/ 1053/ 225 | 224/ 1053/ 225 | 48/ 150/ 49 | 28/ 58/ 29* |
| box | 3 | 580/ 3189/ 581 | 583/ 3201/ 584 | 191/ 779/ 192 | 93/ 296/ 94* |
| sing | 4 | 236/ 1190/ 237 | 237/ 1196/ 238 | 128/ 519/ 129 | 71/ 189/ 72* |
| wood | 4 | 59/ 174/ 60 | 66/ 120/ 67* | 105/ 190/ 106 | 105/ 190/ 106 |
| rose | 2 | 35/ 50/ 36* | 36/ 54/ 37 | 61/ 99/ 62 | 60/ 98/ 61 |
| froth | 2 | 11/ 20/ 12 | 9/ 18/ 10* | 9/ 18/ 10* | 9/ 18/ 10* |
| beale | 2 | 15/ 23/ 16* | 15/ 23/ 16* | 18/ 24/ 19 | 18/ 24/ 19 |
| jensam | 2 | 8/ 14/ 9 | 8/ 14/ 9 | 8/ 14/ 9 | 8/ 14/ 9 |
| gauss | 3 | 3/ 6/ 4 | 3/ 6/ 4 | 3/ 6/ 4 | 3/ 6/ 4 |
| gulf | 3 | 1/ 4/ 2 | 1/ 4/ 2 | 1/ 4/ 2 | 1/ 4/ 2 |
| kowosb | 4 | 487/ 2965/ 488 | 487/ 2965/ 488 | 212/ 956/ 213 | 175/ 798/ 176* |
| bd | 4 | 23/ 71/ 24* | 50/ 99/ 51 | 53/ 95/ 54 | 53/ 95/ 54 |
| osb2 | 11 | 203/ 754/ 204 | 203/ 754/ 204 | 67/ 115/ 68 | 66/ 104/ 67* |
| watson | 2 | 9/ 19/ 10* | 9/ 19/ 10* | 11/ 20/ 12 | 11/ 20/ 12 |
| rosex | 10 | 99/ 164/ 100 | 95/ 159/ 96* | 142/ 202/ 143 | 147/ 209/ 148 |
| band | 10 | NaN | NaN | 61/ 135/ 62 | 61/ 130/ 62* |
| rosex | 50 | 254/ 573/ 255* | 270/ 578/ 271 | 354/ 626/ 355 | 375/ 646/ 376 |
| rosex | 100 | 436/ 1036/ 437 | 408/ 1007/ 409 | 377/ 948/ 378* | 587/ 1127/ 588 |
| singx | 20 | 481/ 2630/ 482 | 637/ 3902/ 638 | 492/ 2170/ 493* | 799/ 3943/ 800 |
| singx | 100 | 1498/ 10449/ 1499 | 1276/ 8714/ 1277* | 1797/ 9418/ 1798 | 2665/ 13455/ 2666 |
| trig | 100 | 63/ 134/ 64 | 63/ 134/ 64 | 43/ 49/ 44 | 42/ 47/ 43* |
| trig | 200 | 59/ 118/ 60 | 59/ 118/ 60 | 45/ 55/ 46 | 45/ 51/ 46* |
| trig | 500 | 57/ 101/ 58 | 57/ 101/ 58 | 46/ 49/ 47* | 46/ 49/ 47* |
| almost | 100 | 4/ 16/ 5* | 6/ 18/ 7 | 6/ 18/ 7 | 6/ 18/ 7 |
| bv | 10 | 60/ 175/ 61 | 60/ 175/ 61 | 19/ 37/ 20 | 19/ 35/ 20* |
| bv | 50 | NaN | NaN | 2414/ 17944/ 2415 | 2025/ 14756/ 2026* |
| bv | 100 | 1052/ 7373/ 1053 | 1052/ 7373/ 1053 | 392/ 1870/ 393 | 298/ 1133/ 299* |
| bv | 200 | 174/ 842/ 175 | 174/ 842/ 175 | 161/ 497/ 162* | 178/ 493/ 179 |
| ie | 100 | 10/ 12/ 11 | 10/ 12/ 11 | 10/ 12/ 11 | 10/ 12/ 11 |
| ie | 500 | 11/ 13/ 12 | 11/ 13/ 12 | 11/ 13/ 12 | 11/ 13/ 12 |
| trid | 100 | 98/ 482/ 99* | 100/ 489/ 101 | 143/ 521/ 144 | 187/ 545/ 188 |
| lin | 100 | 2/ 4/ 3* | 3/ 5/ 4 | 3/ 5/ 4 | 3/ 5/ 4 |
| lin | 500 | 2/ 4/ 3* | 4/ 6/ 5 | 4/ 6/ 5 | 4/ 6/ 5 |

where

$$J_i = \{j : j \neq i, \max\{1, i - m_l\} \leq j \leq \min\{n, i + m_u\}\}$$

and $m_l = 5, m_u = 1$.
If we choose two points $y = (1, \dots, 1)^T$ and $x = (0, \dots, 0)^T$, then we have

$$f(y) - f(x) - \nabla f(x)^T (y - x) = -10 < 0.$$

This shows that the Band function is not convex.

- Jensam function ($n = 2$):

$$f(x) = (4 - e^{x_1} - e^{x_2})^2 + (6 - e^{2x_1} - e^{2x_2})^2.$$

We can get the Hessian matrix of $f(x)$ at $x = (0, 0)^T$

$$\nabla^2 f(0, 0) = \begin{pmatrix} -18 & 10 \\ 10 & -18 \end{pmatrix},$$

which is not a semi-positive definite matrix. So the Jensam function is nonconvex.

Table 1 lists numerical results. In Table 1, "problem" and "$n$" stand for the test problem name and the dimension of the test problem, respectively. "iter/fn/gn" are the total number of the iterations, the function evaluations and the gradient evaluations, respectively.

Table 1 shows that the MBFGS method performs similarly as the standard BFGS method when $M = 0$, but they are failed to solve the problems "band" and "bv" with $n = 50$. The MBFGS method with $M = 3$ or $M = 5$ solves all the given problems successfully even for nonconvex functions such as "band" function. This shows that the nonmonotone methods are more stable than the corresponding monotone methods.

We also can see that the MBFGS method with $M = 5$ performs best since it can solve about 55% (19 out of 34) of all test problems with the smallest number of iterations and function evaluations. But for some problems such as "rose" function, its performance is worse than those of other methods. This also shows that how to choose a suitable $M$ is very important, but this is a relatively difficult problem.

## 5. Conclusions

We have presented the MBFGS method with a new nonmonotone Armijo line search. Under suitable conditions, we proved that the proposed method is globally convergent even for nonconvex functions. Some limited numerical results are also reported, which show that the nonmonotone method is more efficient than the monotone one.

## Acknowledgments

The authors would like to thank referees for giving us many valuable suggestions and comments, which improve this paper greatly.

## References

[1] R.H. Byrd, J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, SIAM J. Numer. Anal. 26 (1989) 727–739.
[2] Y.H. Dai, On the nonmonotone line search, J. Optim. Theory Appl. 112 (2002) 315–330.
[3] Y.H. Dai, Convergence properties of the BFGS algorithm, SIAM J. Optim. 13 (2002) 693–701.
[4] Y.H. Dai, A nonmonotone conjugate gradient algorithm for unconstrained optimization, J Syst. Sci. Complex. 15 (2002) 139–145.
[5] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's method, SIAM J. Numer. Anal. 23 (1986) 707–716.
[6] L. Grippo, F. Lampariello, S. Lucidi, A truncated Newton method with nonmonotone line search for unconstrained optimization, J. Optim. Theory Appl. 60 (1989) 401–419.
[7] J.Y. Han, G.H. Liu, Global convergence analysis of a new nonmonotone BFGS algorithm on convex objective functions, Comput. Optim. Appl. 7 (1997) 277–289.
[8] D.H. Li, M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, J. Comput. Appl. Math. 129 (2001) 15–35.
[9] G.H. Liu, J.Y. Han, D.F. Sun, Global convergence of the BFGS algorithm with nonmonotone line search, Optimization 34 (1995) 147–159.
[10] G.H. Liu, L.L. Jing, Convergence of the nonmonotone Perry and Shanno method for optimization, Comput. Optim. Appl. 16 (2000) 159–172.
[11] G.H. Liu, L.L. Jing, L.X. Han, D. Han, A class of nonmonotone conjugate gradient methods for unconstrained optimization, J. Optim. Theory Appl. 101 (1999) 127–140.
[12] S. Lucidi, F. Rochetich, M. Roma, Curvilinear stabilization techniques for truncated Newton methods in large scale unconstrained optimization, SIAM J. Optim. 8 (1998) 916–936.
[13] W.F. Mascarenhas, The BFGS method with exact line searches fails for non-convex objective functions, Math. Program. 99 (2004) 49–61.
[14] J.J. Moré, B.S. Garbow, K.E. Hillstrom, Testing unconstrained optimization software, ACM Trans. Math. Softw. 7 (1981) 17–41.
[15] E.R. Panier, A.L. Tits, Avoiding the Maratos effect by means of nonmonotone line search I. General constrained problems, SIAM J. Numer. Anal. 28 (1991) 1183–1195.
[16] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problems, SIAM J. Optim. 7 (1997) 26–33.
[17] P.L. Toint, An assessment of nonmonotone line search techniques for unconstrained optimization, SIAM J. Sci. Comput. 17 (1996) 725–739.
[18] P.L. Toint, Non-monotone trust-region algorithms for nonlinear optimization subject to convex constraints, Math. Program. 77 (1997) 69–94.
[19] H.C. Zhang, W.W. Hager, A nonmonotone line search technique and its application to unconstrained optimization, SIAM J. Optim. 14 (2004) 1043–1056.