Available online at www.sciencedirect.com

ScienceDirect

Journal of Computational and Applied Mathematics 219 (2008) 134 – 144

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

A new nonmonotone line search technique for unconstrained optimization☆

Zhensheng Yu[a, b, *], Dingguo Pu[a]

[a] Department of Applied Mathematics, Tongji University, Shanghai 200092, PR China
[b] College of Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China

Received 12 October 2005; received in revised form 2 February 2007

Abstract

In this paper, we propose a new nonmonotone line search technique for unconstrained optimization problems. By using this new technique, we establish the global convergence under conditions weaker than those of the existed nonmonotone line search techniques.

MSC: 65K05; 90C30

Keywords: Unconstrained optimization problem; Nonmonotone F-rule; Global convergence; Memoryless quasi-Newton method

1. Introduction

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

$$\min \quad f(x) \text{ s.t } \quad x \in R^n, \tag{1}$$

where $f(x)$ is a continuously differentiable function from R^n to R.

At current iteration x_k, if $g_k = \nabla f(x_k) \neq 0$, a line search method defines a search direction d_k in some way, finds a step-length by carrying some line search along d_k. Among the most popular line search rules are the Arimijo rule, the Goldstein rule and the Wolfe rule (see [4,5,15]).

The traditional line searches require the function value descent monotonically at every iteration, namely:

$$f(x_{k+1}) \leqslant f(x_k), \tag{2}$$

where $x_{k+1} = x_k + \alpha_k d_k$, d_k is a descent direction, $\alpha_k \geqslant 0$ is obtained by the line search.

Recent research [6–8,10,17,18] indicates that the monotone line search technique may considerably reduce the rate of convergence when the iteration is trapped near a narrow curved valley, which can result in very short steps or zigzagging. The nonmonotone line search technique does not impose the condition (2), as a result, it is helpful to

☆ This work is supported by the National Natural Science Foundation of China (Nos. 10571106, 10671126 and 10571137) and Natural Foundation of Shanghai Education Committee (05EZ50).
* Corresponding author. College of Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China.
E-mail address: zhsh-yu@163.com (Z. Yu).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2007.07.008

overcome this drawback. Several numerical tests show that the nonmonotone line search technique for unconstrained optimization and constrained optimization is effective and competitive.

The first nonmonotone line search technique was proposed by Grippo et al. [6] for unconstrained optimization, the subsequent nonmonotone line search techniques were all based on their method. Recently, by combining forcing function with the nonmonotone line search techniques, Sun et al. [16] proposed a general nonmonotone line search rule, called nonmonotone F-rule, for unconstrained optimization. They showed that the common nonmonotone line search rules, such as the nonmonotone Arimijo rule, the nonmonotone Goldstein rule and the nonmonotone Wolfe rule, are special classes of nonmonotone F-rule.

To obtain the global convergence, Sun et al. required the search direction satisfy the following conditions:

$$\left|\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|}\right| \geqslant \sigma(\|g_k\|), \quad k=1,2,\ldots \tag{3}$$

and

$$\|d_k\| \leqslant c_1 \|g_k\|, \tag{4}$$

where $\sigma(\cdot)$ is a forcing function and $c_1 > 0$.

The conditions (3) and (4) play important role in the nonmonotone line search methods. However, as pointed out in by Toint [18], condition (4) prevents large step where the gradient is small, as in the neighborhood of saddle point and at the bottom of the valley. Although this can be circumvented by choosing a larger c_1, this strategy indicates that condition (4) will not be essential.

In Ref. [3], Dai provided some basic analysis of the nonmonotone Arimijo line search. He showed that when $\nabla f(x)$ is Lipschitz continuous, condition (4) can be relaxed to

$$\|d_k\| \leqslant \beta + \gamma k, \quad k=1,2,\ldots \tag{5}$$

for some positive constant β, γ.

Under condition (5) and the sufficient descent condition

$$g_k^{\mathrm{T}} d_k \leqslant -c_2 \|g_k\|^2, \tag{6}$$

where $c_2 > 0$ is a positive constant, Dai proved the weak convergence property

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{7}$$

In this paper, we propose a new nonmonotone F-rule for unconstrained optimization. The method is motivated by the nonmonotone trust region method proposed in [19]. Using this new nonmonotone technique, we can remove condition (4) and establish the strong convergence property

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{8}$$

Moreover, the compactness of the level set used in [16] and the Lipschitz continuity of $\nabla f(x)$ used in [3] are not needed, either.

The rest of this paper is organized as follows: In Section 2, we present our new nonmonotone F-rule. In Section 3, we analyze the convergence property of the algorithm. The numerical tests are given in Section 4 and we conclude the paper with some final remarks in Section 5.

2. New nonmonotone F-rule

Denote the level set by $\mathscr{L} = \{x \in R^n \mid f(x) \leqslant f(x_0)\}$. The following assumption is imposed throughout this paper.

Assumption 1. $f(x)$ is bounded below on the level set $\mathscr{L}$ and $\nabla f(x)$ is continuous uniformly on an open set that contains the level set $\mathscr{L}$.

In what follows, we give some basic definitions as in Ref. [16].

Definition 1. The function $\sigma : [0, +\infty] \to [0, +\infty]$ is a forcing function (F-function) if for any sequence $\{t_i\} \subset [0, +\infty]$

$$\lim_{i\to\infty} \sigma(t_i) = 0 \quad \text{implies} \quad \lim_{i\to\infty} t_i = 0. \tag{9}$$

Definition 2. Let $\eta = \sup\{\|g(x) - g(y)\| x, y \in \mathscr{L}\} > 0$. Then the mapping $\delta : [0, +\infty] \to [0, +\infty]$ defined by

$$\delta(t) = \begin{cases} \inf\{\|x - y\| \,|\, \|g(x) - g(y)\| \geqslant t\}, & t \in [0, \eta) \\ \lim\limits_{s\to\eta^-} \delta(s), & t \in [\eta, +\infty] \end{cases}$$

is the reverse modulus of continuity of gradient $g(x)$.

The nonmonotone F-rule defined by Sun et al. is as follows:

2.1. The nonmonotone F-rule for line searches

Let M be a nonnegative integer. For each k, let $m(k)$ satisfy

$$m(0) = 0, \quad 0 \leqslant m(k) \leqslant \min[m(k-1), M] \quad \text{for } k \geqslant 1. \tag{10}$$

Let $\alpha_k \geqslant 0$ be bounded above and satisfy

$$f(x_k + \alpha_k d_k) \leqslant \max_{0\leqslant j\leqslant m(k)} [f(x_{k-j})] - \sigma(t_k), \tag{11}$$

where σ is a forcing function and $t_k = -g_k^{\mathrm{T}} d_k / \|d_k\|$. Set

$$x_{k+1} = x_k + \alpha_k d_k.$$

Based on Definitions 1 and 2, Sun et al. proved that the nonmonotone Arimijo rule, the nonmonotone Goldstein rule and the nonmonotone Wolfe rule are special classes of nonmonotone F-rule (see [16, Proposition 2.4]).

In what follows, we give our new nonmonotone F-rule for line searches. Let $\lambda \in (0, 1]$, $M \geqslant 1$ is a positive integer, define $m(k) = \min[k+1, M]$, choose

$$\lambda_{kr} \geqslant \lambda, \quad r = 1, 2, \ldots, m(k) - 1 \quad \text{and} \quad \sum_{r=0}^{m(k)-1} \lambda_{kr} = 1.$$

Let $\alpha_k \geqslant 0$ be bounded above and satisfy

$$f(x_k + \alpha_k d_k) \leqslant \max \left[f(x_k), \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r}) \right] - \sigma(t_k), \tag{12}$$

set

$$x_{k+1} = x_k + \alpha_k d_k. \tag{13}$$

Similar to Proposition 2.4 in [16], it is easy to see that the nonmonotone Arimijo rule, the nonmonotone Goldstein rule and the nonmonotone Wolfe rule are special classes of our new nonmonotone F-rule.

It can be seen that our nonmonotone F-rule is slightly stronger than (11). However, with this new F-rule, we can obtain the global convergence under weaker conditions.

3. Global convergence

In this section, we discuss the global convergence property of unconstrained optimization with the new nonmonotone F-rule.

Lemma 1. *If x_k is generated by* (12, 13). *Then we have*

$$f(x_k) \leqslant f(x_0) - \lambda \sum_{r=0}^{k-2} \sigma(t_r) - \sigma(t_{k-1}) \leqslant f(x_0) - \lambda \sum_{r=0}^{k-1} \sigma(t_r). \tag{14}$$

Proof. We prove (14) by induction.

If $k = 1$, from (12, 13), we have from $\lambda \leqslant 1$ that

$$f(x_1) \leqslant f(x_0) - \sigma(t_0) \leqslant f(x_0) - \lambda\sigma(t_0).$$

Assume (14) holds for $1, 2, \ldots, k$, we consider two cases:

Case 1: $\max[f(x_k), \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})] = f(x_k)$, from (12, 13), we have:

$$\begin{aligned} f(x_{k+1}) = f(x_k + \alpha_k d_k) &\leqslant f(x_k) - \sigma(t_k) \\ &\leqslant f(x_0) - \lambda \sum_{r=0}^{k-1} \sigma(t_r) - \sigma(t_k) \\ &\leqslant f(x_0) - \lambda \sum_{r=0}^{k} \sigma(t_r). \end{aligned}$$

Case 2: $\max[f(x_k), \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})] = \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})$, let $q = \min[k, M-1]$, again from (12, 13), we have

$$\begin{aligned} f(x_{k+1}) = f(x_k + \alpha_k d_k) &\leqslant \sum_{p=0}^{q} \lambda_{kp} f(x_{k-p}) - \sigma(t_k) \\ &\leqslant \sum_{p=0}^{q} \lambda_{kp} \left(f(x_0) - \lambda \sum_{r=0}^{k-p-2} \sigma(t_r) - \sigma(t_{k-p-1}) \right) - \sigma(t_k). \end{aligned}$$

Using $(0, 1, 2, \ldots, q) \times (0, 1, 2, \ldots, k-q-2) \subset \{(p, r); 0 \leqslant p \leqslant q, 0 \leqslant r \leqslant k-q-2\}$. $\sum_{p=0}^{q} \lambda_{kp} = 1,\ \lambda_{kp} \geqslant \lambda$, we have

$$\begin{aligned} f(x_{k+1}) &\leqslant f(x_0) - \lambda \sum_{r=0}^{k-q-2} \left(\sum_{p=0}^{q} \lambda_{kp} \right) \sigma(t_r) - \sum_{p=0}^{q} \lambda_{kp} \sigma(t_{k-p-1}) - \sigma(t_k) \\ &\leqslant f(x_0) - \lambda \sum_{r=0}^{k-q-2} \sigma(t_r) - \lambda \sum_{r=k-q-1}^{k-1} \sigma(t_r) - \sigma(t_k) \\ &= f(x_0) - \lambda \sum_{r=0}^{k-1} \sigma(t_r) - \sigma(t_k) \\ &\leqslant f(x_0) - \lambda \sum_{r=0}^{k} \sigma(t_r). \end{aligned}$$

This completes the proof. □

Now we establish the global convergence properties of optimization with our new nonmonotonte F-rule.

Theorem 1. *Assume Assumption* 1 *hold, if the search direction d_k satisfies condition* (3). *Then $\{x_k\} \subseteq \mathscr{L}$ and*

$$\lim_{k \to \infty} \|g_k\| = 0. \tag{15}$$

Proof. From Lemma 1 we know that $x_k \in \mathscr{L}$ for all k. Since $f(x)$ is bounded below on $\mathscr{L}$, hence Lemma 1 means

$$\lambda \sum_{r=0}^{k} \sigma(t_r) \leqslant f(x_0) - f(x_{k+1}),$$

let $k \to \infty$, we have

$$\lambda \sum_{r=0}^{\infty} \sigma(t_r) < \infty.$$

Hence, we have

$$\lim_{k\to\infty} \sigma(t_k) = 0,$$

which means from Definition 1 that

$$\lim_{k\to\infty} t_k = \lim_{k\to\infty} \frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|} = 0.$$

Using condition (3) we deduce

$$\lim_{k\to\infty} \sigma(\|g_k\|) = 0,$$

which implies (15) holds. The proof is completed. □

As a conclusion of this section, we consider the new nonmonotone memoryless quasi-Newton type methods:

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

where α_k is generated by nonmonotone F-rule (12) and

$$d_k = -B_k^{-1} g_k, \quad d_0 = g_0, \tag{16}$$

where B_k is updated by the following Perry and Shanno formula:

$$B_{k+1} = \frac{\|y_k\|^2}{y_k^{\mathrm{T}} s_k} I + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k} - \frac{\|y_k\|^2}{s_k^{\mathrm{T}} y_k \|s_k\|^2} s_k s_k^{\mathrm{T}} \tag{17}$$

with

$$s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k.$$

If we use reverse form H_k of B_k, then we can obtain the formula of H_k.

$$H_{k+1} = \frac{y_k^{\mathrm{T}} s_k}{\|y_k\|^2} I + 2\frac{s_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k} - \frac{1}{\|y_k\|^2}(y_k s_k^{\mathrm{T}} + s_k y_k^{\mathrm{T}}) \tag{18}$$

and then the next iteration can be written as

$$\begin{aligned} d_{k+1} &= -B_{k+1}^{-1} g_{k+1} = -H_{k+1} g_{k+1} \\ &= \frac{y_k^{\mathrm{T}} s_k}{\|y_k\|^2} g_{k+1} + \left(\frac{y_k^{\mathrm{T}} g_{k+1}}{\|y_k\|^2} - 2\frac{s_k^{\mathrm{T}} g_{k+1}}{y_k^{\mathrm{T}} s_k} \right) s_k + \frac{s_k^{\mathrm{T}} g_{k+1}}{\|y_k\|^2} y_k. \end{aligned}$$

The methods were originated with the work of Perry [11] and Shanno [13,14], and have been developed by many researches, see [1,2] for example. Recently, Liu and Jing [9] analyzed the convergence of the method with nonmonotone

Wolfe line search. In what follows, we will analyze the global convergence of the method with our new nonmonotone method. Compared with [9], we obtain the strong convergence property (8) instead of the weak property (7).

Assumption 2. (1) $f(x)$ is twice continuously differentiable,
(2) There exist positive constants c_3, c_4 such that

$$c_3\|z\|^2 \leqslant z^{\mathrm{T}} G(x) z \leqslant c_4 \|z\|^2,$$

where $G(x) = \nabla^2 f(x)$.

Lemma 2 (*Powell [12]*). *Consider the iteration of form* (13), *if Assumption 2 holds. Then there exists a positive constant c_5 such that*

$$\frac{\|y_k\|^2}{y_k^{\mathrm{T}} s_k} \leqslant c_5, \quad k = 1, 2, \ldots .$$

Theorem 2. *Let $f(x)$ satisfy Assumptions 1 and 2, for any $x_0 \in R^n$, $B_0 \in R^{n\times n}$ symmetric positive defined. Let $\{x_k\}$ be an infinitely sequence generated by memoryless quasi-Newton method with the rule* (12), (13). *Then* (8) *holds.*

Proof. From Lemma 1, we have

$$f(x_k) \leqslant f(x_0) - \lambda \sum_{r=0}^{k-1} \sigma(t_r),$$

let $k \to \infty$, from Assumption 1, we have

$$\lim_{k\to\infty} \sigma(t_k) = 0.$$

Hence, according to Definition 1, we have

$$\lim_{k\to\infty} \frac{g_k^{\mathrm{T}} H_k g_k}{\|d_k\|} = \lim_{k\to\infty} t_k = 0. \tag{19}$$

By the update formula of B_k and H_k, we have

$$\mathrm{tr}(B_{k+1}) = n\frac{\|y_k\|^2}{y_k^{\mathrm{T}} s_k} \tag{20}$$

and

$$\mathrm{tr}(H_{k+1}) = (n-2)\frac{y_k^{\mathrm{T}} s_k}{\|y_k\|^2} + 2\frac{\|s_k\|^2}{y_k^{\mathrm{T}} s_k} \tag{21}$$

From Assumption 2, we have

$$y_k^{\mathrm{T}} s_k = \int_0^1 s_k^{\mathrm{T}} G(x_k + t s_k) s_k \,\mathrm{d}t \geqslant c_3 \|s_k\|^2,$$

from which and the Cauchy–Schwartz inequality, we have

$$\frac{y_k^{\mathrm{T}} s_k}{\|y_k\|^2} \leqslant \frac{\|s_k\|^2}{y_k^{\mathrm{T}} s_k} \leqslant \frac{1}{c_3}.$$

Hence from Lemma 2, (20, 21) we have

$$\operatorname{tr}(B_{k+1}) \leqslant nc_5, \quad \operatorname{tr}(H_{k+1}) \leqslant n\frac{1}{c_3}.$$

Since

$$\frac{\|d_k\|^2}{g_k^{\mathrm{T}} H_k g_k} = \frac{\|H_k g_k\|^2}{g_k^{\mathrm{T}} H_k g_k} \leqslant \operatorname{tr}(H_k),$$

we have

$$\frac{(g_k^{\mathrm{T}} H_k g_k)^2}{\|d_k\|^2} \geqslant \frac{g_k^{\mathrm{T}} H_k g_k}{\operatorname{tr}(H_k)}.$$

Since

$$g_k^{\mathrm{T}} H_k g_k \geqslant \frac{\|g_k\|^2}{\operatorname{tr}(B_k)},$$

we have

$$\frac{(g_k^{\mathrm{T}} H_k g_k)^2}{\|d_k\|^2} \geqslant \frac{\|g_k\|^2}{\operatorname{tr}(B_k)\operatorname{tr}(H_k)} \geqslant \frac{c_3\|g_k\|^2}{n^2 c_5}.$$

Hence from (19), we get the desired result. This completes the proof. □

4. Numerical tests

In this section, we report the numerical results obtained for a set of standard tests problems. The following algorithm model is a modified version Newton's method in [6], here we replace the nonmonotone line search in Step 6 by our new nonmonotone line search technique.

Algorithm model 1

Data: x_0, Positive constant M, constant $\varepsilon > 0$, $c_6 > 0$, $\gamma_1 \in (0, 1)$, $\sigma \in (0, 1)$, $m(0) = 1$, $k = 0$.

Step 1: Compute g_k, if $\|g_k\| \leqslant \varepsilon$, stop, else computation the Hessian matrix H_k of f_k, if H_k is singular, set $d_k = -g_k$, $m(k) = 1$ and go to Step 4.

Step 2: Compute $d_k = -H_k^{-1} g_k$, if $|g_k^{\mathrm{T}} d_k| < c_6\|g_k\|^2$, set $d_k = -g_k$, $m(k) = 1$ and go to Step 4.

Step 3: If $g_k^{\mathrm{T}} d_k > 0$, set $d_k = -d_k$.

Step 4: Set $\alpha = 1$.

Step 5: Compute $f_\alpha = f(x_k + \alpha d_k)$, let $\lambda \in (0, 1]$, choose

$$\lambda_{kr} \geqslant \lambda, \quad r = 1, 2, \ldots, m(k) - 1 \quad \text{and} \quad \sum_{r=0}^{m(k)-1} \lambda_{kr} = 1.$$

If

$$f_\alpha \leqslant \max\left[f(x_k), \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})\right] + \gamma_1 \alpha g_k^{\mathrm{T}} d_k,$$

set $f_{k+1} = f_\alpha$, $x_{k+1} = x_k + \alpha d_k$, $k = k + 1$, $m(k) = \min[k + 1, M]$ and go to Step 1.

Step 6: Set $\alpha = \sigma\alpha$ and go to Step 5.

Next we give the Perry and Shanno algorithm model (see [9]) with our new nonmonotone line search technique.

Algorithm model 2

Step 1: Given $x_0 \in R^n$, positive constant M, constant $\varepsilon > 0$, $c_6 > 0$, $\gamma_1 \in (0, 1)$, $\sigma \in (0, 1)$, $d_0 = -g_0$, $m(0) = 1$, $k = 0$.

Step 2: Let $\lambda \in (0, 1]$, choose

$$\lambda_{kr} \geqslant \lambda, \quad r = 1, 2, \ldots, m(k) - 1 \quad \text{and} \quad \sum_{r=0}^{m(k)-1} \lambda_{kr} = 1.$$

Set $\alpha = 1$, $m(k) = \min[k + 1, M]$ if

$$f(x_k + \alpha d_k) \leqslant \max\left[f(x_k), \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})\right] + \gamma\alpha g_k^{\mathrm{T}} d_k$$

does not hold, set $\alpha := \sigma\alpha$.

Step 3: Set $x_{k+1} = x_k + \alpha_k d_k$ and compute $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$.

Step 4: If $\|g_k\| \leqslant \varepsilon$, stop.

Step 5: Update the matrix H_k by the Perry and Shanno formula

$$H_{k+1} = \frac{y_k^{\mathrm{T}} s_k}{\|y_k\|^2} I + 2\frac{s_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k} - \frac{1}{\|y_k\|^2}(s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}),$$

set $k := k + 1$, compute $d_k = -H_k g_k$, if $|g_k^{\mathrm{T}} d_k| < c_6\|g_k\|^2$, set $d_k = -g_k$, go to Step 2.

The algorithms has been tested on the following set of problems.

Problem 1. Scaled Rosenbrock function

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2,$$

$$x_0 = [-1.2, 1]^{\mathrm{T}}.$$

Problem 2. Wood function

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3^2 - x_4)^2 + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1),$$

$$x_0 = [-3, -1, -3, -1]^{\mathrm{T}}.$$

Problem 3. Powell singular function

$$f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4,$$

$$x_0 = [3, -1, 0, 1]^{\mathrm{T}}.$$

The numerical experiments have been performed for $M = 1$–10. Typical parameters are $c_6 = 10^{-5}$, $\gamma_1 = 10^{-3}$, $\sigma = 0.5$, $\varepsilon = 10^{-5}$, $\lambda_{kr} = 1/m(k)$, $(r = 0, 1, \ldots, m(k) - 1)$. In particular we report, for each problem the number n_g of gradient evaluations, the number n_f of the function evaluations and the value $f(\widehat{x})$ of the objective function at the solution found $\widehat{x}$.

For problem 1, Table 1 shows that the nonmonotone line search techniques with $M = 2, 3, \ldots, 10$ are all better than the monotone technique under algorithm model 1. Under algorithm model 2, although not all results of nonmonotone technique are better than those of monotone technique, there exists a best $M = 3$, which means that the behaviors of the nonmonotone technique depend on the choice of the parameter M.

Figures 1–3 shows the behaviors of $\{f_k\}$ for the cases $M = 1, 4, 10$. From the figures, we can see that the nonmonotone technique for problem 1 permits large step than that of monotone case.

For problem 2, Table 2 shows that the nonmonotone line search techniques with $M = 2, 3, \ldots, 10$ are all better than the monotone technique under algorithm model 1. Under algorithm model 2, some of the results of nonmonotone technique are better than those of monotone technique, and there also exists a best one ($M = 7$) among them.

For problem 3, Table 3 shows for all M, the function sequence $\{f_k\}$ is descent monotonically under algorithm model 1, hence the results are independent on M. Under algorithm model 2, the nonmonotone line search techniques with $M = 2, 3, \ldots, 10$ are all better than the monotone technique.

Table 1
Results for problem 1

M	$(A1)\, n_g/n_f$	$f(\widehat{x})$	$(A2)\, n_g/n_f$	$f(\widehat{x})$
1	21/28	3.7440e − 21	60/124	7.0378e − 16
2	19/27	8.3425e − 20	62/127	1.7700e − 13
3	19/27	8.3425e − 20	46/78	5.6257e − 13
4	15/22	1.8674e − 12	65/99	1.1584e − 13
5	15/22	1.8674e − 12	67/100	2.2657e − 11
6	15/22	1.8674e − 12	73/99	2.8215e − 12
7	15/22	1.8674e − 12	73/99	2.8215e − 12
8	15/22	1.8674e − 12	73/99	2.8215e − 12
9	15/22	1.8674e − 12	76/104	8.1686e − 12
10	13/19	8.1081e − 14	76/100	8.1803e − 12

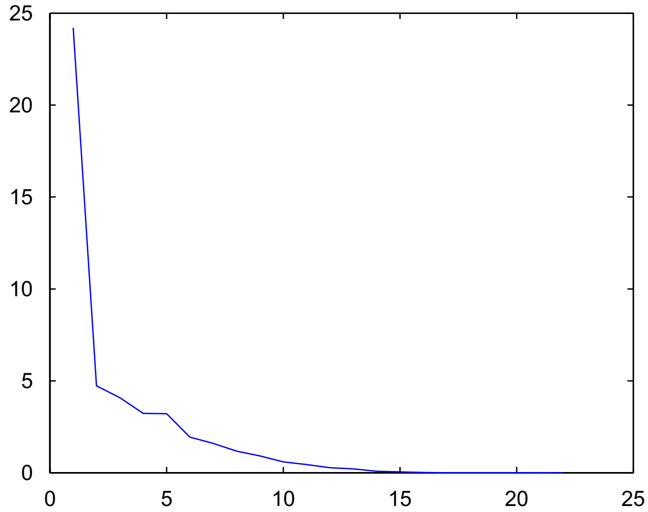

Fig. 1. $M = 1$ for Algorithm 1.

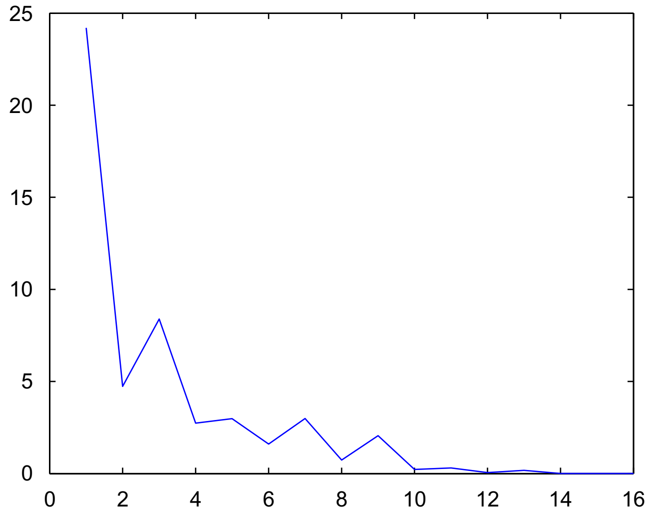

Fig. 2. $M = 4$ for Algorithm 1.

Table 2
Results for problem 2

M	$(A1)\, n_g/n_f$	$f(\widehat{x})$	$(A2)\, n_g/n_f$	$f(\widehat{x})$
1	38/67	6.1558e − 16	140/183	3.0637e − 11
2	38/67	6.1558e − 16	127/166	4.7937e − 11
3	36/51	9.7771e − 15	127/156	3.4821e − 12
4	35/62	1.2084e − 14	140/177	2.6272e − 12
5	36/66	1.1041e − 15	153/190	4.8199e − 13
6	34/53	4.6715e − 16	161/201	3.4610e − 13
7	31/45	1.7486e − 15	118/140	1.3734e − 12
8	31/45	1.7486e − 15	220/262	2.1010e − 13
9	29/37	3.1698e − 16	213/260	3.6950e − 12
10	28/32	2.2741e − 19	213/260	3.6950e − 12

Table 3
Results for problem 3

M	$(A1)\, n_g/n_f$	$f(\widehat{x})$	$(A2)\, n_g/n_f$	$f(\widehat{x})$
1	35/36	2.0135e − 08	357/415	1.8171e − 10
2	35/36	2.0135e − 08	201/235	1.3280e − 09
3	35/36	2.0135e − 08	197/229	1.5518e − 10
4	35/36	2.0135e − 08	122/136	2.2829e − 09
5	35/36	2.0135e − 08	122/229	2.2829e − 09
6	35/36	2.0135e − 08	227/250	5.2056e − 09
7	35/36	2.0135e − 08	157/168	2.6327e − 09
8	35/36	2.0135e − 08	157/168	2.6327e − 13
9	35/36	2.0135e − 08	216/234	1.6598e − 08
10	35/36	2.0135e − 08	319/344	7.1843e − 09

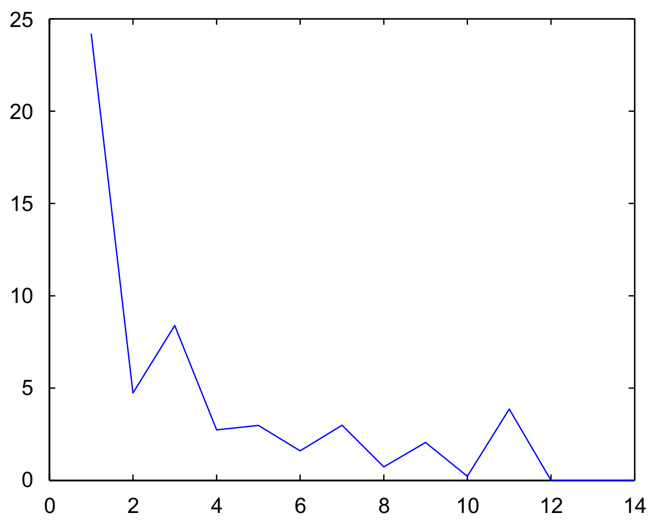

Fig. 3. $M = 10$ for Algorithm 1.

5. Conclusions

In this paper, we propose a new nonmonotone line search technique for unconstrained optimization, and the global convergence can be established under weaker assumptions than those of the existed nonmonotone line search techniques. From the numerical results, we can see that the new nonmonotone technique is very effective.

References

[1] A. Buckley, Combining conjugate gradient and quasi-Newton minimization algorithm, Math. Programming 15 (1978) 200–210.

[2] R.H. Byrd, J. Nocedal, R.B. Schnabel, Representations of quasi-Newton matrices their use in limited memory methods, Math. Programming 63 (1994) 129–156.

[3] Y.H. Dai, On the nonmonotone line search, J. Optim. Theory Appl. 112 (2001) 315–330.

[4] J.E. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[5] R. Fletcher, Practical Methods of Optimization, second ed., Wiley, New York, 1987.

[6] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's methods, SIAM J. Numer. Anal. 23 (1986) 707–716.

[7] L. Grippo, F. Lampariello, S. Lucidi, A truncated Newton method with nonmonotone line search for unconstrained optimization, J. Optim. Theory Appl. 60 (1989) 410–419.

[8] G. Liu, J. Han, D. Sun, Global convergence of BFGS algorithm with nonmonotone line search, Optimization 34 (1995) 147–159.

[9] G.H. Liu, L. Jing, Convergence of nonmonotone Perry and Shanno methods for optimization, Comput. Optim. and Appl. 16 (2000) 159–172.

[10] E.R. Panier, A.L. tits, Avoiding Maratos effect by means of nonmonotone line search constrained problems, SIAM J. Numer. Anal. 28 (1991) 1183–1190.

[11] J.M. Perry, A class of conjugate gradient algorithms with a step variable metric memory, Discussion Paper 269, Center for Mathematical Studies in Economical and Management Science, Northwest University, Evanston, 1977.

[12] M.J.D. Powell, Some global convergence properties of variable metric algorithm for minimization without exact line searches, in: R.W. Cottle, C.E. Lemke, (Eds.), Nonlinear programming, SIAM-AMS proceeding, vol. 9, 1976, pp. 53–72.

[13] D.F. Shanno, On the convergence of a new conjugate gradient algorithm, SIAM J. Numer. Anal. 15 (1978) 1247–1257.

[14] D.F. Shanno, Conjugate gradient method with inexact searches, Math. of Oper. Res. 3 (1978) 244–256.

[15] W. Sun, Y. Yuan, Optimization Theory and Methods, Science Press, Beijing, 1997.

[16] W.Y. Sun, J.Y. Han, J. Sun, Global convergence of non-monotone descent methods for unconstrained optimization problems, J. Comput. Appl. Math. 146 (2002) 89–98.

[17] P.L. Toint, An assessment of nonmonotone line search technique for unconstrained optimization, SIAM J. on Scientific Comput. 17 (1996) 725–739.

[18] P.L. Toint, A nonmonotone trust region algorithm for nonlinear programming subject to convex constraints, Math. Programming 77 (1997) 69–94.

[19] M. Ulbrich, Nonmonotone trust region methods for bound-constrained semi-smooth equation with application to nonlinear complementarity problems, SIAM J. on Optimiz. 11 (2001) 889–917.