Numer Algor (2011) 57:1–25
DOI 10.1007/s11075-010-9408-7

ORIGINAL PAPER

Modified nonmonotone Armijo line search for descent method

Zhenjun Shi · Shengquan Wang

Received: 23 February 2009 / Accepted: 20 June 2010 /
Published online: 3 July 2010

Abstract Nonmonotone line search approach is a new technique for solving optimization problems. It relaxes the line search range and finds a larger step-size at each iteration, so as to possibly avoid local minimizer and run away from narrow curved valley. It is helpful to find the global minimizer of optimization problems. In this paper we develop a new modification of matrix-free nonmonotone Armijo line search and analyze the global convergence and convergence rate of the resulting method. We also address several approaches to estimate the Lipschitz constant of the gradient of objective functions that would be used in line search algorithms. Numerical results show that this new modification of Armijo line search is efficient for solving large scale unconstrained optimization problems.

Keywords Unconstrained optimization · Nonmonotone line search · Convergence

Mathematics Subject Classifications (2010) 90C30 · 65K05 · 49M37

The work was supported in part by Rackham Faculty Research Grant of University of Michigan, USA.

Z. Shi (✉)
Department of Mathematics and Computer Science,
Central State University, Wilberforce, OH 45384, USA
e-mail: zshi@centralstate.edu

S. Wang
Department of Computer and Information Science,
The University of Michigan, Dearborn, MI 48128, USA
e-mail: shqwang@umd.umich.edu

1 Introduction

Let R^n be an n-dimensional Euclidean space and $f : R^n \to R^1$ be a continuously differentiable function, where n is a positive integer. Line search method for the unconstrained minimization problem

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

is defined by the equation

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

where $x_0 \in R^n$ is an initial point, d_k a descent direction of $f(x)$ at x_k, and α_k a step-size. Throughout this paper, x^* refers to the minimizer or stationary point of (1). We denote $f(x_k)$ by f_k, $f(x^*)$ by f^*, and $\nabla f(x_k)$ by g_k, respectively.

Choosing a search direction d_k and determining a step-size α_k along the direction at each iteration are the main task in line search methods [19, 22]. The search direction d_k is generally required to satisfy

$$g_k^T d_k < 0, \tag{3}$$

which guarantees that d_k is a descent direction of $f(x)$ at x_k [19, 35]. In order to guarantee the global convergence, we sometimes require that d_k satisfies the sufficient descent condition

$$g_k^T d_k \le -c\|g_k\|^2, \tag{4}$$

where $c > 0$ is a constant. Moreover, we often choose d_k to satisfy the angle property

$$\cos < -g_k, d_k >= -\frac{g_k^T d_k}{\|g_k\| \cdot \|d_k\|} \ge \tau_0, \tag{5}$$

where $\tau_0 : 1 \ge \tau_0 > 0$ is a constant.

Once a search direction is determined at each step, traditional line search requires the function value to decrease monotonically, namely

$$f(x_{k+1}) < f(x_k), k = 0, 1, 2, ..., \tag{6}$$

whenever $x_k \ne x^*$. Many line searches can guarantee the descent property (6). For example, exact line search ([31, 34–37], Armijo line search (see [1]), Wolfe line search [38, 39]) and Goldstein line search [14], etc. Here are the commonly-used line searches.

(a) **Minimization.** At each iteration, α_k is selected so that

$$f\big(x_k + \alpha_k d_k\big) = \min_{\alpha > 0} f\big(x_k + \alpha d_k\big). \tag{7}$$

(b) **Approximate minimization.** At each iteration, α_k is selected so that

$$\alpha_k = \min\left\{\alpha | g\left(x_k + \alpha d_k\right)^T d_k = 0, \alpha > 0\right\}. \tag{8}$$

(c) **Armijo [30].** Set scalars s_k, β, $L > 0$, and σ with $s_k = -\frac{g_k^T d_k}{L\|d_k\|^2}$, $\rho \in (0, 1)$, and $\sigma \in \left(0, \frac{1}{2}\right)$. Let α_k be the largest one in $\left\{s_k, s_k\rho, s_k\rho^2, ...\right\}$ for which

$$f_k - f\left(x_k + \alpha d_k\right) \geq -\sigma\alpha g_k^T d_k, \tag{9}$$

(d) **Limited minimization.** Set $s_k = -\frac{g_k^T d_k}{L\|d_k\|^2}$, α_k is defined by

$$f\left(x_k + \alpha_k d_k\right) = \min_{\alpha\in[0,s_k]} f\left(x_k + \alpha d_k\right), \tag{10}$$

where $L > 0$ is a constant.

(e) **Goldstein.** A fixed scalar $\sigma \in \left(0, \frac{1}{2}\right)$ is selected and α_k is chosen to satisfy

$$\sigma \leq \frac{f\left(x_k + \alpha_k d_k\right) - f_k}{\alpha_k g_k^T d_k} \leq 1 - \sigma. \tag{11}$$

It is possible to show that, if f is bounded from below, then there exists an interval of step-sizes α_k for which the relation above is satisfied, and there are fairly simple algorithms for finding such a step-size through a finite number of arithmetic operations.

(f) **Strong Wolfe.** α_k is chosen to satisfy simultaneously

$$f_k - f\left(x_k + \alpha_k d_k\right) \geq -\sigma\alpha_k g_k^T d_k \tag{12}$$

and

$$\left|g\left(x_k + \alpha_k d_k\right)^T d_k\right| \leq -\beta g_k^T d_k, \tag{13}$$

where σ and β are some scalars with $\sigma \in \left(0, \frac{1}{2}\right)$ and $\beta \in (\sigma, 1)$.

(g) **Wolfe.** α_k is chosen to satisfy (12) and

$$g\left(x_k + \alpha_k d_k\right)^T d_k \geq \beta g_k^T d_k. \tag{14}$$

Some important global convergence results for various methods using the above mentioned specific line search procedures have been given [3, 7, 20, 21, 31, 38, 39]. In fact, the above-mentioned line searches can make the related methods be monotone descent methods for unconstrained optimization problems [34–37].

Nonmonotone line search methods have also been investigated by many authors [5, 11, 27]. In fact, Barzilai-Borwein method [2, 4, 25, 26] is a non-monotone descent method. The nonmonotone line search does not need to satisfy the condition (6) in many situations; as a result, it is helpful to overcome the case where the sequence of iterates run into the bottom of a curved narrow valley, a common occurrence in practical nonlinear problems.

Grippo et al. [11–13] proposed a new line search technique, nonmonotone line search, for Newton-type method and truncated Newton method. Liu et al. [6] used nonmonotone line search to the BFGS quasi-Newton method. Deng [6], Toint [32, 33], and many other researchers [15–17, 40, 41] studied nonmonotone trust region methods. Moreover, general nonmonotone line search technique has been investigated by many authors [5, 23, 27, 28]. In particular, Sun et al. [27] studied general nonmonotone line search methods by using inverse continuous modulus, which may be described as follows.

(A) **Nonmonotone Armijo line search.** Let $s > 0,\ \sigma \in (0, 1),\ \rho \in (0, 1)$ and let M be a nonnegative integer. For each k, let $m(k)$ satisfy

$$m(0) = 0,\ 0 \le m(k) \le \min[m(k-1), M], \forall k \ge 1. \tag{15}$$

Let α_k be the largest one in $\{s, s\rho, s\rho^2, ...\}$ such that

$$f\left(x_k + \alpha d_k\right) \le \max_{0 \le j \le m(k)} [f(x_{k-j})] + \sigma\alpha g_k^T d_k. \tag{16}$$

(B) **Nonmonotone Goldstein line search.** Let $\sigma \in \left(0, \frac{1}{2}\right)$, α_k is selected to satisfy

$$f\left(x_k + \alpha_k d_k\right) \le \max_{0 \le j \le m(k)} [f_{k-j}] + \sigma\alpha_k g_k^T d_k, \tag{17}$$

and

$$f\left(x_k + \alpha_k d_k\right) \ge \max_{0 \le j \le m(k)} [f_{k-j}] + (1-\sigma)\alpha_k g_k^T d_k. \tag{18}$$

(C) **Nonmonotone Wolfe line search.** Let $\sigma \in \left(0, \frac{1}{2}\right)$, and $\beta \in (\sigma, 1)$, α_k is selected to satisfy

$$f\left(x_k + \alpha_k d_k\right) \le \max_{0 \le j \le m(k)} [f_{k-j}] + \sigma\alpha_k g_k^T d_k, \tag{19}$$

and

$$g\left(x_k + \alpha_k d_k\right)^T d_k \ge \beta g_k^T d_k. \tag{20}$$

Sun et al. [27] established general convergent algorithms with the above three nonmonotone line searches. Armijo-type line search is easier to implement on computer than other line searches do. However, when we use Armijo-type line search to find an acceptable step size at each iteration, two questions need to be answered. How to determine an adequate initial step size s is the first question. If s is too large, then the number of backtracking will be large.

If s is too small, then the resulting method will converge slowly. Shi et al. developed a more general nonmonotone line search method [28] and gave a suitable estimation of s. The second question is how to determine the backtracking ratio $\rho \in (0, 1)$. If ρ is too close to 1, then the number of backtracking will be increased greatly, the computational cost will be increased substantially. If ρ is too close to 0, then the smaller step size will be appeared at iteration and possibly decrease the convergence rate.

In this paper we develop a new modification of matrix-free nonmonotone Armijo line search and analyze the global convergence of resulting line search methods. This development enables us to choose a larger step-size at each iteration and maintain the global convergence. We also address several ways to estimate the Lipschitz constant of the gradient of objective functions that is used to estimate the initial step size at each iteration. Numerical results show that the new modification of nonmonotone Armijo line search is efficient for solving large scale unconstrained optimization problems and corresponding algorithms is superior to some existing similar line search methods.

The rest of this paper is organized as follows. In Section 2 we describe the new modicifcation of nonmonotone Armijo line search. In Section 3 we analyze the global convergence of the resulting method. Section 4 is devoted to the analysis of convergence rate. In Section 5, we propose some ways to estimate the parameters used in the new version of Armijo line search and report some numerical results.

2 Nonmonotone Armijo line search

Throughout this paper we assume

- (H1) The objective function $f(x)$ has a lower bound on R^n.
- (H2) The gradient $g(x) = \nabla f(x)$ of $f(x)$ is uniform continuous in an open convex set B that contains the level set $L(x_0) = \{x \in R^n | f(x) \leq f(x_0)\}$ with x_0 given.
- (H2)$'$ The gradient $g(x) = \nabla f(x)$ of $f(x)$ is Lipschitz continuous in an open convex set B that contains the level set $L(x_0)$, i.e., there exists $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in B.$$

Obviously, if (H2)$'$ holds, then (H2) holds, i.e., (H2) is weaker than (H2)$'$. We describe line search method as follows.

Algorithm (A) Given some parameters and the initial point x_0, $k := 0$.

Step 1. If $\|g_k\| = 0$, then stop; else go to Step 2;
Step 2. $x_{k+1} = x_k + \alpha_k d_k$ where d_k is a descent direction of $f(x)$ at x_k, α_k is selected by some line search.
Step 3. $k := k + 1$, go to Step 1.

In this paper, we would like to investigate how to choose the step-size α_k instead of the choices of d_k. We develop a new modification of nonmonotone Armijo line search approach and find that the step-size defined in the modified nonmonotone line search is larger than that defined in the original nonmonotone line searches.

(D) **Modified nonmonotone Armijo line search.** Set scalars s_k, ρ, μ, $L_k > 0$, and σ with $s_k = -\frac{g_k^T d_k}{L_k \|d_k\|^2}$, $\sigma \in \left(0, \frac{1}{2}\right)$, $\rho \in (0, 1)$ and $\mu \in [0, 2)$. M is a nonnegative integer and $m(k)$ satisfies (15). Let α_k be the largest one in $\left\{s_k, s_k\rho, s_k\rho^2, ...\right\}$ such that

$$f\left(x_k + \alpha d_k\right) - \max_{0 \le j \le m(k)} [f_{k-j}] \le \sigma\alpha \left[g_k^T d_k + \frac{1}{2}\alpha\mu L_k \|d_k\|^2\right]. \qquad (21)$$

Remark 1 Suppose that α_k is defined by line search (A) and α_k' is defined by line search (D), then $\alpha_k' \ge \alpha_k$. In fact, if $L_k \equiv L$ and there exists α_k satisfying (16), then α_k is certainly to satisfy (21). Avoiding the memory and computation of a matrix is the advantage of this modified nonmonotone line search.

Unlike the nonmonotone line search in [28], it is a matrix-free line search and can be used to solve large scale optimization problems. If $M \equiv 0$, then the modified nonmonotone line search will reduce to the monotone line search in [29].

Moreover, if $\mu = 0$, then the line search (D) will reduce to the nonmonotone Armijo line search (A).

In Algorithm (A), the corresponding algorithm with line search (D) is denoted by Algorithm (D). In what follows, we will analyze the global convergence of Algorithm (D).

3 Global convergence

In order to analyze the global convergence, we first describe the general theorem.

Theorem 3.1 *Assume that* $(H1)$ *and* $(H2)$ *hold, the search direction* d_k *satisfies* (3) *and* α_k *is determined by the modified nonmonotone Armijo line search* (D), *Algorithm* (D) *generates an infinite sequence* $\{x_k\}$ *and* $\{L_k\}$ *satisfies*

$$u_0 \le L_k \le U_0, \qquad (22)$$

with $0 < u_0 \le U_0$. *Then, we have*

$$\liminf \left(-\frac{g_k^T d_k}{\|d_k\|}\right) = 0. \qquad (23)$$

In order to prove (23), we assume that the contrary proposition holds, i.e., there exists a positive number $\epsilon > 0$ such that

$$-\frac{g_k^T d_k}{\|d_k\|} \geq \epsilon, \forall k. \tag{24}$$

Therefore, it follows from (21), (22), and (24), that

$$\begin{aligned} f(x_k + \alpha d_k) - \max_{0 \leq j \leq m(k)} [f_{k-j}] &\leq \sigma\alpha \left[g_k^T d_k + \frac{1}{2}\alpha\mu L_k \|d_k\|^2 \right] \\ &\leq \sigma\alpha \left[g_k^T d_k + \frac{1}{2} s_k \mu L_k \|d_k\|^2 \right] \\ &= \frac{\sigma\left(1 - \frac{1}{2}\mu\right)}{L_k} \alpha_k g_k^T d_k \\ &\leq -\frac{\sigma\left(1 - \frac{1}{2}\mu\right)}{U_0} \alpha_k \|d_k\| \cdot \frac{-g_k^T d_k}{\|d_k\|} \\ &\leq -\frac{\sigma\left(1 - \frac{1}{2}\mu\right)\epsilon}{U_0} \alpha_k \|d_k\|, \end{aligned}$$

which results in

$$f(x_k + \alpha d_k) - \max_{0 \leq j \leq m(k)} [f_{k-j}] \leq -\frac{\sigma\left(1 - \frac{1}{2}\mu\right)\epsilon}{U_0} \alpha_k \|d_k\|, \forall k. \tag{25}$$

Lemma 3.1 *Assume that the conditions in Theorem* 3.1 *hold and* (25) *holds. Let*

$$\eta_0 = \frac{\sigma\left(1 - \frac{1}{2}\mu\right)\epsilon}{U_0}.$$

Then,

$$\max_{1 \leq j \leq M} \left[f(x_{Ml+j}) \right] \leq \max_{1 \leq i \leq M} \left[f(x_{M(l-1)+i}) \right] - \eta_0 \min_{0 \leq j \leq M-1} \left(\alpha_{Ml+j-1} \|d_{Ml+j-1}\| \right), \tag{26}$$

and thus

$$\sum_{l=1}^{\infty} \min_{0 \leq j \leq M-1} \left(\alpha_{Ml+j} \|d_{Ml+j}\| \right) < +\infty. \tag{27}$$

Proof By (H1) and (25), it suffices to show that the following inequality holds for $j = 1, 2, ..., M$,

$$f(x_{Ml+j}) \leq \max_{1 \leq i \leq M} \left[f(x_{M(l-1)+i}) \right] - \eta_0 \min_{0 \leq j \leq M-1} \left(\alpha_{Ml+j-1} \|d_{Ml+j-1}\| \right). \tag{28}$$

Since the new modified nonmonotone Armijo line search and (25) imply

$$f\big(x_{Ml+1}\big) \leq \max_{0\leq i\leq m(Ml)} f\big(x_{Ml-i}\big) - \eta_0\big(\alpha_{Ml}\|d_{Ml}\|\big), \tag{29}$$

it follows from this and

$$m(Ml) \leq M$$

that (28) holds for $j = 1$. Suppose that (28) holds for any $1 \leq j \leq M-1$. With the descent property of d_k, we obtain

$$\max_{1\leq i\leq j}\big[f(x_{Ml+i})\big] \leq \max_{1\leq i\leq M}\big[f(x_{M(l-1)+i})\big]. \tag{30}$$

By the new modified Armijo line search, the induction hypothesis,

$$m\big(Ml+j\big) \leq M,$$

(25), and (30), we obtain

$$\begin{aligned} f\big(x_{Ml+j+1}\big) &\leq \max_{0\leq i\leq m(Ml+j)}\big[f(x_{Ml+j-i})\big] - \eta_0\big(\alpha_{Ml+j}\|d_{Ml+j}\|\big) \\ &\leq \max\left\{\max_{1\leq i\leq M}\big[f(x_{M(l-1)+i})\big], \max_{1\leq i\leq j}\big[f(x_{Ml+i})\big]\right\} - \eta_0\big(\alpha_{Ml+j}\|d_{Ml+j}\|\big) \\ &\leq \max_{1\leq i\leq M}\big[f(x_{M(l-1)+i})\big] - \eta_0\big(\alpha_{Ml+j}\|d_{Ml+j}\|\big). \end{aligned}$$

Thus, (28) is also true for $j+1$. By induction, (28) holds for $1 \leq j \leq M$. This shows that (26) holds.

Since $f(x)$ is bounded below by (H1), it follows that

$$\max_{1\leq i\leq M}\big[f(x_{Mj+i})\big] > -\infty.$$

By summing (28) over l, we can get

$$\sum_{l=1}^{\infty} \min_{0\leq j\leq M-1}\big(\alpha_{Ml+j}\|d_{Ml+j}\|\big) < +\infty.$$

Therefore, (27) holds. The proof is completed. □

Proof of Theorem 3.1 From Lemma 3.1, there exists an infinite subset K of $\{0, 1, 2, ..., \}$ such that

$$\lim_{k\in K, k\to\infty} \left(\alpha_k \|d_k\|\right) = 0. \tag{31}$$

If there is an infinite subset $K_1 = \{k \in K | \alpha_k = s_k\}$, then we have

$$\begin{aligned} 0 &= \lim_{k\in K_1, k\to\infty} \left(\alpha_k \|d_k\|\right) \\ &= \lim_{k\in K_1, k\to\infty} \left(s_k \|d_k\|\right) \\ &= \lim_{k\in K_1, k\to\infty} \left(-\frac{g_k^T d_k}{L_k \|d_k\|^2} \cdot \|d_k\|\right) \\ &\geq \frac{1}{U_0}\left(-\frac{g_k^T d_k}{\|d_k\|}\right), \end{aligned}$$

which contradicts (24).

If there is an infinite subset $K_2 = \{k \in K | \alpha_k < s_k\}$, then we have $\alpha_k \rho^{-1} \leq s_k$. By the modified nonmonotone Armijo line search (D), we have

$$f\left(x_k + \alpha_k \rho^{-1} d_k\right) - \max_{0\leq j\leq m(k)} [f_{k-j}] > \sigma \alpha_k \rho^{-1} \left[g_k^T d_k + \frac{1}{2}\alpha_k \rho^{-1} \mu L_k \|d_k\|^2\right].$$

Therefore,

$$f\left(x_k + \alpha_k \rho^{-1} d_k\right) - f_k > \sigma \alpha_k \rho^{-1} g_k^T d_k.$$

Using the mean value theorem on the left-hand side of the above inequality, we can find $\theta_k \in [0, 1]$ such that

$$\alpha_k \rho^{-1} g\left(x_k + \theta_k \alpha_k \rho^{-1} d_k\right)^T d_k > \sigma \alpha_k \rho^{-1} g_k^T d_k.$$

Therefore,

$$g\left(x_k + \theta_k \alpha_k \rho^{-1} d_k\right)^T d_k > \sigma g_k^T d_k. \tag{32}$$

By (H2), the Cauchy-Schwartz inequality, (32) and (31), we obtain

$$\begin{aligned} 0 &= \lim_{k\in K_2, k\to\infty} \left\|g\left(x_k + \theta_k \alpha_k \rho^{-1} d_k\right) - g_k\right\| \\ &\geq \lim_{k\in K_2, k\to\infty} \frac{\left[g\left(x_k + \theta_k \alpha_k \rho^{-1} d_k\right) - g_k\right]^T d_k}{\|d_k\|} \\ &> -(1-\sigma) \lim_{k\in K_2, k\to\infty} \frac{g_k^T d_k}{\|d_k\|}, \end{aligned}$$

which also contradicts (24). The contradictions shows that (23) holds. The proof of Theorem 3.1 is completed. □

Lemma 3.2 *Assume that* $(H1)$ *and* $(H2)'$ *hold, the search direction* d_k *satisfies* (3) *and* α_k *is determined by the modified nonmonotone Armijo line search, Algorithm* (D) *generates an infinite sequence* $\{x_k\}$*, and* L_k *satisfies*

$$uL < L_k \le UL, \tag{33}$$

where $u \in (0, 1)$, $U \in [1, +\infty)$*. Then, there exists* $\eta_0' > 0$ *such that for any* $k \ge M$*,*

$$f\left(x_k + \alpha_k d_k\right) - \max_{0 \le j \le m(k)} [f_{k-j}] \le -\eta_0' \left(\frac{g_k^T d_k}{\|d_k\|}\right)^2. \tag{34}$$

Proof Let

$$K_1 = \left\{k |\ \alpha_k = s_k\right\},\ K_2 = \left\{k |\ \alpha_k < s_k\right\}.$$

If $k \in K_1$, then

$$\begin{aligned} f\left(x_k + \alpha_k d_k\right) - \max_{0 \le j \le m(k)} [f_{k-j}] &\le \sigma \alpha_k \left[g_k^T d_k + \frac{1}{2}\alpha_k \mu L_k \|d_k\|^2\right] \\ &= -\sigma \cdot \frac{g_k^T d_k}{L_k \|d_k\|^2} \left[g_k^T d_k - \frac{1}{2}\mu g_k^T d_k\right] \\ &= -\frac{\sigma\left(1 - \frac{1}{2}\mu\right)}{L_k} \cdot \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2}. \end{aligned}$$

Thus,

$$f\left(x_k + \alpha_k d_k\right) - \max_{0 \le j \le m(k)} [f_{k-j}] \le -\frac{\sigma\left(1 - \frac{1}{2}\mu\right)}{L_k} \cdot \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2}, k \in K_1. \tag{35}$$

Letting

$$\eta_k' = -\frac{\sigma\left(1 - \frac{1}{2}\mu\right)}{L_k},\ k \in K_1,$$

by (33), we have

$$\begin{aligned} \eta_k' &= -\frac{\sigma\left(1 - \frac{1}{2}\mu\right)}{L_k} \\ &\le -\frac{\sigma\left(1 - \frac{1}{2}\mu\right)}{UL} \\ &< 0. \end{aligned}$$

Letting

$$\eta' = \frac{\sigma\left(1 - \frac{1}{2}\mu\right)}{UL},$$

by (35), we obtain $\eta'_k \leq -\eta'$ and

$$f_{k+1} - \max_{0 \leq j \leq m(k)} [f_{k-j}] \leq -\eta' \cdot \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2, \ k \in K_1. \tag{36}$$

If $k \in K_2$, then $\alpha_k < s_k$. Thus, $\alpha_k \rho^{-1} \leq s_k$. By the modified nonmonotone Armijo line search (D), we have

$$f\big(x_k + \alpha_k \rho^{-1} d_k\big) - \max_{0 \leq j \leq m(k)} [f_{k-j}] > \sigma \alpha_k \rho^{-1} \left[g_k^T d_k + \frac{1}{2} \alpha_k \rho^{-1} \mu L_k \|d_k\|^2 \right].$$

Therefore,

$$f\big(x_k + \alpha_k \rho^{-1} d_k\big) - f_k > \sigma \alpha_k \rho^{-1} g_k^T d_k.$$

Using the mean value theorem on the left-hand side of the above inequality, we can find $\theta_k \in [0, 1]$ such that

$$\alpha_k \rho^{-1} g\big(x_k + \theta_k \alpha_k \rho^{-1} d_k\big)^T d_k > \sigma \alpha_k \rho^{-1} g_k^T d_k.$$

Therefore,

$$g\big(x_k + \theta_k \alpha_k \rho^{-1} d_k\big)^T d_k > \sigma g_k^T d_k. \tag{37}$$

By (H2)$'$, the Cauchy-Schwartz inequality, and (37), we obtain

$$\begin{aligned} \alpha_k \rho^{-1} L \|d_k\|^2 &\geq \left\| g\left(x_k + \theta_k \alpha_k \rho^{-1} d_k\right) - g_k \right\| \cdot \|d_k\| \\ &\geq \left[g\big(x_k + \theta_k \alpha_k \rho^{-1} d_k\big) - g_k \right]^T d_k \\ &\geq -(1 - \sigma) g_k^T d_k. \end{aligned}$$

Therefore,

$$\alpha_k > -\frac{\rho(1-\sigma)}{L} \cdot \frac{g_k^T d_k}{\|d_k\|^2}, \ k \in K_2. \tag{38}$$

By letting

$$s'_k = -\frac{\rho(1-\sigma)}{L} \cdot \frac{g_k^T d_k}{\|d_k\|^2}, \ k \in K_2,$$

we have

$$s_k > \alpha_k > s'_k, \ k \in K_2. \tag{39}$$

By (21) and (39), we have

$$\begin{aligned} f_{k+1} - \max_{0\le j\le m(k)}[f_{k-j}] &\le \sigma\alpha_k\left[g_k^T d_k + \frac{1}{2}\alpha_k\mu L_k\|d_k\|^2\right] \\ &\le \sigma\alpha_k\left[g_k^T d_k + \frac{1}{2}s_k\mu L_k\|d_k\|^2\right]\} \\ &\le \sigma s_k'\left[g_k^T d_k - \frac{1}{2}\mu g_k^T d_k\right] \\ &= -\frac{\sigma\rho(1-\sigma)(2-\mu)}{2L_k}\cdot\left(\frac{g_k^T d_k}{\|d_k\|}\right)^2. \end{aligned}$$

Letting

$$\eta_k'' = -\frac{\sigma\rho(1-\sigma)(2-\mu)}{2L_k}, \tag{40}$$

we have

$$f_{k+1} - \max_{0\le j\le m(k)}[f_{k-j}] \le \eta_k''\cdot\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2},\ k\in K_2. \tag{41}$$

Let

$$\eta'' = \frac{\sigma\rho(1-\sigma)(2-\mu)}{2UL}.$$

It follows from (33) that $\eta_k'' \le -\eta''$ and

$$f_{k+1} - \max_{0\le j\le m(k)}[f_{k-j}] \le -\eta''\cdot\left(\frac{g_k^T d_k}{\|d_k\|}\right)^2,\ k\in K_2. \tag{42}$$

Letting $\eta_0' = \min(\eta', \eta'')$, by (36) and (42), we have

$$f_{k+1} - \max_{0\le j\le m(k)}[f_{k-j}] \le -\eta_0'\cdot\left(\frac{g_k^T d_k}{\|d_k\|}\right)^2,\ \forall k\ge M. \tag{43}$$

The proof is completed. □

Remark 3.1 This lemma shows that the sequence $\{f_k\}$ is not a decreasing sequence but a decreasing sequence within every $m(k)$ steps.

Lemma 3.3 *Assume that* $(H1)$ *and* $(H2)'$ *hold, the search direction* d_k *satisfies* (3) *and* α_k *is determined by the modified nonmonotone Armijo line search* (D)*, Algorithm* (D) *generates an infinite sequence* $\{x_k\}$*, and* (33) *holds at each iteration. Then,*

$$\max_{1\le j\le M}\left[f(x_{Ml+j})\right] \le \max_{1\le i\le M}\left[f(x_{M(l-1)+i})\right] - \eta_0' \min_{0\le j\le M-1}\left(\frac{g_{Ml+j-1}^T d_{Ml+j-1}}{\|d_{Ml+j-1}\|}\right)^2, \tag{44}$$

and thus

$$\sum_{l=1}^{\infty} \min_{0\le j\le M-1}\left(\frac{g_{Ml+j}^T d_{Ml+j}}{\|d_{Ml+j}\|}\right)^2 < +\infty. \tag{45}$$

The proof is similar to that of Lemma 3.1.

Corollary 3.1 *Assume that* $(H1)$ *and* $(H2)$ *hold, and that the conditions in Theorem* 3.1 *hold. Then*

$$\overline{\lim_{k\to\infty}} f_k \le \lim_{k\to\infty} f(x_{l(k)}), \tag{46}$$

where $l(k)$ *satisfies*

$$f(x_{l(k)}) = \max_{0\le j\le m(k)}[f_{k-j}].$$

Proof Since $m(k+1) \le m(k) + 1$, we have

$$\begin{aligned} f(x_{l(k+1)}) &= \max_{0\le j\le m(k+1)}\left[f(x_{k+1-j})\right] \\ &\le \max_{0\le j\le m(k)+1}\left[f(x_{k+1-j})\right] \\ &= \max\left\{f(x_{l(k)}),\ f(x_{k+1})\right\} \\ &= f(x_{l(k)}). \end{aligned}$$

Thus $\{f(x_{l(k)})\}$ is a monotone non-increasing sequence. (H1) implies that $\{f(x_{l(k)})\}$ has a bound from below and thus it has a limit. By (26), we obtain that (46) holds. □

Corollary 3.2 *Assume that* $(H1)$ *and* $(H2)$ *hold, the search direction* d_k *satisfies* (5) *and* α_k *is determined by the modified nonmonotone Armijo line search* (D)*, Algorithm* (D) *generates an infinite sequence* $\{x_k\}$*, and* (22) *holds at each iteration. Then,*

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

Proof It is easy to prove by (5) and (46). The proof is completed. □

Theorem 3.2 *Suppose that* $(H1)$ *and* $(H2)'$ *hold,* d_k *satisfies* (4) *and*

$$\|d_k\|^2 \le c_1 + c_2 k, \ \forall k, \tag{47}$$

where c_1 *and* c_2 *are positive constants. The step-size* α_k *is determined by the modified nonmonotone Armijo line search* (D) *and* (22) *holds. Algorithm* (D) *generates an infinite sequence* $\{x_k\}$*. Then,*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{48}$$

Proof We proceed by contradiction. Assume that the relation (48) does not hold. Then, there exists some τ such that

$$\|g_k\| \ge \tau, \ \forall k \ge 1. \tag{49}$$

This and (4) imply

$$\left|g_k^T d_k\right| \ge c\tau^2, \ \forall k \ge 1. \tag{50}$$

Then, it follows from (47) and (50) that

$$\begin{aligned}\sum_{l=1}^{\infty} \min_{0\le i\le M-1} \left(\frac{g_{Ml+i}^T d_{Ml+i}}{\|d_{Ml+i}\|}\right)^2 &\ge \sum_{l=1}^{\infty} \frac{c^2\tau^4}{\|d_{Ml+i_0}\|^2} \\ &\ge \sum_{l=1}^{\infty} \frac{c^2\tau^4}{c_1 + c_2(Ml+i_0)} \\ &= +\infty,\end{aligned}$$

where i_0 satisfies

$$\left(\frac{g_{Ml+i_0}^T d_{Ml+i_0}}{\|d_{Ml+i_0}\|}\right)^2 = \min_{0\le i\le M-1} \left(\frac{g_{Ml+i}^T d_{Ml+i}}{\|d_{Ml+i}\|}\right)^2.$$

The above relation contradicts (45). The contradiction shows the truth of (48). □

Remark 3.2 Theorem 3.2 shows that $\{\|g_k\|\}$ is global convergence even if $\|d_k\| \to \infty$ as $k \to +\infty$.

Theorem 3.3 *Suppose that* $(H1)$ *and* $(H2)'$ *hold,* d_k *satisfies* (5)*. The step-size* α_k *is determined by the modified nonmonotone Armijo line search* (D) *and* (33) *holds. Algorithm* (D) *generates an infinite sequence* $\{x_k\}$*. Then,*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{51}$$

Proof By (45), we have

$$\lim_{l\to+\infty}\min_{0\le j\le M-1}\left(\frac{g_{Ml+j}^T d_{Ml+j}}{\|d_{Ml+j}\|}\right)^2=0.$$

By (5), we have

$$\lim_{l\to+\infty}\min_{0\le j\le M-1}\|g_{Ml+j}\|=0.$$

Letting

$$\|g_{Ml+\phi(l)}\|=\min_{0\le j\le M-1}\|g_{Ml+j}\|,$$

we obtain

$$\lim_{l\to\infty}\|g_{Ml+\phi(l)}\|=0. \tag{52}$$

It follows from (H2)′ and (33) that

$$\begin{aligned}\|g_{k+1}\| &= \|g_{k+1}-g_k+g_k\|\\ &\le \|g_{k+1}-g_k\|+\|g_k\|\\ &\le L\alpha_k\|d_k\|+\|g_k\|\\ &\le Ls_k\|d_k\|+\|g_k\|\\ &= L\|g_k\|\left(-\frac{g_k^Td_k}{L_k\|g_k\|\cdot\|d_k\|}\right)+\|g_k\|\\ &\le \left(\frac{L}{L_k}\right)\|g_k\|+\|g_k\|\\ &\le \left(1+\frac{L}{u}\right)\|g_k\|.\end{aligned}$$

By letting $c_3=\left(1+\frac{L}{u}\right)$, we have

$$\|g_{k+1}\|\le c_3\|g_k\|. \tag{53}$$

Therefore, for $1\le i\le M$,

$$\begin{aligned}\|g_{M(l+1)+i}\| &\le c_3\|g_{M(l+1)+i-1}\|\\ &\le \ldots.\\ &\le c_3^{2M}\|g_{Ml+\phi(l)}\|.\end{aligned}$$

By (52), we obtain that (51) holds and the proof is finished. □

4 Convergence rate

In order to analyze the convergence rate, we restrict our discussion to the case of uniformly convex objective functions. We assume that

(H3) f is twice continuously differentiable and uniformly convex on R^n.

Lemma 4.1 *Assume that* $(H3)$ *holds, then* $(H1)$ *and* $(H2)'$ *hold,* $f(x)$ *has a unique minimizer* x^**, and there exists* $0 < m' \leq L$ *such that*

$$m'\|y\|^2 \leq y^T \nabla^2 f(x) y \leq M'\|y\|^2,\ \forall x, y \in R^n; \tag{54}$$

$$\frac{1}{2}m'\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2}M'\|x - x^*\|^2,\ \forall x \in R^n; \tag{55}$$

$$M'\|x - y\|^2 \geq (g(x) - g(y))^T(x - y) \geq m'\|x - y\|^2,\ \forall x, y \in R^n; \tag{56}$$

and thus

$$M'\|x - x^*\|^2 \geq g(x)^T(x - x^*) \geq m'\|x - x^*\|^2,\ \forall x \in R^n. \tag{57}$$

By (57) *and* (56) *we can also obtain from the Cauchy-Schwartz inequality, that*

$$M'\|x - x^*\| \geq \|g(x)\| \geq m'\|x - x^*\|,\ \forall x \in R^n, \tag{58}$$

and

$$\|g(x) - g(y)\| \leq M'\|x - y\|, \forall x, y \in R^n. \tag{59}$$

By (55) *and* (58) *we can also obtain the following relation*

$$\frac{m'}{2M'^2}\|g(x)\|^2 \leq f(x) - f(x^*) \leq \frac{M'}{2m'^2}\|g(x)\|^2. \tag{60}$$

Its proof can be seen from [3, 31].

Theorem 4.1 *Assume that* $(H3)$ *holds, the search direction* d_k *satisfies the angle property* (5)*, the step-size* α_k *is determined by the modified nonmonotone Armijo line search* (D) *and* (33) *holds. Algorithm* (D) *generates an infinite sequence* $\{x_k\}$*. Then,* $\{x_k\} \to x^*$ *at least R-linearly, i.e., there exists* $0 < \omega < 1$ *such that*

$$R_1\{x_k\} = \lim_{k\to+\infty} \|x_k - x^*\|^{\frac{1}{k}} = \omega. \tag{61}$$

Proof Since (H3) implies (H1) and (H2)$'$, the conditions of Theorem 3.3 hold in this case. Thus (51) holds. By (58) it follows that

$$\lim_{k\to\infty} x_k = x^*. \tag{62}$$

At first, we obtain from (60) and (53), that

$$f(x_{k+1}) - f(x^*) \le b\big[f(x_k) - f(x^*)\big], \ \forall k \ge 1, \tag{63}$$

where

$$b = \frac{c_3^2 M'^3}{m'^3} > 1.$$

For any $l \ge 0$, let $\psi(l)$ be any index in $[Ml+1, M(l+1)]$ for which

$$f(x_{\psi(l)}) = \max_{1\le i\le M}\big[f(x_{Ml+i})\big]. \tag{64}$$

By the definition of $\psi(l)$, (5), and (44), we can get

$$f(x_{\psi(l)}) \le f(x_{\psi(l-1)}) - c_6 \min_{0\le j\le M-1} \|g_{Ml+j}\|^2, \tag{65}$$

where

$$c_6 = \tau_0^2 \eta_0'$$

is a positive constant.

Similarly as in the proof of Theorem 3.1 in [5], we can prove that there exist constants c_4 and $c_5 \in (0, 1)$ such that

$$f(x_k) - f(x^*) \le c_4 c_5^k\big[f(x_1) - f(x^*)\big]. \tag{66}$$

By (55) and (66) we have

$$\frac{1}{2} m' \|x_k - x^*\|^2 \le f(x_k) - f(x^*) \le c_4 c_5^k\big[f(x_1) - f(x^*)\big],$$

and thus

$$\|x_k - x^*\| \le \sqrt{\frac{2c_4\big[f(x_1) - f(x^*)\big]}{m'}} \cdot \big(\sqrt{c_5}\big)^k. \tag{67}$$

Letting

$$\omega = \sqrt{c_5},$$

we have

$$\begin{aligned} R_1\{x_k\} &= \lim_{k\to+\infty} \|x_k - x^*\|^{\frac{1}{k}} \\ &= \lim_{k\to+\infty} \left(\sqrt{\frac{2c_4[f(x_1) - f(x^*)]}{m'}}\right)^{\frac{1}{k}} \omega \\ &= \omega < 1 \end{aligned}$$

which shows that $\{x_k\}$ converges to x^* at least R-linearly. □

5 Numerical results

In this section we will discuss the implementation of the new algorithm. The technique of choosing parameters is reasonable and effective for solving practical problems both in theoretical and numerical aspect.

5.1 Parametric estimation

In the modified Armijo line search, the parameter L_k needs to be estimated. As we know, L_k should approximate Lipschitz constant M' of the gradient g of objective function f. If M' is given, we should certainly take $L_k = M'$. However, M' is not known a prior in many situations.

First of all, we let $\delta_k = x_{k+1} - x_k,\ y_k = g_{k+1} - g_k$, and estimate

$$L_k = \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|}, \tag{68}$$

or

$$L_k = \max\left\{\frac{\|y_{k-i}\|}{\|\delta_{k-i}\|} \mid i = 1, 2, ..., M\right\} \tag{69}$$

whenever $k \geq M + 1$, and M is a positive integer.

Next, BB method [2, 4, 25, 26] also motivates us to find a way of estimating M'. Solving the minimization

$$\min_{L \in R^1} \|L\delta_{k-1} - y_{k-1}\|,$$

we can obtain

$$L_k = \frac{\left|\delta_{k-1}^T y_{k-1}\right|}{\|\delta_{k-1}\|^2}. \tag{70}$$

Obviously, if $K \geq M + 1$, we can also take

$$L_k = \max\left\{\frac{|\delta_{k-i}^T y_{k-i}|}{\|\delta_{k-i}\|^2} \mid i = 1, 2, ..., M\right\}. \tag{71}$$

On the other hand, we can take

$$L_k = \frac{\|y_{k-1}\|^2}{\left|\delta_{k-1}^T y_{k-1}\right|}, \tag{72}$$

or

$$L_k = \max\left\{\frac{\|y_{k-i}\|^2}{\left|\delta_{k-i}^T y_{k-i}\right|} \mid i = 1, 2, ..., M\right\}, \tag{73}$$

whenever $k \geq M + 1$.

There are many other techniques of estimating the Lischitz constant M' [35]. We will use (68)–(73) to estimate the M' and the corresponding algorithms are denoted by Algorithms (68)–(73) respectively.

5.2 Numerical results

In what follows, we will discuss the numerical performance of the new line search method. The test problems are chosen from [18] and the implementable algorithm is stated as follows.

Algorithm (A)

Step 0. Given some parameters $\sigma \in \left(0, \frac{1}{2}\right)$, $\rho \in (0, 1)$, $\mu \in [0, 2)$, and $L_0 = 1$; let $x_1 \in R^n$ and set $k := 0$;
Step 1. If $\|g_k\| = 0$ then stop, else go to step 3;
Step 3. Choose d_k satisfies angle property (5), for example, taking $d_k = -g_k$;
Step 4. $x_{k+1} = x_k + \alpha_k d_k$, where α_k is defined by the modified nonmonotone Armijo line search;
Step 5. $\delta_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$, and L_{k+1} is determined by one of (68)–(73);
Step 6. $k := k + 1$, go to step 1.

In the above algorithm, we set $\sigma = 10^{-4}$, $\rho = 0.87$, $\mu = 1$, $M = 2$ and compare with original Armijo line search method with the same parameters. We will find that the step-size in the new line search method is easier to find than in the original one. In other words, the new line search method needs less evaluations of gradient and objective function at each iteration. We tested the new line search methods and original Armijo line search method with double precision in a portable computer. The codes were written in the visual C++ language. Our test problems and the initial points used are drawn from [18]. For each problem, the limiting number of function evaluations is set to 1,000, and the stopping condition is

$$\|g_k\| \leq 10^{-6}. \tag{74}$$

We report our numerical results in Table 1, where 'Armijo', 'N-Armijo', 'New(68)', 'New(70)', and 'New(72)' stand for the Armijo line search method, nonmonotone Armijo line search method, new line search methods with L_k taken by (68), (70), and (72), respectively. The symbols n, IN, N_f mean the dimension of the problem, the number of iterations, and the number of function evaluations.

The unconstrained optimization problems are numbered in the same way as [18]. For example, P5 means Problem 5 in [18]. We will set $d_k = -g_k$ at each

Table 1 IN and N_F for reaching the same precision, $M = 2$

Problem	n	Armijo	N-Armijo	New(68)	New(70)	New(72)
P5	2	8/12	8/9	6/9	7/11	7/8
P11	3	22/32	19/22	16/19	14/17	17/23
P14	4	36/50	33/45	28/34	26/33	30/42
P16	4	14/72	12/56	16/63	12/58	11/56
P20	9	12/17	13/13	12/13	12/13	11/11
P21	16	18/21	19/19	16/23	12/14	11/15
P21	100	21/30	18/27	17/22	16/25	15/20
P23	8	30/42	30/38	28/34	26/36	26/38
P23	100	36/58	32/49	30/34	28/32	30/52
P23	200	55/87	59/61	43/67	48/72	42/61
P24	20	52/67	49/64	45/59	38/38	47/52
P25	50	11/121	14/34	11/32	16/78	9/83
P26	50	14/30	14/17	14/19	12/16	15/18
P35	20	17/28	15/24	14/21	16/19	12/18
CPU	–	23s	19s	17s	14s	16s

iteration. In this case, if $\alpha_k = s_k$ at each iteration, then 'New(70)' and 'New(72)' will reduce to BB methods [2, 4, 25, 26] because of

$$\alpha_k = s_k = -\frac{g_k^T d_k}{L_k \|d_k\|^2} = \frac{1}{L_k} = \frac{\|\delta_{k-1}\|^2}{\left|\delta_{k-1}^T y_{k-1}\right|}$$

corresponding to (70), or

$$\alpha_k = \frac{\left|\delta_{k-1}^T y_{k-1}\right|}{\|y_{k-1}\|^2}$$

corresponding to (72).

As we know, BB method is an efficient method without line search [24]. In fact, a powerful inexact line search should reduce the function evaluation number at each iteration. The new line search method proposed in the paper needs less evaluation number of function value than other line search methods. For example, Armijo line search method, nonmonotone Armijo line search method, etc. Moreover, how to estimate the parameter L_k is still a valuable problem. We should seek other ways to estimate L_k so as to cut down the number of function evaluations at each iteration. We use "CPU" to denote the average computational time [8] for solving each corresponding problems.

From Table 1, we can see that, for some problems, the three new algorithms need less number of function evaluations than original Armijo line search methods. However, for some problems, Armijo line search method performs as well as the new line search methods. Therefore, our numerical results show that the new line search methods are superior to original Armijo line search methods in many situations. In particular, the new method needs less function evaluations than original Armijo line search methods, i.e, in many cases, α_k always takes s_k in the new line search. Moreover, the estimation of L_k (and essentially s_k) is very important in the new algorithm. In numerical experiment, we take $d_k = -g_k$, which is a special case of descent directions. We may take other descent directions as d_k at each step.

If we take $\mu = 1$, $\rho = 0.001$ and $U = 1000$ then the convergence condition (33) reduces to

$$\frac{L}{1000} < L_k \leq 1000L. \tag{75}$$

Thus the algorithm will converge for rough estimation of L_k. All the above mentioned problems are very small scale problems. The advantage of the new modification of nonmonotone Armijo line search seems not obvious. We should implement it for large scale problems.

For large scale problems [8–10], we implement the modified nonmonone Armijo line search in the same software environment. In Table 2, '– –' denotes the number of iteration $IN \geq 10{,}000$. The numerical results are listed in Table 2.

From Table 2, we can see that, the previous Armjio-type line searches, such as 'Armijo' and 'N-Armijo', need more iterations and more functional evaluations than the new modifications New(68), New(70) and New(72) when reaching the same precision. We found that the previous Armijo-type line search can result in zigzag phenomena in solving some large scale problems. The corresponding method cannot stop in the local area of the minimizer because of the constant choice of initial possible acceptable step size at each iteration. The shows that the new modification of nonmonotone Armijo line search can avoid zigzag phenomena due to the good estimation of initial possible acceptable step size s_k.

Table 2 IN and N_f for reaching the same precision, $M = 2$

Problem	n	Armijo	N-Armijo	New(68)	New(70)	New(72)
TORSION1	15,625	335/8,334	249/8,132	146/6,739	145/6,921	148/7,120
CLPLATEA	4,970	265/5,289	213/2,322	132/1,891	141/1,756	145/1,542
JNLBRNGA	15,625	382/7,291	338/6,245	321/2,360	287/3,533	315/2,163
SVANBERG	5,041	512/8,127	389/4,287	367/3,986	326/3,812	382/3,480
BRATU2D	5,184	142/4,517	156/1,563	134/1,473	152/1,453	123/1,311
BRATU3D	4,913	148/2,331	149/1,329	134/2,543	122/1,984	131/1,345
TRIGGER	5,002	231/3,340	128/2,347	137/3,422	134/2,345	134/2,230
BDVALUE	5,002	310/4,222	323/3,438	228/3,234	226/3,236	276/1,768
VAREIGVL	5,000	–/–	324/1,249	310/1,214	218/1,232	183/1,352
MSQRTA	10,000	–/–	–/–	413/4,523	438/6,772	242/4,551
NLMSURF	15,625	–/–	–/–	965/8,867	956/9,823	847/8,252
OBSTCLBU	15,625	–/–	814/7,534	711/7,323	618/6,871	459/3,487
PRODPL1	500	32/381	28/127	26/106	26/106	25/98
ODNAMUR	11,276	–/–	515/7,832	314/3,721	216/2,917	167/1,927
CPU	–	–	–	196s	143s	158s

Table 3 IN and N_F for reaching the same precision, $M = 15$

Problem	n	Armijo	N-Armijo	New(69)	New(71)	New(73)
P5	2	8/12	9/13	10/14	9/18	12/15
P11	3	22/32	13/20	17/17	14/16	15/21
P14	4	36/50	32/41	26/35	26/32	31/43
P16	4	14/72	14/17	16/24	12/34	11/43
P20	9	12/17	14/15	16/17	12/15	12/18
P21	16	18/21	19/19	16/23	12/14	11/15
P21	100	21/30	23/23	17/17	18/22	15/28
P23	8	30/42	34/34	31/39	28/32	25/28
P23	100	36/58	32/32	28/32	28/32	30/32
P23	200	55/87	52/52	43/43	45/52	42/54
P24	20	52/67	53/65	48/65	42/37	48/55
P25	50	11/121	14/14	12/12	16/18	9/23
P26	50	14/30	15/15	14/17	12/14	15/15
P35	20	17/28	15/15	14/16	14/17	16/16
CPU	–	23s	19s	21s	18s	18s

As to the nonmonotone strategy, taking $M = 2$ seems to be too small. We should do more numerical experiments to test the performance of the new nonmonotone line search when M takes big numbers. We use (69), (71) and (73) to replace (68), (70) and (72) in the algorithm when $k \geq M + 1$.

From Table 3, we found that we should take small M for small scale problems. When $M = 15$ and $n \leq 20$, the performance of the nonmonotone line search seems to be bad while the performance seems to be better when n is large. We are not sure how to take M can make the nonmonotone line search have a good performance in practical computation. We try to find some relationship between M and n but not yet right now. It may depend on the specific objective function. We used $M = 5$ and $M = 10$ to test the nonmontone

Table 4 IN and N_f for reaching the same precision, $M = 15$

Problem	n	Armijo	N-Armijo	New(69)	New(71)	New(73)
TORSION1	15,625	335/8,334	236/2,632	132/4,265	134/5,128	128/5,641
CLPLATEA	4,970	265/5,289	187/2,014	125/1,569	136/1,527	118/1,241
JNLBRNGA	15,625	382/7,291	321/5,297	321/2,187	279/2,563	283/1,982
SVANBERG	5,041	512/8,127	342/3,257	325/3,216	309/3,173	325/3,241
BRATU2D	5,184	142/4,517	132/1,269	133/1,424	128/1,243	121/1,289
BRATU3D	4,913	148/2,331	138/1,247	128/2,172	119/1,573	116/1,426
TRIGGER	5,002	231/3,340	136/2,529	132/3,326	136/2,523	129/2,374
BDVALUE	5,002	310/4,222	283/3,542	216/2,860	208/3,029	226/1,528
VAREIGVL	5,000	–/–	317/1,183	284/1,073	238/1,541	188/1,456
MSQRTA	10,000	–/–	424/5,253	411/4,217	405/6,263	352/4,825
NLMSURF	15,625	–/–	934/8,925	923/8,377	921/9,345	825/8,153
OBSTCLBU	15,625	–/–	834/7,734	731/7,453	623/6,452	426/3,723
PRODPL1	500	32/381	26/147	28/121	26/98	22/95
ODNAMUR	11,276	–/–	525/7,421	306/3,021	208/2,254	125/1,906
CPU	–	–	183s	172s	147s	153s

line search and found that the situation of $M = 5$ is very similar to that of $M = 2$, while the situation of $M = 10$ is very similar to that of $M = 15$. See Table 4 to find the performance of the nonmonotone line search when $M = 15$.

In Tables 1, 2, 3 and 4, Algorithm New(70) and New(71) seem to be the best in all mentioned algorithms based on the average computational time. Actually, we want to find the smallest Lipschitz constant of the gradient of objective function. In fact,

$$\frac{|y_{k-1}^T \delta_{k-1}|}{\|\delta_{k-1}\|^2} \leq \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|} \leq \frac{\|y_{k-1}\|^2}{|y_{k-1}^T \delta_{k-1}|}.$$

Therefore, (70) and (71) are the best choices of Lipschitz constant estimation and thus result in the best estimations of step sizes at the k-th iteration.

6 Conclusion

A new modification of matrix-free nonmonotone Armijo line search was developed and the related descent method was investigated. It was proved that the new modification of nonmonotone Armijo line search method has global convergence and linear convergence rate under some mild conditions. We can choose a larger step-size in each line search procedure and sometimes avoid the iterates running into the narrow curved valley of objective functions. The new modification of nonmontone Armijo line search can avoid zigzag phenomena in some cases. The new line search approach can make us design new line search methods in some wider sense. Especially, the new modified line search method can reduce to BB method [2, 4, 25, 26] in some special cases. As we can see, Lipschitz constant M' of gradient $g(x)$ of the objective function $f(x)$ needs to be estimated at each step. we have discussed some techniques for choosing L_k. In numerical experiment, we take $d_k = -g_k$ at each step. Indeed, we can take other descent directions as d_k, such as conjugate gradient direction, limited quasi-Newton directions, etc.

For the future research, we can establish global convergence of other line search methods with the modified nonmonotone Goldstein or the modified nonmonotone Wolfe line searches. Of course, we expect to decrease the number of gradients and functional evaluations for some line searches. Since the new line search needs to estimate L_k, we can find other ways to estimate L_k and choose diverse parameters such as σ, μ, and ρ, so as to find available parameters in solving special unconstrained optimization problems. Moreover, we can investigate the L-BFGS and CG method with the modified nonmonotone Armijo line search.

Acknowledgements The authors are very grateful to the anonymous referees and the editor for their careful reading and valuable comments and suggestions that greatly improved the paper.

References

1. Armijo, L.: Minimization of functions having Lipschitz continuous first partial derivatives. Pac. J. Math. **16**, 1–3 (1966)
2. Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA J. Numer. Anal. **8**, 141–148 (1988)
3. Cohen, A.I.: Stepsize analysis for descent methods. J. Optim. Theory Appl. **33**(2), 187–205 (1981)
4. Dai, Y.H., Liao, L.Z.: R-linear convergence of the Barzilai and Borwein gradient method. IMA J. Numer. Anal. **22**, 1–10 (2002)
5. Dai, Y.H.: On the nonmonotone line search. J. Optim. Theory Appl. **112**(2), 315–330 (2002)
6. Deng, N., Xiao, Y., Zhou, F.: Nonmonotone trust region algorithms. J. Optim. Theory Appl. **76**, 259–285 (1993)
7. Dennis J.E. Jr., Schnable, R.B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs (1983)
8. Dolan, E., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**, 201–213 (2002)
9. Gould, N.I.M., Orban, D., Toint, Ph.L.: GALAHAD, a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization. ACM Trans. Math. Softw. **29**, 353–372 (2003)
10. Bongartz, I., Conn, A.R., Gould, N., Toint, Ph.L.: CUTE: constrained and unconstrained testing environment, ACM Trans. Math. Softw. **21**, 123–160 (1995)
11. Grippo, L., Lampariello, F., Lucidi, S.: A class of nonmonotone stability methods in unconstrained optimization. Numer. Math. **62**, 779–805 (1991)
12. Grippo, L., Lampariello, F., Lucidi, S.: A truncated Newton method with nonmonotone line search for unconstrained optimization. J. Optim. Theory Appl. **60**, 401–419 (1989)
13. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newton's method. SIAM J. Numer. Anal. **23**, 707–716 (1986)
14. Goldstein, A.A.: On steepest descent. SIAM J. Control **3**, 147–151 (1965)
15. Ke, X., Han, J.: A nonmonotone trust region algorithm for equality constrained optimization. Sci. China **38A**, 683–695 (1995)
16. Ke, X., Liu, G., Xu, D.: A nonmonotone trust region algorithm for unconstrained optimization. Chin. Sci. Bull. **41**, 197–201 (1996)
17. Liu, G., Han, J., Sun, D.: Global convergence of BFGS algorithm with nonmonotone line search. Optimization **34**, 147–159 (1995)
18. Moré, J.J., Garbow, B.S., Hillstron, K.E.: Testing unconstrained optimization software. ACM Trans. Math. Softw. **7**, 17–41 (1981)
19. Nocedal, J., Wright, J.S.: Numerical Optimization. Springer, New York (1999)
20. Nocedal, J.: Theory of algorithms for unconstrained optimization. Acta Numer. **1**, 199–242 (1992)
21. Powell, M.J.D.: Direct search algorithms for optimization calculations. Acta Numer. **7**, 287–336 (1998)
22. Polak, E.: Optimization: Algorithms and Consistent Approximations. Springer, New York (1997)
23. Panier, E., Tits, A.: Avoiding the Maratos effect by means of a nonmonotone line search, I: general constrained problems. SIAM J. Numer. Anal. **28**, 1183–1195 (1991)
24. Raydan, M., Svaiter, B.E.: Relaxed steepest descent and Cauchy-Barzilai-Borwein method. Comput. Optim. Appl. **21**, 155–167 (2002)
25. Raydan, M.: The Barzilai and Borwein method for the large scale unconstrained minimization problem. SIAM J. Optim. **7**, 26–33 (1997)
26. Raydan, M.: On the Barzilai Borwein gradient choice of steplength for the gradient method. IMA J. Numer. Anal. **13**, 321–326 (1993)
27. Sun, W.Y., Han, J.Y., Sun, J.: Global convergence of nonmonotone descent methods for unconstrained optimization problems. J. Comput. Appl. Math. **146**, 89–98 (2002)
28. Shi, Z.J., Shen, J.: Convergence of nonmonotone line search method. J. Comput. Appl. Math. **193**, 397–412 (2006)
29. Shi, Z.J., Shen, J.: New inexact line search method for unconstrained optimization. J. Optim. Theory Appl. **127**, 425–446 (2005)

30. Shi, Z.J., Shen, J.: Convergence of descent method without line search. Appl. Math. Comput. **167**, 94–107 (2005)
31. Shi, Z.J.: Convergence of line search methods for unconstrained optimiztaion. Appl. Math. Comput. **157**, 393–405 (2004)
32. Toint, P.L.: An assessment of nonmonotone line search techniques for unconstrained optimization. SIAM J. Sci. Comput. **17**, 725–739 (1996)
33. Toint, P.L.: A nonmonotone trust-region algorithm for nonlinear optimization subject to convex constraints. Math. Program. **77**, 69–94 (1997)
34. Todd, M.J.: On convergence properties of algorithms for unconstrained minimization. IMA J. Numer. Anal. **9**, 435–441 (1989)
35. Vrahatis, M.N., Androulakis, G.S., Lambrinos, J.N., Magoulas, G.D.: A class of gradient unconstrained minimization algorithms with adaptive stepsize. J. Comput. Appl. Math. **114**, 367–386 (2000)
36. Vrahatis, M.N., Androulakis, G.S., Manoussakis, G.E.: A new unconstrained optimization method for imprecise function and gradient values. J. Math. Anal. Appl. **197**, 586–607 (1996)
37. Wei, Z., Qi, L., Jiang, H.: Some convergence properties of descent methods. J. Optim. Theory Appl. **95**(1), 177–188 (1997)
38. Wolfe, P.: Convergence condition for ascent methods II: some corrections. SIAM Rev. **13**, 185–188 (1971)
39. Wolfe, P.: Convergence condition for ascent methods. SIAM Rev. **11**, 226–235 (1969)
40. Xiao, Y., Zhou, F.J.: Nonmonotone trust region methods with curvilinear path in unconstrained minimization. Computing **48**, 303–317 (1992)
41. Zhang, J.L., Zhang, X.S.: A nonmonotone adaptive trust region method and its convergence. Comput. Math. Appl. **45**, 1469–1477 (2003)