Optimization: A Journal of Mathematical Programming and Operations Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gopt20

A class of nonmonotone Armijo-type line search method for unconstrained optimization

Masoud Ahookhosh [a] , Keyvan Amini [a] & Somayeh Bahrami [a]
[a] Department of Mathematics , Faculty of Science, Razi University , Kermanshah , Iran
Published online: 06 Jan 2012.

Optimization
Vol. 61, No. 4, April 2012, 387–404

A class of nonmonotone Armijo-type line search method for unconstrained optimization

Masoud Ahookhosh, Keyvan Amini* and Somayeh Bahrami

Department of Mathematics, Faculty of Science, Razi University, Kermanshah, Iran

(*Received 5 October 2010; final version received 8 November 2011*)

In this article, we propose a new line search method for solving unconstrained optimization problems in that we combine a nonmonotone strategy into a modified Armijo rule and design a new algorithm that possibly chooses a larger steplength. This can decrease the number of iterations and function evaluations and can improve the efficiency of the algorithm. The global convergence and convergence rate are analysed under some suitable conditions. Preliminary numerical experiments establish that the new approach is robust and efficient for unconstrained optimization problems.

Keywords: unconstrained optimization; line search method; nonmonotone strategy; global convergence

AMS Subject Classifications: 90C30; 65k05; 65k10

1. Introduction

Consider the following unconstrained optimization problem:

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

where f: $\mathbf{R}^n \to \mathbf{R}$ is continuously differentiable function. There are various methods to solve the problem (1) most of which are iterative methods. In general, iterative algorithms for nonlinear optimization proceed as follows: given a point x_k, find a descent direction d_k such that $d_k^T \nabla f(x_k) < 0$, a suitable steplength α_k and construct the new point as follows:

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

The line search is a subproblem to find α_k in iterative process (2) [14,15]. In this process to find a steplength α_k, we must solve the following one-dimensional minimization problem

$$\min_{\alpha \geq 0} \phi(\alpha) = f(x_k + \alpha d_k). \tag{3}$$

*Corresponding author. Email: kamini@razi.ac.ir

ISSN 0233–1934 print/ISSN 1029–4945 online

http://dx.doi.org/10.1080/02331934.2011.641126
http://www.tandfonline.com

Very often in practice, we cannot afford the luxury of performing an exact line search because of the expense of excessive function evaluations, even if we terminate algorithm with some small accuracy tolerance $\epsilon > 0$. Therefore, in the most cases, one tries to avoid solving problem (3) exactly. On the other hand, if we forfeit accuracy, we might impair the convergence of the overall algorithm that iteratively employs such a line search.

More practical strategies perform an inexact line search to identify a steplength that achieves adequate reductions in f at minimal cost. These strategies choose the steplength α_k guaranteeing a sufficient reduction in function values while this might induce the overall algorithm to converge. Some conditions proposed for acceptance of steplength α_k, namely the Armijo, Wolfe and Goldstein conditions [3,14]. Among these rules, the Armijo rule is the most popular condition stating as follows:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma \alpha_k \nabla f(x_k)^T d_k, \tag{4}$$

where $\sigma \in (0, 1/2)$ and α_k is the largest α in $\{s, \rho s, \rho^2 s, \ldots\}$ for $s > 0$ and $\rho \in (0, 1)$. At a glance to (4), we can conclude $f(x_{k+1}) < f(x_k)$, so this schema is called a monotone line search. For the sake of simplicity, we abbreviate $f(x_k)$, $\nabla f(x_k)$ and $\nabla^2 f(x_k)$ with f_k, g_k and G_k, respectively.

In [8], Grippo et al. presented the new line search technique for unconstrained optimization that permitted f_{k+1} to be greater than f_k, so they called their method as the nonmonotone line search. Numerical results show that their nonmonotone technique decreases both the number of line searches and function evaluations. Motivated by this subject, many researchers worked on nonmonotone techniques. They pointed out that nonmonotone strategies can improve the likelihood of finding a global optimum [1,4,6,9,10,16,18,19]. Some numerical results have indicated that nonmonotone strategies can raise the speed convergence of algorithms, especially in the presence of narrow curved valley [4,18]. Grippo et al. defined their rule as follows

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq m(k)} \{f_{k-j}\} + \sigma \alpha_k g_k^T d_k, \tag{5}$$

where $N \geq 0$ is an integer constant, $m(0) = 0$ and for all $k \geq 1$ we have $0 \leq m(k) \leq \min\{m(k-1) + 1, N\}$. Although this nonmonotone technique has many advantages, it contains some drawbacks [4,17,20]. We can list some of these drawbacks as follows:

- Although an iterative method is generating R-linearly convergent iterations for a strongly convex function, the iterates may not satisfy the condition (5) for k sufficiently large, for any fixed bound N on the memory (see introduced example by Dai [4]).
- Initial steplength is a constant and cannot be adjusted according to the characteristics of the objective function in the current iteration [17].
- A good function value generated in any iteration is essentially discard due to the max in (5) [20].
- In some cases, the numerical performances are very dependent on the choice of N [20].

Different approaches have been proposed to overcome some of these drawbacks [11,17,20]. For example, Zhang and Hager in [20] proposed a nonmonotone line search algorithm based on a weighted average of successive function values. In detail, they relaxed Armijo condition as follows:

$$f(x_k + \alpha_k d_k) \leq C_k + \sigma \alpha_k g_k^T d_k, \tag{6}$$

where C_k is defined as follows

$$C_k = \begin{cases} f(x_k), & \text{if } k = 0, \\ (\eta_{k-1} Q_{k-1} C_{k-1} + f(x_k))/Q_k, & \text{if } k \geq 1, \end{cases} \tag{7}$$

with

$$Q_k = \begin{cases} 1, & \text{if } k = 0, \\ \eta_{k-1} Q_{k-1} + 1, & \text{if } k \geq 1, \end{cases}$$

where $\eta_{k-1} \in [\eta_{\min}, \eta_{\max}]$ for $\eta_{\min} \in [0, 1)$ and $\eta_{\max} \in [\eta_{\min}, 1)$. Numerical results showed the efficiency of their nonmonotone line search method [20]. It is clear that this method tries to overcome 1st, 3rd and 4th drawbacks.

In this article, we propose a new line search algorithm for solving unconstrained optimization problems. In the algorithm, we combine a nonmonotone strategy into a modified Armijo rule and design a new algorithm that possibly chooses a larger steplength in each step. It may decrease the number of iterations and function evaluations and can improve the efficiency of the new approach. The global convergence and convergence rate are analysed under some suitable conditions. Preliminary numerical results show that the new approach is robust and efficient for solving unconstrained optimization problems.

The rest of this article is organized as follows: in Section 2, we describe a new nonmonotone line search algorithm and present its properties. The global convergence of the proposed algorithm is proved in Section 3. Section 4 presents results on convergence rates of the proposed algorithm. Numerical experiments are given in Section 5. Finally, we draw some conclusions in Section 6.

2. New nonmonotone line search algorithm

In this section, we present a new nonmonotone line search algorithm with automatically adjusted steplength and give some properties about it.

It is well-known that if objective function $f(x)$ is a convex quadratic function, then the exact line search steplength can be computed by

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T G_k d_k}.$$

Thus, for a general function, it could possibly be a good initial steplength for Armijo-type conditions. Due to this reason and to exploit the second-order characteristic of objective function at current step, we choose the automatically adjustable initial steplength s_k as follows

$$s_k = -\frac{g_k^T d_k}{d_k^T B_k d_k}, \tag{8}$$

where B_k is a symmetric approximation of Hessian matrix G_k. Like [17], if $d_k^T B_k d_k \leq 0$, then we set $B_k = B_k + iI$ where i is the smallest nonnegative integer such that

$$i > -(d_k^T B_k d_k)/\|d_k\|^2,$$

where $\|\cdot\|$ represents Euclidean norm.

We also choose the steplength α_k to be the largest α in $\{s_k, \rho s_k, \rho^2 s_k, \ldots\}$ satisfying the condition

$$f(x_k + \alpha_k d_k) \leq D_k + \sigma\alpha_k\left[g_k^T d_k + \gamma\|g_k\|^2\right], \tag{9}$$

where $\gamma > 0$ is a constant, and similar to [11], we define

$$D_k = \begin{cases} f(x_0), & \text{if } k = 0, \\ f(x_k) + \eta_{k-1}(D_{k-1} - f(x_k)), & \text{if } k \geq 1, \end{cases} \tag{10}$$

where $\eta_{k-1} \in [\eta_{\min}, \eta_{\max}]$ for $\eta_{\min} \in [0, 1)$ and $\eta_{\max} \in [\eta_{\min}, 1)$.

In order to analyse the convergence properties of the proposed method, throughout this article, we consider the following assumptions:

(i) The level set $L(x_0) = \{x \in \mathbf{R}^n \mid f(x) \leq f(x_0),\ x_0 \in \mathbf{R}^n\}$ is bounded.
(ii) There exist constants $0 < m \leq M$ such that for all k

$$m\|d_k\|^2 \leq d_k^T B_k d_k \leq M\|d_k\|^2. \tag{11}$$

Furthermore, due to guaranteeing the global convergence of the iterative schema (2), we need the following assumption concerning the search direction that has been extensively used in optimization literature (e.g. [4,8,20]).

Direction assumption: There are positive constants c_1 and c_2 such that

$$g_k^T d_k \leq -c_1\|g_k\|^2 \tag{12}$$

and

$$\|d_k\| \leq c_2\|g_k\|. \tag{13}$$

Note that this assumption is not a strong condition because there are many directions satisfying these conditions, for example the steepest descent direction fulfills (12) and (13) with $c_1 = c_2 = 1$.

Now, we can outline our new algorithm as follows:

Algorithm 2.1 (A new nonmonotone algorithm)

(i) **Initialization.** An initial point $x_0 \in \mathbf{R}^n$ and a symmetric positive matrix $B_0 \in \mathbf{R}^{n\times n}$ are given. The constants $0 < \rho < 1$, $0 < \sigma < \frac{1}{2}$, $\gamma > 0$, $\epsilon > 0$, $0 \leq \eta_{\min} \leq \eta_0 \leq \eta_{\max} < 1$ are also given. Compute $D_0 = f(x_0)$ and set $k = 0$.
(ii) **Stopping criterion.** Compute g_k. If $\|g_k\| \leq \epsilon$, then stop.
(iii) **Direction and steplength calculation.** Choose the descent direction d_k satisfying in (12) and (13). Then find the steplength $\alpha_k \in \{s_k, \rho s_k, \rho^2 s_k, \ldots\}$ that is the largest among these number satisfying the condition (9). Set $x_{k+1} = x_k + \alpha_k d_k$.

(iv) **Parameters update.** Compute D_{k+1} by (10), update B_{k+1} by a quasi-Newton formula and choose $\eta_{\min} \le \eta_k \le \eta_{\max}$. Set $k = k+1$ and go to Step (ii).

Remark 1 Assume that $f(x)$ is a twice continuously differentiable function and (H1) holds, then $\|\nabla^2 f(x)\|$ is uniformly continuous and bounded on the open bounded convex set Ω that contains $L(x_0)$. Hence, there exists a constant L such that $\|\nabla^2 f(x)\| \le L$. Using the mean value theorem, one can conclude

$$\| g(x) - g(y)\| \le L\|x - y\| \quad \forall x, y \in L(x_0).$$

It follows from this remark that there exists a constant $\gamma_1 > 0$ such that

$$\| g(x)\| \le \gamma_1 \quad \forall x \in L(x_0). \tag{14}$$

LEMMA 2.1 *Suppose that the sequence* $\{x_k\}$ *is generated by Algorithm* 2.1, *then for* $0 < \gamma < c_1$, *we have*

$$f_{k+1} \le D_{k+1} \le D_k \quad \forall k \in \mathbf{N} \cup \{0\}. \tag{15}$$

Proof Let iteration k be a successive iteration. From the definition α_k, (9) and (12) we have

$$f_{k+1} - D_k \le \sigma\alpha_k\big[g_k^T d_k + \gamma\| g_k\|^2\big] \le \sigma\alpha_k\big[g_k^T d_k + c_1\| g_k\|^2\big] \le 0. \tag{16}$$

Now, (10) together with (16) imply that

$$\begin{aligned} D_{k+1} - D_k &= f_{k+1} + \eta_k(D_k - f_{k+1}) - D_k \\ &= (1 - \eta_k)(f_{k+1} - D_k) \le 0. \end{aligned} \tag{17}$$

On the other hand, if $\eta_k \neq 0$ from (10) and (16) we have

$$\begin{aligned} D_{k+1} - f_{k+1} &= f_{k+1} + \eta_k(D_k - f_{k+1}) - f_{k+1} \\ &= \eta_k(D_k - f_{k+1}) \ge 0. \end{aligned} \tag{18}$$

Thus, (17) and (18) indicate that

$$f_{k+1} \le D_{k+1} \le D_k \quad \forall k \in \mathbf{N} \cup \{0\}.$$

Also, if $\eta_k = 0$, we have $D_{k+1} = f_{k+1}$. Therefore, (15) holds. ■

COROLLARY 2.2 *Suppose that assumption* (H1) *holds, then the sequence* $\{D_k\}$ *is convergent*.

Proof First, we show that $x_k \in L(x_0)$ for all $k \in \mathbf{N} \cup \{0\}$. Obviously, from (15) we have

$$f_k \le D_k \le D_{k-1} \le \cdots \le D_0 = f_0 \quad \forall k \in \mathbf{N} \cup \{0\}.$$

Thus, we obtain $x_k \in L(x_0)$ for all $k \in \mathbf{N} \cup \{0\}$. Now, this fact along with (15) and (H1) imply that

$$\exists \lambda \text{ s.t. } \forall n \in \mathbf{N} \cup \{0\}\colon \lambda \le f_{k+n} \le D_{k+n} \le \cdots \le D_{k+1} \le D_k,$$

i.e., the sequence $\{D_k\}$ has a lower bound. This implies that the sequence $\{D_k\}$ is a convergent. ∎

At a glance to the standard Armijo rule and the new Armijo-type line search, firstly, we can see that the term $\gamma\|g_k\|^2$ is added to right-hand side of the Armijo rule. Secondly, we substitute D_k instead of f_k which is possibly greater than f_k. Therefore, we can see the right-hand side of the approach is greater than the right-hand side of the standard Armijo rule, so it is possibly permitted to the algorithm to gain a larger steplength. These changes may reduce the number of iterations and function evaluations due to attaining the same optimum. Details are represented in the following lemma.

LEMMA 2.3 *Suppose that the sequence $\{x_k\}$ is generated by Algorithm* 2.1. *If $\tilde{\alpha}$ and α are steplengths which satisfy in the standard Armijo rule and Algorithm* 2.1, *respectively, then $\tilde{\alpha} \le \alpha$, and the new nonmonotone line search is well-defined.*

Proof If $\tilde{\alpha}$ and α are the steplengths which satisfy in the standard Armijo rule and the new Armijo-type line search method, respectively, then we have

$$f(x_k + \tilde{\alpha}_k d_k) - D_k \le f(x_k + \tilde{\alpha}_k d_k) - f_k \le \sigma\tilde{\alpha}_k g_k^T d_k \le \sigma\tilde{\alpha}_k\left[g_k^T d_k + \gamma\|g_k\|^2\right].$$

This implies that $\tilde{\alpha} \le \alpha$. From (13) and (14), we have

$$\|d_k\| < \infty \quad \forall k \in \mathbf{N} \cup \{0\}. \tag{19}$$

Now, by Taylor's theorem, (9) and (15) we obtain

$$\begin{aligned}
&\lim_{\alpha\to 0^+} \frac{D_k - f(x_k + \alpha d_k) + \sigma\alpha\left[g_k^T d_k + \gamma\|g_k\|^2\right]}{\alpha}\\
&\quad\ge \lim_{\alpha\to 0^+} \frac{f_k - f(x_k + \alpha d_k) + \sigma\alpha\left[g_k^T d_k + \gamma\|g_k\|^2\right]}{\alpha}\\
&\quad= \lim_{\alpha\to 0^+} \frac{f_k - (f_k + \alpha g_k^T d_k + o(\alpha\|d_k\|)) + \sigma\alpha\left[g_k^T d_k + \gamma\|g_k\|^2\right]}{\alpha}\\
&\quad= \left[-(1-\sigma)\, g_k^T d_k + \gamma\sigma\|g_k\|^2\right] > 0.
\end{aligned}$$

So, there exists a $\hat{\alpha}_k > 0$ such that

$$f(x_k + \alpha d_k) \le D_k + \sigma\alpha\left[g_k^T d_k + \gamma\|g_k\|^2\right] \quad \forall \alpha \in [0, \hat{\alpha}_k].$$

Therefore, the new nonmonotone line search is well-defined. ∎

3. Global convergence

In this section, we discuss some convergence properties of the proposed Armijo-type line search algorithm and prove the global convergence to first-order critical points under some suitable conditions.

LEMMA 3.1 *Suppose that* (H1) *holds and the sequence* $\{x_k\}$ *is generated by Algorithm* 2.1, *then we have*

$$\lim_{k\to\infty} D_k = \lim_{k\to\infty} f(x_k). \tag{20}$$

Proof At first, notice that from Steps (i) and (iv) of Algorithm 2.1, we can see that $\eta_{\max} \in [0, 1)$ and $\eta_k \in [\eta_{\min}, \eta_{\max}]$ for all k. From (17), we have

$$D_{k+1} - D_k = (1 - \eta_k)(f_{k+1} - D_k) \le 0$$

and

$$1 - \eta_k \ge 1 - \eta_{\max} > 0.$$

Using Corollary 2.2, as $k \to \infty$, implies that

$$\lim_{k\to\infty}(f_{k+1} - D_k) = \lim_{k\to\infty}(D_{k+1} - D_k) = 0.$$

This completes the proof. ■

LEMMA 3.2 *Assume that* (H2) *holds*, $0 < \gamma < c_1$, *and the sequence* $\{x_k\}$ *is generated by Algorithm* 2.1. *Then there exists a positive constant* δ *such that*

$$D_k - f_{k+1} \ge \delta \| g_k\|^2 \quad \forall k \in \mathbf{N}. \tag{21}$$

Proof Set $K_1 = \{k \in K| \ \alpha_k = s_k\}$ and $K_2 = \{k \in K| \ \alpha_k < s_k\}$. We must prove that (21) holds for both K_1 and K_2.

If $k \in K_1$, then (11) and (12) together with (8), (9) and (13) imply that

$$\begin{aligned} D_k - f_{k+1} &\ge -\sigma\alpha_k\left[g_k^T d_k + \gamma\| g_k\|^2\right] \\ &= \sigma \frac{-g_k^T d_k}{d_k^T B_k d_k}\left[-g_k^T d_k - \gamma\| g_k\|^2\right] \\ &\ge \frac{c_1\sigma}{M}\frac{\| g_k\|^2}{\|d_k\|^2}\left[c_1\| g_k\|^2 - \gamma\| g_k\|^2\right] \\ &\ge \frac{c_1\sigma}{c_2^2 M}(c_1 - \gamma)\| g_k\|^2 = \delta_1\| g_k\|^2 \quad \forall k \in K_1, \end{aligned} \tag{22}$$

where $\delta_1 = \frac{c_1\sigma}{c_2^2 M}(c_1 - \gamma)$.

On the other hand, if $k \in K_2$, we have $\alpha_k < s_k$. We define $\hat{\alpha}_k = \alpha_k/\rho$, so, by definition of the new Armijo rule, we get

$$D_k - f(x_k + \hat{\alpha}_k d_k) < -\sigma\hat{\alpha}_k\left[g_k^T d_k + \gamma\| g_k\|^2\right] \quad \forall k \in K_2.$$

From (15), we have

$$f_k - f(x_k + \hat{\alpha}_k d_k) < -\sigma\hat{\alpha}_k\left[g_k^T d_k + \gamma\| g_k\|^2\right] \quad \forall k \in K_2.$$

Using the mean value theorem on the left-hand side of previous inequality, there exists a constant $\mu \in [0, 1]$ such that

$$f_k - f(x_k + \hat{\alpha}_k d_k) = -\hat{\alpha}_k g(x_k + \mu\hat{\alpha}_k d_k)^T d_k \quad \forall k \in K_2.$$

Thus,

$$g(x_k + \mu\hat{\alpha}_k d_k)^T d_k > \sigma\big[\, g_k^T d_k + \gamma \|\, g_k\|^2\big] \quad \forall k \in K_2. \tag{23}$$

Now, by the Cauchy–Schwartz inequality and Remark 1, one can obtain

$$\begin{aligned}\hat{\alpha}_k L \|d_k\|^2 &\geq \|\, g(x_k + \mu\hat{\alpha}_k d_k) - g_k\| \|d_k\| \\ &\geq [g(x_k + \mu\hat{\alpha}_k d_k) - g_k]^T d_k \\ &> \sigma\big[\, g_k^T d_k + \gamma \|\, g_k\|^2\big] - g_k^T d_k \\ &> -(1-\sigma)\, g_k^T d_k \quad \forall k \in K_2.\end{aligned}$$

If set $\kappa = \frac{\rho(1-\sigma)}{L}$, then we have

$$\alpha_k \geq -\kappa \frac{g_k^T d_k}{\|d_k\|^2} \quad \forall k \in K_2. \tag{24}$$

Using (11) and (12) together with (9), (13) and (24), we obtain

$$\begin{aligned} D_k - f_{k+1} &\geq -\sigma\alpha_k\big[\, g_k^T d_k + \gamma\|\, g_k\|^2\big] \\ &= -\kappa\sigma \frac{g_k^T d_k}{\|d_k\|^2}\big[-g_k^T d_k - \gamma \|\, g_k\|^2\big] \\ &\geq c_1\kappa\sigma \frac{\|\, g_k\|^2}{\|d_k\|^2}\big[c_1\|\, g_k\|^2 - \gamma\|\, g_k\|^2\big] \\ &\geq \frac{c_1\kappa\sigma}{c_2^2}(c_1-\gamma)\|\, g_k\|^2 = \delta_2\|\, g_k\|^2 \quad \forall k \in K_2, \end{aligned} \tag{25}$$

where $\delta_2 = \frac{c_1\kappa\sigma}{c_2^2}(c_1-\gamma)$. Now, we set

$$\delta = \max\{\delta_1, \delta_2\}. \tag{26}$$

Therefore, (22), (25) together with (26) conclude the result. ■

THEOREM 3.3 *Suppose that* (H1) *and* (H2) *hold, and the sequence* $\{x_k\}$ *is generated by Algorithm* 2.1. *Then we have*

$$\lim_{k\to\infty} \|\, g_k\| = 0. \tag{27}$$

Proof Using (21), we know that there is a positive constant δ such that

$$D_k - f_{k+1} \geq \delta\|\, g_k\|^2 \geq 0 \quad \forall k \in \mathbf{N}.$$

Then Lemma 3.1, as $k \to \infty$, completes the proof. ■

4. Convergence rate analysis

Recently, Dai in [4,5] proved *R*-linearly convergence of the nonmonotone max-based line search, when the objective function $f(x)$ is uniformly convex. More recently, Zhang and Hager in [20] extended this property for their proposed nonmonotone line search algorithm for strongly convex functions. In this section, similar to [20], we establish the *R*-linearly convergence of our proposed algorithm for strongly convex functions.

Recall that the objective function $f(x)$ is strongly convex if there exists a scalar ω such that

$$f(x) \geq f(y) + \nabla f(y)^T(x-y) + \frac{1}{2\omega}\|x-y\|^2 \quad \forall x, y \in \mathbf{R}^n. \tag{28}$$

After some manipulations, we have

$$(\nabla f(x) - \nabla f(y))^T(x-y) \geq \frac{1}{\omega}\|x-y\|^2 \quad \forall x, y \in \mathbf{R}^n.$$

By choosing $y = x^*$, it can be written as

$$\|x - x^*\| \leq \omega \|\nabla f(x)\|. \tag{29}$$

For $t \in [0,1]$, we define $x(t) = x^* + t(x - x^*)$. Since $f(x)$ is a convex function, $f(x(t))$ is a convex function of t, and the derivative $f'(x(t))$ is an increasing function of $t \in [0,1]$ with $f'(x(0)) = 0$. Hence, for $t \in [0,1]$, $f'(x(t))$ attains its maximum value at $t = 1$. This observation combined with (29) gives

$$\begin{aligned} f(x) - f(x^*) &= \int_0^1 f'(x(t))\mathrm{d}t \leq f'(x(1)) = \nabla f(x)^T(x - x^*) \\ &\leq \|\nabla f(x)\|\|x - x^*\| \leq \omega \|\nabla f(x)\|^2. \end{aligned} \tag{30}$$

Lemma 4.1 *Assume that $f(x)$ is a twice continuously differentiable and strongly convex function on $\mathbf{R}^n$, the sequence $\{x_k\}$ is generated by Algorithm* 2.1 *converging to x^* and $\nabla f(x)$ is Lipschitz continuous on $L(x_0)$. Then there exists a constant $\theta \in (0,1)$ such that*

$$f(x_k) - f(x^*) \leq \theta^k(f(x_0) - f(x^*)) \quad \forall k \in \mathbf{N} \cup \{0\}. \tag{31}$$

Proof From (12) and the Cauchy–Schwartz inequality, we have

$$c_1\|g_k\|^2 \leq -g_k^T d_k \leq \|g_k\|\|d_k\|.$$

This leads to

$$\|d_k\| \geq c_1\|g_k\|. \tag{32}$$

Using (8), (11) and (32), we can write

$$\alpha_k \leq -\frac{g_k^T d_k}{d_k^T B_k d_k} \leq \frac{\|g_k\|.\|d_k\|}{m\|d_k\|^2} \leq \frac{1}{mc_1} \triangleq \mu. \tag{33}$$

Thanks to Lemma 2.1, we obtain

$$f(x_{k+1}) \leq D_{k+1} \leq D_k \leq \cdots \leq D_0 = f(x_0).$$

This implies that all iterates x_k are contained in the level set $L(x_0)$. Since $f(x)$ is strongly convex, it follows that $L(x_0)$ is bounded and $\nabla f(x)$ is Lipschitz continuous on $L(x_0)$.

Using (13) along with (33), we get

$$\|x_{k+1} - x_k\| = \alpha_k \|d_k\| \leq \mu c_2 \| g_k \|.$$

Thus,

$$\| g_{k+1} - g_k \| \leq L \|x_{k+1} - x_k\| \leq \mu c_2 L \| g_k \|.$$

Therefore,

$$\| g_{k+1} \| \leq \| g_{k+1} - g_k \| + \| g_k \| \leq (1 + \mu c_2 L) \| g_k \| = b \| g_k \|, \tag{34}$$

where $b = 1 + \mu c_2 L$.

In the rest of the proof, we show that, for each k, the following inequality holds.

$$D_{k+1} - f(x^*) \leq \theta (D_k - f(x^*)), \tag{35}$$

where

$$\theta = 1 - \delta b_2 (1 - \eta_{\max}) \quad \text{and} \quad b_2 = \frac{1}{\delta + \omega b^2}. \tag{36}$$

We divide the rest of the proof into two parts:

(i) If $\|g_k\|^2 \geq b_2(D_k - f(x^*))$. Definition of D_k, (10) and (21) imply that

$$\begin{aligned} D_{k+1} - f(x^*) &= f_{k+1} + \eta_k (D_k - f_{k+1}) - f(x^*) \\ &= \eta_k (D_k - f(x^*)) + (1 - \eta_k)(f_{k+1} - f(x^*)) \\ &\leq \eta_k (D_k - f(x^*)) + (1 - \eta_k)(D_k - f(x^*)) - (1 - \eta_k) \delta \| g_k \|^2 \\ &= D_k - f(x^*) - (1 - \eta_k) \delta \| g_k \|^2 \\ &\leq D_k - f(x^*) - (1 - \eta_{\max}) \delta \| g_k \|^2, \end{aligned}$$

since $\|g_k\|^2 \geq b_2(D_k - f(x^*))$, (35) has been established.

(ii) If $\|g_k\|^2 < b_2(D_k - f(x^*))$.

From (30) and (34), we have

$$f(x_{k+1}) - f(x^*) \leq \omega \| g_{k+1} \|^2 \leq \omega b^2 \| g_k \|^2 \leq \omega b^2 b_2 (D_k - f(x^*)). \tag{37}$$

Using (10), (21) and (37), we obtain

$$\begin{aligned} D_{k+1} - f(x^*) &= \eta_k (D_k - f(x^*)) + (1 - \eta_k)(f_{k+1} - f(x^*)) \\ &\leq [\eta_k + \omega b^2 b_2 (1 - \eta_k)](D_k - f(x^*)) \\ &= [\eta_k + (1 - \delta b_2)(1 - \eta_k)](D_k - f(x^*)) \\ &= [1 - \delta b_2 (1 - \eta_k)](D_k - f(x^*)) \\ &\leq [1 - \delta b_2 (1 - \eta_{\max})](D_k - f(x^*)). \end{aligned}$$

So, in both the cases, (35) holds. Therefore, we can write

$$D_k - f(x^*) \leq \theta(D_{k-1} - f(x^*)) \leq \theta^2(D_{k-2} - f(x^*)) \leq \cdots \leq \theta^k(f(x_0) - f(x^*)).$$

From (15), we have

$$f_k - f(x^*) \leq D_k - f(x^*) \leq \theta^k(f(x_0) - f(x^*)),$$

for all $k \in \mathbf{N} \cup \{0\}$.

On the other hand, by choosing $\eta_{\max}$ as enough as near to 1, we can conclude

$$0 < \delta b_2(1 - \eta_{\max}) < 1.$$

Therefore, $\theta \in (0, 1)$ and this completes the proof. ■

THEOREM 4.2 *Assume that all the conditions of Lemma* 4.1 *hold, then the sequence* $\{x_k\}$ *converges to* x^* *at least R-linearly.*

Proof Recall that the sequence $\{x_k\}$ converges to x^* R-linearly if there exists a sequence of nonnegative scalars $\{\nu_k\}$ such that

$$\|x_k - x^*\| \leq \nu_k \quad \forall k \in \mathbf{N} \cup \{0\},$$

where the sequence $\{\nu_k\}$ converges Q-linearly to zero. We firstly introduce the sequence $\{\nu_k\}$, and then we prove its Q-linearly convergence. From (28) together with choosing $y = x^*$, $x = x_k$ and using Lemma 4.1, we obtain

$$\|x_k - x^*\|^2 \leq 2\omega(f_k - f^*) \leq [2\omega(f_0 - f^*)]\theta^k = \bar{c}\theta^k,$$

where $\bar{c} = [2\omega(f_0 - f^*)]$. Setting $\nu_k = \bar{c}\theta^k$, we have that $\nu^* = 0$, and we get

$$\lim_{k\to\infty} \frac{\nu_{k+1} - \nu^*}{\nu_k - \nu^*} = \theta < 1,$$

as well. Therefore, the sequence $\{x_k\}$ converges to x^* at least R-linearly. ■

In the rest of this section, we prove that if we take the quasi-Newton or Newton direction, then the proposed algorithm is reduced to the quasi-Newton or Newton algorithm, respectively. Therefore, the algorithm can take the superlinear and quadratic convergence rate under some suitable conditions.

THEOREM 4.3 *Assume that* $f(x)$ *is a twice continuously differentiable function on* $\mathbf{R}^n$, *and all conditions of Theorem* 3.3 *hold. Furthermore, if* $d_k = -B_k^{-1}g_k$ *and* B_k *satisfies in the following condition*:

$$\lim_{k\to\infty} \frac{\|[B_k - G(x^*)]d_k\|}{\|d_k\|} = 0, \tag{38}$$

then the sequence $\{x_k\}$ *converges to* x^* *superlinearly.*

Proof Since $d_k = -B_k^{-1}g_k$, for sufficiently large k we get

$$s_k = -\frac{g_k^T d_k}{d_k^T B_k d_k} = 1.$$

Using (38), we have

$$\lim_{k\to\infty} \frac{\| g_k + G(x^*)d_k\|}{\|d_k\|} = 0, \tag{39}$$

i.e.,

$$d_k = -G(x^*)^{-1} g_k + o(\|d_k\|).$$

This suggests that

$$\|d_k\| \le \|G(x^*)^{-1}\| \| g_k\| + o(\|d_k\|).$$

Theorem 3.3 implies that $\|g_k\| \to 0$, as $k \to \infty$. Hence, we have

$$\lim_{k\to\infty} \|d_k\| = 0. \tag{40}$$

From (38), we have that $[B_k - G(x^*)]d_k = o(\|d_k\|)$ and $[G_k - G(x^*)]d_k = o(\|d_k\|)$. Using Taylor's expansion, one can obtain

$$\begin{aligned}
f(x_k + d_k) - f_k &= \left(f_k + g_k^T d_k + \frac{1}{2} d_k^T G_k d_k + o(\|d_k\|^2) \right) - f_k \\
&= \left[g_k^T d_k + \frac{1}{2} d_k^T B_k d_k \right] \\
&\quad + \frac{1}{2}\left[d_k^T (G_k - G(x^*)) d_k \right] + \frac{1}{2}\left[d_k^T (G(x^*) - B_k) d_k \right] + o(\|d_k\|^2) \\
&= \left[g_k^T d_k + \frac{1}{2} d_k^T B_k d_k \right] + o(\|d_k\|^2) = \frac{1}{2} g_k^T d_k + o(\|d_k\|^2) \\
&\le \frac{1}{2}\left[g_k^T d_k + \gamma \| g_k\| \right] + o(\|d_k\|^2) \le \sigma\left[g_k^T d_k + \gamma \| g_k\|^2 \right] + o(\|d_k\|^2).
\end{aligned}$$

Lemma 2.1 implies that

$$f(x_k + d_k) - D_k \le \sigma\left[g_k^T d_k + \gamma \| g_k\|^2 \right] + o(\|d_k\|^2).$$

Thus,

$$\lim_{k\to\infty} \frac{f(x_k + d_k) - D_k - \sigma\left[g_k^T d_k + \gamma \| g_k\|^2 \right]}{\|d_k\|^2} \le \lim_{k\to\infty} \frac{o(\|d_k\|^2)}{\|d_k\|^2} = 0.$$

So, for sufficiently large k, we have

$$f(x_k + d_k) - D_k \le \sigma\left[g_k^T d_k + \gamma \| g_k\|^2 \right]. \tag{41}$$

Therefore, for sufficiently large k, $x_{k+1} = x_k + d_k$ and the line search method reduced to standard quasi-Newton methods. It is known that the quasi-Newton methods, in the presence of (38), converge superlinearly [14]. So the sequence $\{x_k\}$ converges to x^* superlinearly. ■

THEOREM 4.4 *Assume that $f(x)$ is a twice continuously differentiable function on $\mathbf{R}^n$, and all conditions of Theorem 3.3 hold. Furthermore, if $d_k = -B_k^{-1} g_k$, $B_k = G_k$, and* *there exists a neighbourhood* $N(x^*, \epsilon)$ *of* x^* *such that* $G(x)$ *is Lipschitz continuous on* $N(x^*, \epsilon)$, *i.e., there exists* $L(\epsilon)$ *such that*

$$\|G(x) - G(y)\| \leq L(\epsilon)\|x - y\|, \tag{42}$$

then the sequence $\{x_k\}$ *converges to* x^* *quadratically.*

Proof Since $G(x)$ is Lipschitz continuous, we have

$$\|G(x) - G(x^*)\| \leq L(\epsilon)\|x - x^*\| \quad \forall x \in N(x^*, \epsilon).$$

Thus (38) holds, and all the hypotheses of Theorem 4.3 are fulfilled. So similar to Theorem 4.3, we can prove that (41) holds. Hence, the new method is reduced to the standard Newton method for sufficiently large k. This fact shows that the sequence $\{x_k\}$ is convergent to x^* quadratically. ■

5. Preliminary numerical experiments

In this section, we report some numerical experiments to illustrate the efficiency of the proposed line search algorithm (NMLS-M), in comparison with the nonmonotone line search of Grippo et al. (NMLS-G) [8], the nonmonotone line search of Shi and Shen (NMLS-S) [17] and the nonmonotone line search of Hager and Zhang (MNLS-H) [20]. We use unconstrained test collection of Andrei [2] and Moré et al. [13]. We have implemented our experiments in MATLAB 7.4 programming environment on a 3.0 GHz Intel Pentium IV WinXP PC with 1G RAM and double precision format.

In our experiments, the initial value of α_k for NMLS-G and NMLS-H is always set to 1, and for NMLS-S and NMLS-M the initial value is determined automatically by the following formula:

$$\alpha_0 = -\frac{g_k^T d_k}{g_k^T B_k g_k},$$

where $B_0 = I$. In all algorithms, we update the approximate Hessian matrix B_k by the BFGS quasi-Newton formula where $y_k^T s_k > 0$, $y_k = g_{k+1} - g_k$, and set $d_k = -B_k^{-1} g_k$. Our termination criterion was

$$\|\nabla f(x_k)\| \leq 10^{-8} \|\nabla f(x_0)\|.$$

For proper comparison on the algorithms, we used the same subroutine. This line search subroutine, similar to [17], computes the steplength α_k by the variant Armijo-type conditions with the parameters $\sigma = 0.38$ and $\rho = 0.618$. In addition, our preliminary experiments show that the best convergence results are obtained by η_k closer to 1 when the iterates are far from the optimum and by η_k closer to 0, when the iterates are near the optimum. Therefore, similar to [12], we update η_k dynamically like

$$\eta_k = \begin{cases} \eta_0/2, & \text{if } k = 1, \\ (\eta_{k-1} + \eta_{k-2})/2, & \text{if } k \geq 2, \end{cases}$$

where $\eta_0 = 0.85$. Like [20], we select $\eta_0 = 0.85$ for NMLS-H algorithm.

Table 1. Numerical results.

Prob. name	Dim.	NMLS-G n_i/n_f	NMLS-S n_i/n_f	NMLS-H n_i/n_f	NMLS-M n_i/n_f
Powell b. scal.	2	285/424	347/477	284/401	621/662
Brown b. scal.	2	13/72	13/72	13/72	13/72
Beale	2	13/26	13/26	13/26	14/26
E. Hilbert	2	3011/4500	3498/4793	2526/3990	1675/2652
Full Hess. FH1	2	28/39	28/39	28/39	28/39
Full Hess. FH2	2	5/9	5/9	5/9	5/9
MCCORMCK	2	9/13	15/19	9/13	10/14
Helical valley	3	42/84	42/84	46/82	35/69
Gaussian func.	3	6/11	6/11	6/11	6/11
Box three-dim.	3	36/63	36/63	39/69	39/71
Gulf r. and d.	3	36/51	36/51	32/50	32/50
Bro. a. Dennis	4	25/110	25/110	22/100	22/103
Wood	4	116/218	136/232	137/229	100/181
Biggs EXP6	6	Up to 20,000	60/75	60/75	47/67
GENHUMPS	6	51/248	108/488	22/216	22/216
SINQUAD	20	139/306	241/1762	125/586	141/286
Staircase 1	30	39/74	40/76	35/64	34/63
Staircase 2	30	54/80	54/82	49/78	34/63
Watson	31	66/124	65/122	64/123	55/122
Variably dim.	100	22/88	22/88	22/88	22/88
Penalty II	100	239/2525	264/2579	253/2400	280/1114
E. trid. 2	100	65/72	65/71	66/73	63/93
G. trid. 1	100	147/633	144/632	141/612	130/638
Diag. 3	100	134/746	135/748	126/749	120/778
Penalty I	200	232/1467	245/1265	95/485	85/223
E. Cliff	200	161/5010	159/4951	142/5103	134/5026
Hager	400	504/1654	568/1847	375/1209	117/439
VARDIM	600	17/108	111/215	17/108	18/109
Raydan 1	800	955/5521	960/5534	682/4066	236/1569
Diag. 9	800	1165/10,835	1166/10,673	1117/9960	906/9657
Diag. 1	900	1452/12,613	1485/12,596	1245/11,188	1063/11,311
Trigonometric	1000	83/89	86/100	83/84	83/84
E. Beale	1000	111/258	322/939	168/497	63/147
E. Wood	1000	3181/16,573	2778/16,206	2630/15,140	2130/15,279
E. F. a. Roth	1000	753/7620	764/7636	268/1993	73/494
E. Rosenbrock	1000	83/89	86/100	83/84	83/84
G. Rosenbrock	1000	4437/15,448	5725/17,080	1689/13,915	1367/14,119
E. W. a. Holst	1000	4002/18,343	3400/13,966	1676/13,914	2334/11,942
G. W. a. Holst	1000	6703/20,635	12947/28,999	13651/24,703	7920/19,803
P. quad.	1000	1105/13,222	1105/13,221	988/12,162	602/7814
G. Quartic	1000	255//640	155/232	115/249	75/134
P. trid. Quad.	1000	1105/13,253	1106/13,249	987/12,155	598/7763
Broyden trid.	1000	1475/8067	1339/7921	1138/7688	425/3290
E. trid. 1	1000	24/33	23/32	26/34	26/34
G. trid. 2	1000	1515/9217	1746/8030	1402/9161	566/7570
E. TET	1000	11/20	11/20	8/16	8/16
E. Himmelblau	1000	17/33	17/33	19/33	21/34
G. PSC1	1000	887/2333	891/2286	463/919	407/707
E. Maratos	1000	2053/13,764	2185/13,833	2001/13,598	1976/12,583
Full Hess. FH3	1000	3/20	3/20	3/20	3/20
Par. p. quad.	1000	1104/13,235	1104/13,234	961/11,868	558/7247
Alm p. quad.	1000	1105/13,233	1105/13,231	988/12,160	602/7813
E. BD1	1000	29/34	29/34	17/22	16/20

(*continued*)

Table 1. Continued.

Prob. name	Dim.	NMLS-G n_i/n_f	NMLS-S n_i/n_f	NMLS-H n_i/n_f	NMLS-M n_i/n_f
p. quad. diag.	1000	490/1221	427/1066	294/760	77/225
E. QP1	1000	410/939	411/873	371/678	190/327
E. QP2	1000	153/235	113/154	153/196	114/182
E. EP1	1000	4/18	4/18	4/18	4/18
DENSCHNB	1000	8/13	8/13	8/12	8/12
DENSCHNF	1000	1174/10,181	1169/10,128	469/3394	142/1025
HARKERP2	1000	397/1911	392/1897	331/1813	148/995
NONSCOMP	1000	1561/3133	1577/3230	1733/7338	1356/7366
DIXON3DQ	1000	1142/1565	1209/1569	1086/1539	610/1185
BIGGSB1	1000	1143/1563	1208/1562	1090/1546	609/1180
HIMMELH	1000	7/14	7/14	6/11	6/11
FLETCBV3	1000	643/646	Up to 20,000	643/646	643/646
FLETCHCR	1000	4690/17,805	2254/15,390	1814/14,727	1481/14,787
BDQRTIC	1000	1426/9153	1449/9176	919/5961	538/3956
TRIDIA	1000	1076/16,222	1076/16,222	1063/16,216	791/13,167
ARWHEAD	1000	7/31	7/31	8/28	8/28
NONDIA	1000	442/1854	1124/6675	589/3291	28/185
NONDQUAR	1000	820/1200	779/1176	736/888	748/843
EG2	1000	102/132	108/139	123/156	81/113
LIARWHD	1000	771/3306	717/3409	383/1587	131/589
POWER	1000	1009/25,836	1009/25,836	1011/25,890	1010/26,181
ENGVAL1	1000	1089/3689	1239/4112	504/1680	295/1095
EDENSCH	1000	718/2757	845/3253	435/1647	146/599
CUBE	1000	2804/15,141	3207/15,568	3188/15,369	1770/11,008
SINCOS	1000	14/26	12/23	20/30	20/30
Raydan 2	1000	5/7	5/7	5/7	5/7
Diag. 2	1000	284/285	284/285	284/285	284/285
Diag. 7	1000	5/8	5/8	6/8	6/8
Diag. 8	1000	5/9	5/9	5/9	5/9
Quad. QF1	1000	1114/11,818	1113/11,819	1016/11,040	358/4284
Quad. QF2	1000	1206/15,175	1207/15,184	1137/14,833	924/13,144
E. Powell sing.	1024	1231/5685	1237/5739	938/3980	533/2346
DIXMAANA	3000	6/9	6/8	6/9	6/9
DIXMAANB	3000	12/15	11/13	12/15	12/15
DIXMAANC	3000	8/11	12/14	8/11	8/11
DIXMAAND	3000	9/14	9/13	8/11	8/11
DIXMAANE	3000	291/293	291/292	291/293	291/293
DIXMAANF	3000	260/262	261/262	260/262	260/262
DIXMAANG	3000	252/254	252/254	252/254	252/254
DIXMAANH	3000	218/220	218/220	218/220	218/220
DIXMAANI	3000	5561/5569	5561/5569	5560/5569	5560/5569
DIXMAANJ	3000	1281/1293	1281/1288	873/982	857/929
DIXMAANK	3000	1449/1451	1465/1466	1449/1451	1449/1451
DIXMAANL	3000	1451/1453	1451/1453	1451/1453	1451/1453
Diag. 4	3000	3/14	3/14	3/14	3/14
Diag. 5	3000	5/7	5/6	5/6	5/6
BDEXP	5000	23/24	23/24	23/24	23/24
QUARTC	5000	21/26	21/26	21/26	21/26
HIMMELBG	5000	26/31	26/31	26/31	26/31
ARGLINB	5000	2/106	2/106	2/106	2/106
ARGLINC	5000	2/106	2/106	2/106	2/106
DQDRTIC	5000	21/56	23/56	23/53	21/55

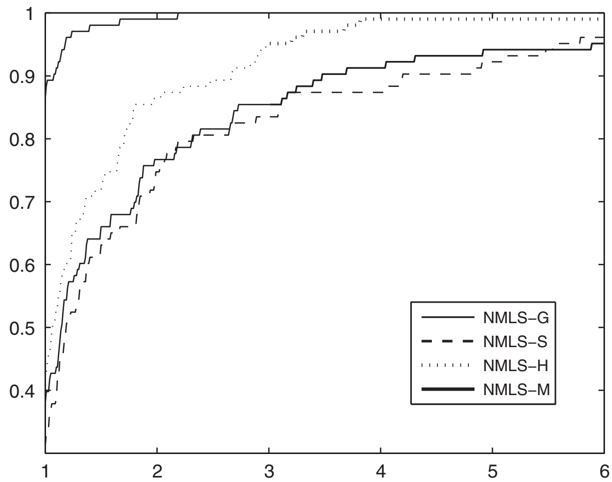

Figure 1. Performance profile for the total number of iterations.

In Table 1, we give an overview of the numerical performance while the first column 'Prob. name' and the second column 'Dim.' present the test problem name and the dimension of the test problem, respectively. Furthermore, the symbols n_i, n_f and n_g stand for the number of iterations, the number of function evaluations and the number of gradient evaluations, respectively. Since in Algorithm 2.1 the number of iterations and gradient evaluations are the same, we use the number of iterations and function evaluations as measure of performance.

Table 1 shows that in most cases both the number of iterations and function evaluations of the proposed algorithm are less than the other algorithms. Although, in some cases, this algorithm is not the best algorithm, but generally the proposed algorithm has better numerical results in comparison with the other algorithms.

Recently, for comparison of iterative algorithms, Dolan and Moré [7] proposed a new technique comparing the considered algorithms with statistical process by demonstration of performance profiles. In this process it is known that a plot of the performance profile reveals all of the major performance characteristics, which is a common tool to graphically compare effectiveness as well as robustness of the algorithms. In this technique, one can choose a performance index as measure of comparison among considered algorithms and can illustrate the results with performance profile. We can use two measures n_i and n_f to compare these algorithms. Hence, we use these two indexes for all of the present algorithms separately. Figures 1 and 2 show the performance of the above algorithms relative to these metrics, respectively.

In Figures 1 and 2, firstly, we observe that the proposed algorithm is the best algorithm on more than 85% and 75% of 105 mentioned test functions, respectively. Secondly, the proposed algorithm solves all of test functions. We also can see that the new algorithm grows up faster than the other algorithms. It means in the cases that the new algorithm is not the best algorithm closing to the performance index of the best algorithm. Therefore, we can deduce that the new algorithm is more efficient and robust than the other considered algorithms.

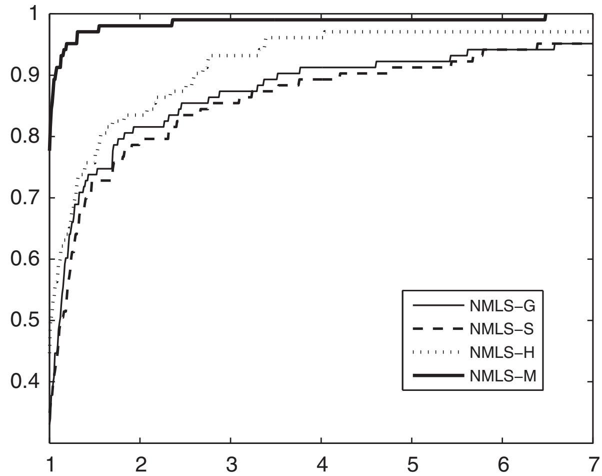

Figure 2. Performance profile for the total number of function evaluations.

6. Conclusions

In this article, we propose a new family of the Armijo-type line search approach. In the sense, we first relax the right-hand side of the standard Armijo rule while the new rule accepts a larger step-size especially whenever iterates are far from the optimum. On the other hand, we incorporate a nonmonotone strategy into this idea to attain a more robustness algorithm in the sense of the global convergence. We also prove that all limit points of the produced sequence are first-order critical points. Moreover, R-linear, superlinear and quadratic convergence of the new algorithm are investigated under some suitable conditions. Finally, we give detailed computational experiments and numerical comparisons to show that our approach is potentially efficient.

Acknowledgements

The authors are very grateful to the two anonymous referees for their valuable comments, which have improved this article greatly.

References

[1] M. Ahookhosh and K. Amini, *A nonmonotone trust region method with adaptive radius for unconstrained optimization*, Comput. Math. Appl. 60 (2010), pp. 411–422.
[2] N. Andrei, *An unconstrained optimization test functions collection*, Adv. Model. Optim. 10 (2008), pp. 147–161.
[3] L. Armijo, *Minimization of functions having Lipschitz continuous first partial derivatives*, Pac. J. Math. 16 (1966), pp. 1–3.
[4] Y.H. Dai, *On the nonmonotone line search*, J. Optim. Theory Appl. 112 (2002), pp. 315–330.
[5] Y.H. Dai, *R-linear convergence of the Barzilai and Borwein gradient method*, IMA J. Numer. Anal. 22 (2002), pp. 1–10.

[6] N.Y. Deng, Y. Xiao, and F.J. Zhou, *Nonmonotonic trust region algorithm*, J. Optim. Theory Appl. 76 (1993), pp. 259–285.
[7] E. Dolan and J.J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program. 91 (2002), pp. 201–213.
[8] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal. 23 (1986), pp. 707–716.
[9] L. Grippo, F. Lampariello, and S. Lucidi, *A truncated Newton method with nonmonotone line search for unconstrained optimization*, J. Optim. Theory Appl. 60 (1989), pp. 401–419.
[10] L. Grippo, F. Lampariello, and S. Lucidi, *A class of nonmonotone stabilization methods in unconstrained optimization*, Numer. Math. 62 (1991), pp. 779–805.
[11] N. Gu and J. Mo, *Incorporating nonmonotone strategies into the trust region method for unconstrained optimization*, Comput. Math. Appl. 55 (2008), pp. 2158–2172.
[12] J. Mo, C. Liu, and S. Yan, *A nonmonotone trust region method based on nonincreasing technique of weighted average of the successive function values*, J. Comput. Appl. Math. 209 (2007), pp. 97–108.
[13] J.J. Moré, B.S. Garbow, and K.E. Hillstrom, *Testing unconstrained optimization software*, ACM Transform. Math. Soft. 7 (1981), pp. 17–41.
[14] J. Nocedal and J.S. Wright, *Numerical Optimization*, Springer, New York, 1999.
[15] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
[16] E.R. Panier and A.L. Tits, *Avoiding the Maratos effect by means of a nonmonotone line search*, SIAM J. Numer. Anal. 28 (1991), pp. 1183–1195.
[17] Z.J. Shi and J. Shen, *Convergence of nonmonotone line search method*, J. Comput. Appl. Math. 193 (2006), pp. 397–412.
[18] Ph.L. Toint, *An assessment of nonmonotone linesearch technique for unconstrained optimization*, SIAM J. Sci. Comput. 17 (1996), pp. 725–739.
[19] Ph.L. Toint, *Non-monotone trust-region algorithm for non-linear optimization subject to convex constraints*, Math. Program. 77 (1997), pp. 69–94.
[20] H.C. Zhang and W.W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim. 14 (2004), pp. 1043–1056.