

Several New Line Search Methods and Their Convergence

Zhenjun Shi¹, Kimberly Kendricks¹, Zhiwei Xu², Yongning Tang³

¹Department of Mathematics and Computer Science, Central State University, Wilberforce, USA

²Department of Computer and Information Science, The University of Michigan, Dearborn, USA

³School of Information Technology, Illinois State University, Normal, USA

Email: zshi@centralstate.edu, kkendricks@centralstate.edu, zwxu@umich.edu, ytang@ilstu.edu

Received September 28, 2012; revised January 7, 2013; accepted January 14, 2013

Copyright © 2013 Zhenjun Shi *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

In this paper, we propose several new line search rules for solving unconstrained minimization problems. These new line search rules can extend the accepted scope of step sizes to a wider extent than the corresponding original ones and give an adequate initial step size at each iteration. It is proved that the resulting line search algorithms have global convergence under some mild conditions. It is also proved that the search direction plays an important role in line search methods and that the step size approaches mainly guarantee global convergence in general cases. The convergence rate of these methods is also investigated. Some numerical results show that these new line search algorithms are effective in practical computation.

Keywords: Unconstrained Minimization; Line Search Method; Global Convergence; Convergence Rate

1. Introduction

Consider an unconstrained minimization problem

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

where R^n is an n-dimensional Euclidean space, $f: R^n \rightarrow R^1$ a continuously differentiable function. Throughout this paper, we use $g(x) = \nabla f(x)$ as the gradient function of $f(x)$. Given an initial point x_0 , line search methods for solving (1) take the form

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \dots, \quad (2)$$

where x_k is the current iterate, d_k a search direction, and α_k a step size. Let x^* be a minimizer of (1) and thus be a stationary point that satisfies $g(x^*) = 0$. We denote $f(x_k)$ by f_k , $f(x^*)$ by f^* , and $\nabla f(x_k)$ by g_k , respectively.

In line search methods, there are two things to do at each iteration. One thing is to find a search direction d_k , and the other is to choose the step size α_k along the search direction d_k .

On the one hand, the search direction d_k is generally required to satisfy the descent condition

$$g_k^T d_k < 0, \quad (3)$$

which guarantees that d_k is a descent direction of $f(x)$ at x_k . In order to ensure the global convergence

of line search methods, we often require that the following condition holds,

$$-\frac{g_k^T d_k}{\|g_k\| \cdot \|d_k\|} \geq c, \quad (4)$$

where $c \in (0, 1]$ is a constant. The condition (4) is sometimes called the angle property (e.g., [1,2]). The choice of search direction d_k plays an important role in designing line search methods (e.g., [3]). There are many techniques for choosing the search direction d_k at the k th iteration (e.g., [2,4,5]).

On the other hand, we should choose α_k to satisfy some line search rules. Line search rules can be classified into two types, exact line search rules and inexact line search rules. This paper is devoted to the case of inexact line search rules. There are three well-known inexact line search rules [6-9].

Armijo rule. Set scalars $s_k > 0, \beta \in (0, 1)$, and $\sigma \in (0, 1/2)$. Choose α_k to be the largest α in $\{s_k, s_k \beta, s_k \beta^2, \dots\}$ such that

$$f_k - f(x_k + \alpha d_k) \geq -\sigma \alpha g_k^T d_k. \quad (5)$$

Remark. The original Armijo line search rule is to set $s_k = s$ with s being a constant [8].

Goldstein rule. A fixed scalar $\sigma \in \left(0, \frac{1}{2}\right)$ is selected,

and α_k is chosen to satisfy

$$\sigma \leq \frac{f(x_k + \alpha_k d_k) - f_k}{\alpha_k g_k^T d_k} \leq 1 - \sigma. \tag{6}$$

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

Wolfe Rule. Choose α_k to satisfy

$$f_k - f(x_k + \alpha_k d_k) \geq -\sigma \alpha_k g_k^T d_k \tag{7}$$

and

$$g(x_k + \alpha_k d_k)^T d_k \geq \gamma g_k^T d_k, \tag{8}$$

where σ and γ are some scalars with $\sigma \in (0, \frac{1}{2})$

and $\gamma \in (\sigma, 1)$.

Lemma 1.1 ([1]) For the Wolfe rule, we assume that there is a scalar M such that $f(x) \geq M$ for all

$x \in R^n$. Let $\sigma \in (0, \frac{1}{2})$ and $\gamma \in (\sigma, 1)$, and assume that

$g_k^T d_k < 0$. Then there exists an interval $[b_1, b_2]$ with $0 < b_1 < b_2$, such that every $\alpha \in [b_1, b_2]$ satisfies (7) and (8).

The above three line search rules can guarantee the existence of α_k under some mild conditions. However, how to find α_k is still a question. Especially, how to choose the initial step size s_k in the Armijo rule is also very important in practical computation. In fact, how to solve the inequalities (6), (7), (8) is also a problem in computation. Some implementable modified Armijo rules were proposed [10-13]. Moreover, some nonmonotonic line search methods were also investigated [14-17]. However, can we find an approach to solve (6), (7), and (8) easily and economically? Sometimes, we first set an initial step size s_k and substitute the test step size $\alpha = s_k$ into the inequalities (6), (7), or (8); if this α satisfies the inequalities, then we stop and find a step size $\alpha_k = \alpha$; otherwise, we need to use back-tracking or forward-tracking to adjust the test step size until we find an accepted step size α_k .

In order to find α_k easily and economically, we need to solve three problems. One problem is how to estimate the initial step size s_k . The second problem is how to adjust the test step size when the test step size doesn't satisfy the inequalities. The third problem is whether we can extend the accepted scope of step sizes to a wider extent. Our research is focused on the second and third questions.

In this paper, we propose several line search rules for solving unconstrained minimization problems. The modi-

fied line search rules used in the methods can extend the range of acceptable step sizes and give a suitable initial step size at each iteration. It is proved that the resulting line search methods have global convergence under some mild conditions. It is also proved that the search direction plays an important role in line search methods and that the step size rule mainly guarantees the global convergence in general cases. Numerical results show that the resulting algorithms are effective in practical computation.

The remainder of the paper is organized as follows. In Section 2, we describe the modified line search rules and its properties. In Section 3, we analyze the global convergence of resulting line search methods, and in Section 4, we study further the convergence rate of the new line search methods. Numerical results are reported in Section 5.

2. Modified Line Search Rules

We first assume that

(H1). The function f has a lower bound on $L_0 = \{x \in R^n \mid f(x) \leq f(x_0)\}$, where $x_0 \in R^n$ is given.

(H2). The gradient $g(x) = \nabla f(x)$ is uniformly continuous in an open convex set B that contains L_0 .

Sometimes, we further assume that

(H3). The gradient g is Lipschitz continuous in an open convex set B that contains L_0 , i.e., there exists $L > 0$ such that

$$\|g(x) - g(y)\| \leq L \|x - y\|, \quad \forall x, y \in B. \tag{9}$$

It is apparent that (H3) implies (H2).

In the following three modified line search rules, we define

$$w_k(\alpha) = \min \left\{ \frac{1}{2} \alpha \|d_k\|, -\frac{g_k^T d_k}{\|d_k\|} \right\}, \tag{10}$$

where $\alpha \in (0, +\infty)$ is a variable.

Modified Armijo Rule. Set scalars $L_k > 0, \beta \in (0, 1)$ and $\sigma \in (0, \frac{1}{2})$, and set $s_k = -\frac{g_k^T d_k}{L_k \|d_k\|^2}$. Let α_k be

the largest α in $\{s_k, \beta s_k, \beta^2 s_k, \dots\}$ such that

$$f_k - f(x_k + \alpha d_k) \geq \sigma \alpha \|d_k\| w_k(\alpha). \tag{11}$$

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

$$-(1 - \sigma) \alpha_k g_k^T d_k \geq f_k - f(x_k + \alpha_k d_k) \geq \sigma \alpha_k \|d_k\| w_k(\alpha_k). \tag{12}$$

Modified Wolfe Rule. The step size α_k is chosen to

satisfy

$$f_k - f(x_k + \alpha_k d_k) \geq \sigma \alpha_k \|d_k\| w_k(\alpha_k) \tag{13}$$

and

$$g(x_k + \alpha_k d_k)^T d_k \geq \gamma g_k^T d_k, \tag{14}$$

where σ and γ are some scalars with $\sigma \in (0, \frac{1}{2})$

and $\gamma \in (\sigma, 1)$.

Apparently, if α_k satisfies (5), or (6), or (7) and (8), then α_k also satisfies (11), or (12), or (13) and (14). We may obtain the conclusion from the fact that

$$\alpha_k \|d_k\| w_k(\alpha_k) \leq -\alpha_k g_k^T d_k.$$

Therefore, if (H1) holds, then the three modified line searches are feasible. As a result, the modified line searches can extend the range of acceptable step sizes α_k .

For the above three modified line search rules, we denote the three corresponding algorithms by Algorithm (NA), Algorithm (NG), and Algorithm (NW), respectively.

3. Global Convergence

In this section, we will prove that if (H1) and (H2) hold, d_k satisfies (3), α_k is chosen so that (11), or (12), or (13) and (14) hold. The related algorithms generate an infinite sequence $\{x_k\}$, then

$$\lim_{k \rightarrow +\infty} \frac{-g_k^T d_k}{\|d_k\|} = 0. \tag{15}$$

Theorem 3.1 Assume that (H1), (H2), and (3) hold, $\{L_k\}$ satisfies $0 < m_0 \leq L_k \leq M_0$ with m_0 and M_0 being positive numbers. Algorithm (NA) with modified Armijo rule generates an infinite sequence $\{x_k\}$. Then (15) holds.

Proof. For Algorithm (NA), by setting

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

and by (11), we have

$$f_k - f_{k+1} \geq s_k \sigma \|d_k\| w_k(s_k), \forall k \in K_1, \tag{16}$$

$$f_k - f_{k+1} \geq \alpha_k \sigma \|d_k\| w_k(\alpha_k), \forall k \in K_2. \tag{17}$$

Since α_k is the largest one to satisfy the modified Armijo rule, we will have $\alpha_k < \alpha_k / \beta \leq s_k, \forall k \in K_2$, and then $\alpha = \alpha_k / \beta$ makes the inequality sign of (11) change, i.e.,

$$f_k - f(x_k + \alpha_k / \beta d_k) < \alpha_k \sigma \|d_k\| w_k(\alpha_k / \beta), \forall k \in K_2.$$

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

$$\begin{aligned} & -\alpha_k g(x_k + \alpha_k \theta_k d_k / \beta)^T d_k / \beta \\ & = f_k - f(x_k + \alpha_k / \beta d_k) \\ & < \alpha_k \sigma \|d_k\| w_k(\alpha_k / \beta) / \beta \\ & \leq -\sigma \alpha_k g_k^T d_k / \beta, \forall k \in K_2. \end{aligned}$$

Therefore,

$$g(x_k + \theta_k \alpha_k d_k / \beta)^T d_k > \sigma g_k^T d_k, k \in K_2. \tag{18}$$

By (H1), (3), and (11), it follows that $\{f_k\}$ is a non-increasing number sequence and bounded from below, and it has a limit. Furthermore, we get from (11) that

$$\sum_{k=0}^{\infty} \alpha_k \|d_k\| w_k(\alpha_k) < +\infty, \tag{19}$$

and thus,

$$\sum_{k \in K_1} s_k \|d_k\| w_k(s_k) + \sum_{k \in K_2} \alpha_k \|d_k\| w_k(\alpha_k) < +\infty. \tag{20}$$

In order to prove (15), we use contrary proof to absurdity. Assume that there exists an $\epsilon > 0$ and an infinite subset $K_3 \subseteq \{0, 1, \dots\}$ such that

$$-\frac{g_k^T d_k}{\|d_k\|} \geq \epsilon, \forall k \in K_3. \tag{21}$$

Since $K_1 \cup K_2 = \{0, 1, \dots\}$, it is obvious that at least one of $K_1 \cap K_3$ or $K_2 \cap K_3$ is an infinite subset.

If $K_1 \cap K_2$ is an infinite subset then by (21) and (20) we have

$$\begin{aligned} & \sum_{k \in K_1 \cap K_3} \frac{\epsilon}{M_0} \min\left(\frac{\epsilon}{2M_0}, \epsilon\right) \\ & \leq \sum_{k \in K_1 \cap K_3} s_k \|d_k\| w_k(s_k) \\ & \leq \sum_{k \in K_1} s_k \|d_k\| w_k(s_k) \\ & \leq \sum_{k \in K_1} s_k \|d_k\| w_k(s_k) + \sum_{k \in K_2} \alpha_k \|d_k\| w_k(\alpha_k) \\ & < +\infty. \end{aligned}$$

The contradiction shows that $K_1 \cap K_3$ is not an infinite subset and $K_2 \cap K_3$ must be an infinite subset.

By (21) and (20), we have

$$\begin{aligned} & \sum_{k \in K_2 \cap K_3} \alpha_k \|d_k\| \min\left(\frac{1}{2} \alpha_k \|d_k\|, \epsilon\right) \\ & \leq \sum_{k \in K_2 \cap K_3} \alpha_k \|d_k\| w_k(\alpha_k) \\ & \leq \sum_{k \in K_2} \alpha_k \|d_k\| w_k(\alpha_k) \\ & \leq \sum_{k \in K_1} s_k \|d_k\| w_k(s_k) + \sum_{k \in K_2} \alpha_k \|d_k\| w_k(\alpha_k) \\ & < +\infty. \end{aligned}$$

Therefore,

$$\alpha_k \|d_k\| \rightarrow 0 (k \in K_2 \cap K_3). \tag{22}$$

By the Cauchy-Schwartz inequality and (18), we obtain

$$\begin{aligned} & \|g(x_k + \alpha_k \theta_k d_k / \beta) - g_k\| \\ &= \frac{\|g(x_k + \alpha_k \theta_k d_k / \beta) - g_k\| \cdot \|d_k\|}{\|d_k\|} \\ &\geq \frac{(g(x_k + \alpha_k \theta_k d_k / \beta) - g_k)^T d_k}{\|d_k\|} \\ &\geq (1 - \sigma) \frac{-g_k^T d_k}{\|d_k\|}, \quad k \in K_2 \cap K_3. \end{aligned}$$

By (H2), (22), and the above inequality, we have

$$-\frac{g_k^T d_k}{\|d_k\|} \rightarrow 0 (k \in K_2 \cap K_3, k \rightarrow +\infty),$$

which also contradicts (21). The contradiction shows that (15) holds.

Theorem 3.2 Assume that (H1), (H2), and (3) hold. Algorithm (NG) with the modified Goldstein line search generates an infinite sequence $\{x_k\}$. Then (15) holds.

Proof. By using the mean value theorem on the left-hand side inequality of (12), there exists $\theta_k \in [0, 1]$ such that

$$-\alpha_k g(x_k + \alpha_k \theta_k d_k)^T d_k = f_k - f_{k+1} \geq -(1 - \sigma) \alpha_k g_k^T d_k,$$

and thus,

$$g(x_k + \alpha_k \theta_k d_k)^T d_k \geq (1 - \sigma) g_k^T d_k. \tag{23}$$

By the right-hand side of (12) and (H1), it follows that $\{f_k\}$ is a monotone decreasing sequence and bounded below, and thus it has a limit. This shows that

$$\sum_{k=0}^{\infty} \alpha_k \|d_k\| w_k(\alpha_k) < +\infty. \tag{24}$$

Using the contrary proof, if (15) doesn't hold, then there exists an infinite subset $K \subseteq \{0, 1, \dots\}$ and an $\epsilon > 0$ such that

$$-\frac{g_k^T d_k}{\|d_k\|} > \epsilon, \forall k \in K. \tag{25}$$

By (25) and (24), we obtain

$$\begin{aligned} & \sum_{k \in K} \alpha_k \|d_k\| \min\left(\frac{1}{2} \alpha_k \|d_k\|, \epsilon\right) \\ &\leq \sum_{k \in K} \alpha_k \|d_k\| w_k(\alpha_k) \\ &\leq \sum_{k=0}^{\infty} \alpha_k \|d_k\| w_k(\alpha_k) < +\infty, \end{aligned}$$

and thus,

$$\alpha_k \|d_k\| \rightarrow 0 (k \in K, k \rightarrow +\infty). \tag{26}$$

By the Cauchy-Schwartz inequality and (23), we have

$$\begin{aligned} & \|g(x_k + \alpha_k \theta_k d_k) - g_k\| \\ &= \frac{\|g(x_k + \alpha_k \theta_k d_k) - g_k\| \cdot \|d_k\|}{\|d_k\|} \\ &\geq \frac{(g(x_k + \alpha_k \theta_k d_k) - g_k)^T d_k}{\|d_k\|} \\ &\geq \sigma \frac{-g_k^T d_k}{\|d_k\|}. \end{aligned}$$

By (26) and (H2), and noting that $\|\alpha_k \theta_k d_k\| \leq \alpha_k \|d_k\| \rightarrow 0 (k \in K, k \rightarrow \infty)$ and the above inequality, we have

$$\frac{-g_k^T d_k}{\|d_k\|} \rightarrow 0 (k \in K, k \rightarrow +\infty),$$

which contradicts (25). The contradiction shows that (15) holds.

Theorem 3.3 Assume that (H1), (H2), and (3) hold. Algorithm (NW) with modified Wolfe rule generates an infinite sequence $\{x_k\}$. Then (15) holds.

Proof. Using contrary proof, if (15) doesn't hold, then there exists an infinite subset $K \subseteq \{0, 1, \dots\}$ and an $\epsilon > 0$ such that (25) holds. By (H1) and (13), it follows that (24) holds, and thus (26) holds.

By the Cauchy-Schwartz inequality and (14), we have

$$\begin{aligned} & \|g_{k+1} - g_k\| \\ &= \frac{\|g_{k+1} - g_k\| \cdot \|d_k\|}{\|d_k\|} \geq \frac{(g_{k+1} - g_k)^T d_k}{\|d_k\|} \geq (1 - \gamma) \frac{-g_k^T d_k}{\|d_k\|}. \end{aligned}$$

By (H2), (26), and the above inequality, we have

$$-\frac{g_k^T d_k}{\|d_k\|} \rightarrow 0 (k \in K, k \rightarrow \infty),$$

which is a contradiction to (25). The contradiction shows that (15) holds.

Theorem 3.4 Assume that (H1), (H3), and (3) hold, Algorithm (NA) with $0 < m_0 \leq L_k \leq M_0$, or Algorithm (NG), or Algorithm (NW) generates an infinite sequence $\{x_k\}$. Then there exists $k_1 \in \{0, 1, \dots\}$ and $\eta_0 > 0$ such that

$$f_k - f_{k+1} \geq \eta_0 \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2, \forall k \geq k_1. \tag{27}$$

Proof. Since (H3) implies (H2), the conclusions in Theorems 3.1, 3.2, and 3.3 are also true. We will use these conclusions and notations in the proofs of these theorems to our proof.

For Algorithm (NA), by (H3), the Cauchy-Schwartz

inequality, and (18), we have

$$\begin{aligned} \alpha_k L \|d_k\|^2 / \beta &\geq \|g(x_k + \alpha_k \theta_k d_k / \beta) - g_k\| \cdot \|d_k\| \\ &\geq (g(x_k + \alpha_k \theta_k d_k / \beta) - g_k)^T d_k \\ &> -(1 - \sigma) g_k^T d_k, \quad k \in K_2, \end{aligned}$$

where K_2 and θ_k are defined in the proof of Theorem 3.1. Thus,

$$\alpha_k > -\beta(1 - \sigma) \frac{g_k^T d_k}{L \|d_k\|^2}, \quad k \in K_2. \tag{28}$$

By (17), (28), and the proof of Theorem 3.1, we obtain

$$f_k - f_{k+1} \geq \frac{\sigma\beta(1 - \sigma)}{L} \min\left(\frac{\beta(1 - \sigma)}{2L}, 1\right) \cdot \left(\frac{-g_k^T d_k}{\|d_k\|}\right)^2, \tag{29}$$

where $k \in K_2, k \geq k_1$. By (16) and the proof of Theorem 3.1, we have

$$f_k - f_{k+1} \geq \frac{\sigma}{M_0} \min\left(\frac{1}{2M_0}, 1\right) \left(\frac{-g_k^T d_k}{\|d_k\|}\right)^2, \quad k \in K_1, k \geq k_1, \tag{30}$$

where K_1 is defined in the proof of Theorem 3.1. Set

$$\begin{aligned} \eta_0 &= \sigma \min\left\{\frac{\beta - \beta\sigma}{L} \min\left(\frac{\beta - \beta\sigma}{2L}, 1\right), \frac{1}{M_0} \min\left(\frac{1}{2M_0}, 1\right)\right\}, \end{aligned}$$

it follows that (27) holds.

For Algorithm (NG), by (H3), the Cauchy-Schwartz inequality, and (23), we have

$$\begin{aligned} \alpha_k L \|d_k\|^2 &\geq \|g(x_k + \alpha_k \theta_k d_k) - g_k\| \cdot \|d_k\| \\ &\geq (g(x_k + \alpha_k \theta_k d_k) - g_k)^T d_k \\ &\geq -\sigma g_k^T d_k, \end{aligned}$$

and thus,

$$\alpha_k \geq -\frac{\sigma}{L} \frac{g_k^T d_k}{\|d_k\|^2}. \tag{31}$$

By (12), (31), and Theorem 3.4, we have

$$f_k - f_{k+1} \geq \frac{\sigma^2}{L} \min\left(\frac{\sigma}{2L}, 1\right) \left(\frac{-g_k^T d_k}{\|d_k\|}\right)^2, \quad k \geq k_1.$$

By setting

$$\eta_0 = \frac{\sigma^2}{L} \min\left(\frac{\sigma}{2L}, 1\right),$$

it follows that (27) holds.

For Algorithm (NW), by (H3), the Cauchy-Schwartz inequality, and (14), we can prove that

$$\alpha_k \geq -\frac{(1 - \beta) g_k^T d_k}{L \|d_k\|^2}. \tag{32}$$

By (13), (32), and Theorem 3.1, we have

$$f_k - f_{k+1} \geq \frac{\sigma(1 - \beta)}{L} \min\left(\frac{1 - \beta}{2L}, 1\right) \left(\frac{-g_k^T d_k}{\|d_k\|}\right)^2, \quad k \geq k_1. \tag{33}$$

By setting

$$\eta_0 = \frac{\sigma(1 - \beta)}{L} \min\left(\frac{1 - \beta}{2L}, 1\right),$$

it follows that (27) holds.

Theorem 3.5 Assume that (H1), (H2), and (4) hold, Algorithm (NA) with $0 < m_0 \leq L_k \leq M_0$, or Algorithm (NG), or Algorithm (NW) generates an infinite sequence $\{x_k\}$. Then

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \tag{34}$$

Proof. By Theorems 3.1, 3.2, and 3.3, we obtain that (15) holds. By (4), when $k \rightarrow \infty$, we have

$$c^2 \|g_k\|^2 \leq \left(\frac{-g_k^T d_k}{\|d_k\| \cdot \|g_k\|}\right)^2 \cdot \|g_k\|^2 = \left(\frac{-g_k^T d_k}{\|d_k\|}\right)^2 \rightarrow 0.$$

Therefore, (15) holds.

Corollary 3.1 Assume that (H1), (H3), and (4) hold, Algorithm (NA) with $0 < m_0 \leq L_k \leq M_0$, or Algorithm (NG), or Algorithm (NW) generates an infinite sequence $\{x_k\}$. Then (34) holds.

Proof. By Theorem 3.5 and (4), we have

$$\begin{aligned} f_k - f_{k+1} &\geq \eta_0 \left(\frac{-g_k^T d_k}{\|d_k\|}\right)^2 \\ &= \eta_0 \left(\frac{-g_k^T d_k}{\|d_k\| \cdot \|g_k\|}\right)^2 \cdot \|g_k\|^2 \geq \eta_0 c^2 \|g_k\|^2. \end{aligned}$$

By (H1) and the above inequality, we obtain that (34) holds.

Remark: Because (H3) implies (H2), it is obvious that the conditions of Corollary 3.1 imply the conditions of Theorem 3.5. Thus, Corollary 3.1 holds.

4. Convergence Rate

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

(H4): f is twice continuously differentiable and uniformly convex on R^n .

Lemma 4.1 Assume that (H4) holds. Then (H1) and (H3) hold, $f(x)$ has a unique minimizer x^* , and there exists $0 < m \leq M$ such that

$$m \|y\|^2 \leq y^T \nabla^2 f(x) y \leq M \|y\|^2, \quad \forall x, y \in R^n; \tag{35}$$

$$\frac{1}{2}m\|x-x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2}M\|x-x^*\|^2, \forall x \in R^n; \tag{36}$$

$$M\|x-y\|^2 \geq (g(x) - g(y))^T(x-y) \geq m\|x-y\|^2, \tag{37}$$

where $\forall x, y \in R^n$; and thus

$$M\|x-x^*\|^2 \geq g(x)^T(x-x^*) \geq m\|x-x^*\|^2, \forall x \in R^n. \tag{38}$$

By (37) and (38), we can also obtain from the Cauchy-Schwartz inequality that

$$M\|x-y\| \geq \|g(x) - g(y)\| \geq m\|x-y\| \tag{39}$$

and

$$M\|x-x^*\| \geq \|g(x)\| \geq m\|x-x^*\|, \forall x \in R^n. \tag{40}$$

Its proof can be seen from (e.g., [18,19], etc.). In this case, the Lipschitz constant L of the gradient function $\nabla f(x)$ is M .

Theorem 4.1 Suppose that (H4) and (4) hold, any of the three algorithms generates an infinite sequence $\{x_k\}$. Then $\{x_k\}$ converges to x^* at least R -linearly.

Proof. By Corollary 3.1 and Lemma 4.1, it follows that $\{x_k\}$ converges to x^* . By Theorem 3.5, (4) and (40), we have

$$\begin{aligned} f_k - f_{k+1} &\geq \eta_0 \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2 \\ &= \eta_0 \left(-\frac{g_k^T d_k}{\|g_k\| \cdot \|d_k\|} \right)^2 \|g_k\|^2 \\ &\geq \eta_0 c^2 \|g_k\|^2 \geq \eta_0 c^2 m^2 \|x_k - x^*\|^2 \\ &\geq 2 \frac{\eta_0 c^2 m^2}{M} (f_k - f(x^*)). \end{aligned}$$

By the above inequality and setting $\rho = \sqrt{2 \frac{\eta_0 c^2 m^2}{M}}$,

we can prove that $\rho < 1$. In fact, by Theorem 3.5 and (39), and noting that $L = M$, we can obtain

$$\eta_0 \leq \frac{\sigma}{L} = \frac{\sigma}{M}, \tag{41}$$

and thus,

$$\rho^2 = \frac{2\eta_0 c^2 m^2}{M} \leq \frac{2\sigma c^2 m^2}{M^2} \leq 2\sigma c^2 < 1.$$

By setting $\theta = \sqrt{1 - \rho^2}$, we have

$$\begin{aligned} f_k - f(x^*) &\leq (1 - \rho^2)(f_{k-1} - f(x^*)) \\ &= \theta^2 (f_{k-1} - f(x^*)) \leq \dots \leq \theta^{2k} (f_0 - f(x^*)). \end{aligned}$$

From the above first inequality, we can say that $\{f_k\}$

converges to f^* at least Q -linearly. From the last inequality, we get that $\{f_k\}$ converges to f^* at least R -linearly [18].

By setting $\omega = \sqrt{\frac{2(f_0 - f(x^*))}{m}}$, it follows that

$$\|x_k - x^*\|^2 \leq \frac{2}{m} (f_k - f(x^*)) \leq \frac{2(f_0 - f(x^*))}{m} \theta^{2k} = \omega^2 \theta^{2k},$$

and thus

$$\|x_k - x^*\| \leq \omega \theta^k,$$

which shows that $\{x_k\}$ converges to x^* at least R -linearly [18].

5. Numerical Results

In the above sections, we investigated the theoretical properties and analyzed the global convergence and convergence rate of related line search methods under some mild conditions. In this section, we will study the numerical performance of algorithms with the new line search approaches.

First, we choose some numerical examples to test the Algorithms (NA), (NG), and (NW) and make some comparisons to the algorithms with the original line searches. The original line search methods are denoted by OA, OG, and OW, respectively.

The numerical examples are from [12]. We use the same symbols to denote the problems. For example, (P5) denotes Problem (P5) in [12]. For each problem, the limiting number of functional evaluations is set to 10,000, and the stopping condition is

$$\|g_k\| \leq 10^{-6}. \tag{42}$$

We choose a portable computer with a Pentium 1.2 MHz CPU and Matlab 6.1 to test our algorithms. The parameters are set to $\sigma = 0.38$, $\beta = 0.87$, $\gamma = 0.618$, $L_0 = 1$, and

$$L_k = \frac{|(x_k - x_{k-1})^T (g_k - g_{k-1})|}{|x_k - x_{k-1}|^2}, \tag{43}$$

for $k \geq 1$.

Algorithm (NA): Step 0. Given parameters $\sigma = 0.38$, $\beta = 0.87$, and set $k := 0$;

Step 1. If $\|g_k\| < 10^{-6}$ then stop else taking

$d_k = -g_k$;

Step 2. Find step size α_k by using the modified Armijo rule;

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$ and set $k := k + 1$, go to Step 1.

For the Goldstein and the modified Goldstein rules, we use the following procedure to find the step size.

Algorithm (NG): Step 0. Given parameters $\rho_0 = 2$ and $\rho = 0.5$, set $k := 0$;

Step 1. If $\|g_k\| < 10^{-6}$ then stop else taking $d_k = -g_k$;

Step 2. Find step size α_k by using the following procedure:

Procedure 5.1. Set $\alpha = s_k = -\frac{g_k^T d_k}{L_k \|d_k\|^2}$ in which L_k

is defined by (43). Set two indices $dec = 0$ and $inc = 0$.

Step 21. If $inc * dec = 1$, then $\rho := \sqrt{\rho}$, $\rho_0 := \sqrt{\rho_0}$, $\alpha_k = s_k$, and set $inc = 0$, $dec = 0$;

Step 22. If (12) holds then $\alpha_k = \alpha$ and stop;

Step 23. If the right-hand inequality of (12) holds and the left doesn't hold, then $\alpha := \rho_0 \alpha$ and set $inc = 1$, and go to Step 21;

Step 24. If the left-hand inequality of (12) holds and the right doesn't hold, then $\alpha := \rho \alpha$ and set $dec = 1$, and go to Step 21;

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$ and set $k := k + 1$, go to Step 1.

For Wolfe and modified Wolfe rules, we use the following procedure to find the step size.

Algorithm (NW): Step 0. Given parameters $\rho_0 = 2$ and $\rho = 0.5$, set $k := 0$;

Step 1. If $\|g_k\| < 10^{-6}$ then stop else taking $d_k = -g_k$;

Step 2. Find step size α_k by using the following procedure:

Procedure 5.2. Set $\alpha = s_k = -\frac{g_k^T d_k}{L_k \|d_k\|^2}$ in which L_k

is defined by (43). Set two indices $dec = 0$ and $inc = 0$.

Step 21. If $inc * dec = 1$, then $\rho := \sqrt{\rho}$, $\rho_0 := \sqrt{\rho_0}$, $\alpha_k = s_k$, and set $inc = 0$, $dec = 0$;

Step 22. If (13) and (14) hold, then $\alpha_k = \alpha$ and stop;

Step 23. If (13) holds and (14) doesn't hold, then $\alpha := \rho_0 \alpha$ and set $inc = 1$, and go to Step 21;

Step 24. If (14) holds and (13) doesn't hold, then $\alpha := \rho \alpha$ and set $dec = 1$, and go to Step 21;

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$ and set $k := k + 1$, go to Step 1.

In numerical experiments, we take $d_k = -g_k$.

We can see from **Table 1** that the modified Wolfe line approach is the best one in practical computation. In fact, NA, NG and NW are all better than OA, OG, and OW. This shows that the modified line searches are effective in practical computation and significantly reduce the number of functional evaluations and iterations when reaching the same precision. Moreover, we found that the new modified line approaches can be used to any descent methods. For example, we can take quasi-Newton direction $d_k = -H_k g_k$ in the line search methods and use these modified line search approaches to find a step size.

Table 1. The number of iterations and functional evaluations.

| P | n | OA | OG | OW | NA | NG | NW |
|-----|------|---------|---------|---------|--------|--------|--------|
| P5 | 2 | 8/12 | 7/11 | 7/10 | 6/7 | 6/7 | 6/7 |
| P13 | 4 | 23/35 | 21/32 | 20/29 | 17/20 | 16/21 | 15/22 |
| P14 | 4 | 36/52 | 32/48 | 32/48 | 23/32 | 23/35 | 21/29 |
| P16 | 4 | 14/63 | 16/67 | 14/62 | 16/43 | 17/51 | 11/29 |
| P20 | 9 | 12/17 | 12/19 | 13/15 | 11/14 | 11/14 | 11/12 |
| P21 | 16 | 18/25 | 18/27 | 18/26 | 12/24 | 12/27 | 11/20 |
| P23 | 8 | 30/41 | 33/49 | 31/45 | 25/34 | 23/36 | 24/32 |
| P24 | 20 | 52/64 | 55/71 | 50/58 | 35/46 | 33/42 | 28/38 |
| P25 | 50 | 11/123 | 12/118 | 10/67 | 11/21 | 12/23 | 11/18 |
| P26 | 50 | 14/30 | 16/42 | 15/38 | 12/28 | 12/24 | 11/19 |
| P30 | 20 | 13/22 | 13/26 | 13/23 | 11/26 | 11/21 | 10/18 |
| P21 | 1000 | 98/563 | 89/483 | 76/428 | 67/310 | 54/258 | 48/211 |
| P21 | 5000 | 143/721 | 126/565 | 118/475 | 76/426 | 78/384 | 68/236 |
| P23 | 1000 | 120/798 | 117/695 | 120/572 | 73/275 | 76/248 | 64/198 |
| P23 | 5000 | 185/775 | 167/883 | 151/663 | 68/85 | 61/83 | 42/65 |
| CPU | - | 189s | 198s | 135s | 121s | 117s | 93s |

If we take L_k like $L_0 = 1$, and

$$L_k = \frac{\|g_k - g_{k-1}\|}{\|x_k - x_{k-1}\|}, \tag{44}$$

for $k \geq 1$, and use (44) to replace (43) in Procedure 5.2, we have the results in **Table 2**. From **Table 3**, we found that the estimation (43) seems better than (44) in many situations. Actually, by Cauchy-Schwartz inequality, we

Table 2. The number of iterations and functional evaluations.

| P | n | OA | OG | OW | NA | NG | NW |
|-----|------|---------|---------|---------|--------|--------|--------|
| P5 | 2 | 8/12 | 7/11 | 7/10 | 7/11 | 7/12 | 7/11 |
| P13 | 4 | 23/35 | 21/32 | 20/29 | 19/23 | 18/25 | 18/25 |
| P14 | 4 | 36/52 | 32/48 | 32/48 | 25/37 | 25/37 | 27/31 |
| P16 | 4 | 14/63 | 16/67 | 14/62 | 16/45 | 18/53 | 12/32 |
| P20 | 9 | 12/17 | 12/19 | 13/15 | 11/18 | 11/19 | 12/19 |
| P21 | 16 | 18/25 | 18/27 | 18/26 | 14/31 | 12/25 | 14/28 |
| P23 | 8 | 30/41 | 33/49 | 31/45 | 26/37 | 25/37 | 26/38 |
| P24 | 20 | 52/64 | 55/71 | 50/58 | 39/48 | 32/45 | 29/43 |
| P25 | 50 | 11/123 | 12/118 | 10/67 | 11/32 | 12/34 | 12/28 |
| P26 | 50 | 14/30 | 16/42 | 15/38 | 12/35 | 13/44 | 14/27 |
| P30 | 20 | 13/22 | 13/26 | 13/23 | 12/27 | 13/26 | 12/26 |
| P21 | 1000 | 98/563 | 89/483 | 76/428 | 69/354 | 58/263 | 53/265 |
| P21 | 5000 | 143/721 | 126/565 | 118/475 | 78/433 | 81/420 | 73/263 |
| P23 | 1000 | 120/798 | 117/695 | 120/572 | 94/321 | 85/291 | 69/238 |
| P23 | 5000 | 185/775 | 167/883 | 151/663 | 72/120 | 63/72 | 43/69 |
| CPU | - | 189s | 198s | 135s | 132s | 141s | 128s |

Table 3. The number of iterations and functional evaluations.

| P | n | OA | OG | OW | NA | NG | NW |
|----------|-----------------|--------|--------|--------|--------|-------|-------|
| ARWHEAD | 10 ⁴ | 46/198 | 42/137 | 32/120 | 26/64 | 25/47 | 25/54 |
| DQDR TIC | 10 ⁴ | 18/43 | 17/38 | 15/32 | 16/31 | 15/43 | 14/28 |
| ENGVAL1 | 10 ⁴ | 18/36 | 17/29 | 15/28 | 12/25 | 12/25 | 12/25 |
| VAREIGVL | 10 ⁴ | 21/28 | 18/32 | 17/31 | 16/25 | 17/32 | 15/28 |
| WOODS | 10 ⁴ | 19/29 | 17/26 | 15/24 | 17/32 | 16/43 | 16/41 |
| LIARWHD | 10 ⁴ | 32/76 | 35/88 | 31/45 | 28/46 | 28/52 | 28/34 |
| MOREBV | 10 ⁴ | 76/86 | 72/124 | 73/97 | 72/97 | 72/97 | 60/81 |
| NONDIA | 10 ⁴ | 32/44 | 28/52 | 26/42 | 22/36 | 25/48 | 23/34 |
| TQUARTIC | 10 ⁴ | 29/43 | 27/38 | 25/31 | 24/36 | 24/43 | 24/43 |
| POWELLSG | 10 ⁴ | 57/126 | 59/119 | 53/97 | 48/126 | 53/92 | 42/77 |
| QUARTC | 10 ⁴ | 53/127 | 42/116 | 34/98 | 32/74 | 33/68 | 31/66 |
| SCHMVETT | 10 ⁴ | 28/75 | 28/92 | 25/82 | 24/38 | 25/42 | 23/38 |
| SPARSQUR | 10 ⁴ | 54/93 | 47/110 | 48/90 | 42/73 | 45/73 | 40/65 |
| SROSENBR | 10 ⁴ | 31/95 | 29/85 | 25/82 | 29/66 | 25/78 | 24/64 |
| TOINTGSS | 10 ⁴ | 26/67 | 28/73 | 26/75 | 24/56 | 21/52 | 24/48 |
| CPU | | 125s | 116s | 108s | 97s | 88s | 72s |

have

$$\frac{\left| (x_k - x_{k-1})^T (g_k - g_{k-1}) \right|}{\|x_k - x_{k-1}\|^2} \leq \frac{\|g_k - g_{k-1}\|}{\|x_k - x_{k-1}\|},$$

which shows that (43) can produce a larger initial step size s_k than (44) does.

In fact, in the beginning of iterations, larger step size can quicken the convergence of resulting algorithms for solving well conditioned middle scale problems. For large scale or even ill-conditioned problems, what is the performance of the resulting algorithms with new modified line search rules? We chose standard test problems from [20-22] to test the new modified line search rules. The numerical results are reported in **Table 3**.

For large scale problems, numerical results show that the resulting algorithms with the modified line search rules are more effective than the original ones in many cases. The reason is that step size estimation is more useful for such large scale problems that have sparse Hessian matrix. Larger step size plays an important role in the convergence of resulting algorithms with modified line searches. Therefore, initial step size estimation and extending the range of acceptable step sizes is necessary in line search design and algorithm design.

6. Conclusions

We proposed several new line search algorithms for solving unconstrained minimization problems. The modified line search rules used in the methods can extend the range of acceptable step sizes to a wider extent and give a suitable initial step size at each iteration. It is clarified that the resulting line search algorithms have global convergence under weak conditions. It is also proved that the search direction plays an important role in these methods and that the step size mainly guarantees global convergence in some cases. The convergence rate of these algorithms is investigated. These theoretical results can help us design new algorithms in practice. Furthermore, we extended the line search methods theoretically in some broad sense. Numerical results show that these new modified line search rules are useful and effective in practical computation.

For the future research, we should study the numerical performance of special line search methods for large-scale practical problems. Moreover, we can generalize these modified line search approaches to nonmonotone cases [14,16]. We can also investigate step-size in different ways [23-25].

7. Acknowledgements

The authors are very grateful to the referees and editors for their helpful and valuable comments and suggestions

that greatly improved the paper.

REFERENCES

- [1] D. P. Bertsekas, "Constrained Optimization and Lagrange Multiplier Methods," Academic Press Inc., Waltham, 1982.
- [2] J. Nocedal and J. W. Stephen, "Numerical Optimization," Springer-Verlag New York, Inc., New York, 1999. [doi:10.1007/b98874](https://doi.org/10.1007/b98874)
- [3] Y. G. Evtushenko, "Numerical Optimization Techniques," Optimization Software Inc., Publications Division, New York, 1985.
- [4] J. P. Dussault, "Convergence of Implementable Descent Algorithms for Unconstrained Optimization," *Journal of Optimization Theory and Applications*, Vol. 104, No. 3, 2000, pp. 749-750. [doi:10.1023/A:1004602012151](https://doi.org/10.1023/A:1004602012151)
- [5] B. Rhanizar, "Hybrid Procedures for Solving Some Unconstrained Nonlinear Optimization Problems," *Applied Numerical Mathematics*, Vol. 30, No. 4, 1999, pp. 459-474. [doi:10.1016/S0168-9274\(98\)00068-3](https://doi.org/10.1016/S0168-9274(98)00068-3)
- [6] L. Armijo, "Minimization of Functions Having Lipschitz Continuous First Partial Derivatives," *Pacific Journal of Mathematics*, Vol. 16, No. 1, 1966, pp. 1-3. [doi:10.2140/pjm.1966.16.1](https://doi.org/10.2140/pjm.1966.16.1)
- [7] A. A. Goldstein, "On Steepest Descent Method," *Journal on Society for the Industrial and Applied Mathematics Series A Control*, Vol. 3, No. 1, 1965, pp. 147-151.
- [8] W. Y. Sun and Y. X. Yuan, "Optimization Theory and Methods—Nonlinear Programming," Springer Optimization and Its Applications, Vol. 1. Springer, New York, 2006.
- [9] P. Wolfe, "Convergence Conditions for Ascent Methods," *SIAM Review*, Vol. 11, No. 2, 1969, pp. 226-235. [doi:10.1137/1011036](https://doi.org/10.1137/1011036)
- [10] Z. J. Shi, "Convergence of Line Search Methods for Unconstrained Optimization," *Applied Mathematics and Computation*, Vol. 157, No. 2, 2004, pp. 393-405. [doi:10.1016/j.amc.2003.08.058](https://doi.org/10.1016/j.amc.2003.08.058)
- [11] Z. J. Shi and J. Shen, "Convergence of Descent Method without Line Search," *Applied Mathematics and Computation*, Vol. 167, No. 1, 2005, pp. 94-107. [doi:10.1016/j.amc.2004.06.097](https://doi.org/10.1016/j.amc.2004.06.097)
- [12] Z. J. Shi and J. Shen, "New Inexact Line Search Method for Unconstrained Optimization," *Journal of Optimization Theory and Applications*, Vol. 127, No. 2, 2005, pp. 425-446. [doi:10.1007/s10957-005-6553-6](https://doi.org/10.1007/s10957-005-6553-6)
- [13] Z. J. Shi and J. Shen, "Step Size Estimation for Unconstrained Optimization Methods," *Journal of Computational and Applied Mathematics*, Vol. 24, No. 3, 2005, pp. 399-417.
- [14] Y. H. Dai, "On the Nonmonotone Line Search," *Journal of Optimization Theory and Applications*, Vol. 112, No. 2, 2002, pp. 315-330. [doi:10.1023/A:1013653923062](https://doi.org/10.1023/A:1013653923062)
- [15] Z. J. Shi and J. Shen, "Convergence of Nonmonotone Line Search Method," *Journal of Computational and Applied Mathematics*, Vol. 193, No. 2, 2006, pp. 397-412. [doi:10.1016/j.cam.2005.06.033](https://doi.org/10.1016/j.cam.2005.06.033)

- [16] W. Y. Sun, J. Y. Han and J. Sun, "Global Convergence of Nonmonotone Descent Methods for Unconstrained Optimization Problems," *Journal of Computational and Applied Mathematics*, Vol. 146, No. 1, 2002, pp. 89-98. [doi:10.1016/S0377-0427\(02\)00420-X](https://doi.org/10.1016/S0377-0427(02)00420-X)
- [17] H. C. Zhang and W. W. Hager, "A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization," *The SIAM Journal on Optimization*, Vol. 14, No. 4, 2004, pp. 1043-1056. [doi:10.1137/S1052623403428208](https://doi.org/10.1137/S1052623403428208)
- [18] J. M. Ortega and W. C. Rheinboldt, "Iterative Solution of Nonlinear Equations in Several Variables," Academic Press, Waltham, 1970.
- [19] R. T. Rockafellar, "Convex Analysis," Princeton University Press, Princeton, 1970.
- [20] I. Bongartz, A. R. Conn, N. I. M. Gould and P. L. Toint, "CUTE: Constrained and Unconstrained Testing Environments," *ACM Transactions on Mathematical Software*, Vol. 21, No. 1, 1995, pp. 123-160. [doi:10.1145/200979.201043](https://doi.org/10.1145/200979.201043)
- [21] W. W. Hager and H. C. Zhang, "A New Conjugate Gradient Method with Guaranteed Descent and an Efficient Line Search," *The SIAM Journal on Optimization*, Vol. 16, No. 1, 2005, pp. 170-192. [doi:10.1137/030601880](https://doi.org/10.1137/030601880)
- [22] W. W. Hager and H. C. Zhang, "Algorithm 851: CG DESCENT, A Conjugate Gradient Method with Guaranteed Descent," *ACM Transactions on Mathematical Software*, Vol. 32, No. 1, 2006, pp. 113-137. [doi:10.1145/1132973.1132979](https://doi.org/10.1145/1132973.1132979)
- [23] A. I. Cohen, "Stepsize Analysis for Descent Methods," *Journal of Optimization Theory and Applications*, Vol. 33, No. 2, 1981, pp. 187-205. [doi:10.1007/BF00935546](https://doi.org/10.1007/BF00935546)
- [24] J. C. Gilbert and J. Nocedal, "Global Convergence Properties of Conjugate Gradient Methods for Optimization," *The SIAM Journal on Optimization*, Vol. 2, No. 1, 1992, pp. 21-42. [doi:10.1137/0802003](https://doi.org/10.1137/0802003)
- [25] Z. J. Shi, "A New Memory Gradient Method under Exact Line Search," *Asia-Pacific Journal of Operational Research*, Vol. 20, No. 2, 2003, pp. 275-284.