From Line Search Method to Trust Region Method*

Zhen-Jun Shi[1,2,†]   Xiang-Sun Zhang[3]

[1] College of Operations Research and Management
Qufu Normal University, Rizhao, Shandong 276826, PRC
[2] Institute of Computational Mathematics and Scientific/Engineering Computing
Academy of Mathematics and Systems Science, CAS, Beijing 100080, PRC
[3] Institute of Applied Mathematics
Academy of Mathematics and Systems Science, CAS, Beijing 100080, PRC

Abstract *Line search method and trust region method are two important classes of techniques for solving optimization problems and have their advantages respectively. In this paper we use the Armijo line search rule in a more accurate way and propose a new line search method for unconstrained optimization problems. Global convergence and convergence rate of the new method are analyzed under mild conditions. Furthermore, the new Armijo-type line search strategy is shown to be equivalent to an approximation of a trust region method then has both advantages of line search strategy and trust region strategy.*

Keywords unconstrained optimization, line search method, global convergence, convergence rate.

1 Introduction

Consider the unconstrained minimization problem

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

where R^n denotes an n-dimensional Euclidean space and $f : R^n \to R^1$ is a continuously differentiable function.

Traditional iterative methods for solving (1) are either line search method or trust region method. Line search method is based on searching a new iterative point along a descent direction at each iteration and trust region method is based on finding a new iterative point within a ball centered at the current iterate.

*The work was supported in part by NSF of China, Grant 10171054, Postdoctoral Fund of China and K. C. Wong Postdoctoral Fund of CAS, Grant 6765700.

†Corresponding author. E-mail address: zjshi@lsec.cc.ac.cn or zjshi@qrnu.edu.cn

Generally, line search method takes the form

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

where d_k is a descent direction of $f(x)$ at x_k and α_k is a step size. For convenience, we denote $\nabla f(x_k)$ by g_k, $f(x_k)$ by f_k, $\nabla^2 f(x_k)$ by G_k and $f(x^*)$ by f^*, respectively. If G_k is available and inverse, then $d_k = G_k^{-1} g_k$ leads to the Newton method while $d_k = -g_k$ results in the steepest descent method (e.g.[2, 3]). 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(e.g.[3, 9, 10]). In order to guarantee the global convergence, we sometimes require d_k to satisfy the sufficient descent condition

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

where $c > 0$ is a constant. Instead of (4), the angle property is often used in proving the global convergence of related line search methods, that is

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

where $1 \ge \tau > 0$.

Once the descent direction d_k is determined we should seek a step size along the descent direction and complete one iterate.

There are many approaches to find an available step size. It is well known that the exact line search is time-consuming, then inexact line search rules, such as Armijo rule([1]), Goldstein rule and Wolfe rule, etc., see [3, 4, 5], are generally used. Convergence analysis on line search methods can be seen in the literatures (e.g.[4, 5, 7, 8]). Among them the Armijo rule is most useful and easy to implement in practical computation.

Armijo Rule. Let $s > 0$ be a constant, $\rho \in (0, 1)$ and $\mu \in (0, 1)$. Take α_k to be the largest α in $\{s, s\rho, s\rho^2, ...,\}$ such that

$$f_k - f(x_k + \alpha d_k) \ge -\alpha\mu g_k^T d_k.$$

On the other hand, unlike the line search method, the trust region requires to solve the following subproblem at each iteration

$$\min m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p, \quad s.t.\ \|p\| \le \Delta_k, \tag{6}$$

with B_k being an approximation to G_k and Δ_k being a trust region radius and obtain a solution p_k. By observing the value

$$r_k = \frac{f_k - f(x_k + p_k)}{m_k(0) - m_k(p_k)}, \tag{7}$$

we can assert whether or not the new point $\overline{x}_k = x_k + p_k$ is accepted. Trust region method avoids line search but needs to solve a subproblem (see [6, 9, 12]). Given $\mu \in [0, \frac{1}{4})$, one has the accepted strategy

$$x_{k+1} = \begin{cases} \overline{x}_k, & \text{if } r_k \geq \mu; \\ x_k, & \text{otherwise.} \end{cases} \tag{8}$$

In this paper we extend the Armijo line search rule to a general form and propose a new line search method for unconstrained optimization problems. The global convergence and convergence rate of the new method are analyzed under mild conditions. Furthermore, the new Armijo-type line search rule enables us to reveal the relationship between line search method and trust region method.

The rest of this paper is organized as follows. In the next section we introduce a novel usage of Armijo rule and develop a line search method. In Sections 3 and 4 we analyze its global convergence and convergence rate respectively. In Section 5 we reveal some relationships between the new line search method and trust region method. Conclusion remarks are given in Section 6.

2 A Novel Usage of Armijo Rule

We first assume that

(**H1**). The objective function $f(x)$ is continuously differentiable and has a lower bound on R^n.

(**H2**). The gradient $g(x)$ of $f(x)$ is uniformly continuous on an open convex set B that contains the level set $L_0 = \{x \in R^n | f(x) \leq f(x_0)\}$, where x_0 is given.

(**H2**)$'$. The gradient $g(x)$ of $f(x)$ is Lipschitz continuous on an open convex set B that contains the level set L_0, i.e., there exists M' such that

$$\|g(x) - g(y)\| \leq M' \|x - y\|, \ \forall x, y \in B.$$

It is apparent that $(H2)'$ implies $(H2)$.

Now we define a new usage of the Armijo rule or simply call it new Armijo search.

New Armijo Search. Given $\mu \in (0, \frac{1}{2})$ and $\rho \in (0, 1)$, B_k is an approximation to $G_k = \nabla^2 f(x_k)$ and $\hat{B}_k$ is defined by the following procedure: take i to be the smallest integer such that $d_k^T B_k d_k + i\|d_k\|^2 > 0$. Set $s_k = -\frac{g_k^T d_k}{d_k^T \hat{B}_k d_k}$ and α_k is the largest α in $\{s_k, s_k\rho, s_k\rho^2, ...,\}$ such that

$$f_k - f(x_k + \alpha d_k) \geq -\alpha\mu[g_k^T d_k + \frac{1}{2}\alpha d_k^T B_k d_k].$$

Algorithm (A).

Step 0. Choose $x_0 \in R^n$ and set $k := 0$.

Step 1. If $\|g_k\| = 0$ then stop; else go to Step 2;

Step 2. Set $x_{k+1} = x_k + \alpha_k d_k$ where d_k is a descent direction of $f(x)$ at x_k and α_k is selected by the new Armijo search;

Step 3. Set $k := k + 1$ and go to Step 1.

Lemma 2.1. If $(H1)$ holds and $g_k^T d_k < 0$, then the new Armijo search is well-defined.

Proof. By $(H1)$ we have

$$\lim_{\alpha \to 0+} [\frac{f(x_k + \alpha d_k) - f_k - \frac{1}{2}\mu\alpha^2 d_k^T B_k d_k}{\alpha}] = g_k^T d_k < \mu g_k^T d_k.$$

Therefore, there exists an $\overline{\alpha}_k > 0$ such that

$$\frac{f(x_k + \alpha d_k) - f_k - \frac{1}{2}\mu\alpha^2 d_k^T B_k d_k}{\alpha} \leq \mu g_k^T d_k, \ \forall \alpha \in [0, \overline{\alpha}_k],$$

which implies that the new Armijo search is well-defined.

□

3 Global convergence

Theorem 3.1. If $(H1)$ and $(H2)$ hold, d_k satisfies (3) and α_k is defined by the new Armijo search. Algorithm (A) generates an infinite sequence $\{x_k\}$ with a bounded sequence $\{B_k\}$, that is, there is a β such that $\|B_k\| \leq \beta$, $\forall k$. Then

$$\lim_{k \to \infty} (-\frac{g_k^T d_k}{\|d_k\|}) = 0. \tag{9}$$

Proof. For contrary, if there exist an infinite subset $K \subseteq \{0, 1, 2, 3, ...\}$ and an $\epsilon > 0$ such that

$$-\frac{g_k^T d_k}{\|d_k\|} \geq \epsilon, \ k \in K, \tag{10}$$

then

$$-g_k^T d_k \geq \epsilon \|d_k\|, \forall k \in K. \tag{11}$$

By the new Armijo search and (11), in the case of $d_k^T B_k d_k \leq 0$ $(k \in K)$, we have

$$f_k - f_{k+1} \geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2}\alpha_k d_k^T B_k d_k] \geq -\alpha_k \mu g_k^T d_k \geq \alpha_k \mu \epsilon \|d_k\|;$$

and in the case of $d_k^T B_k d_k > 0$ $(k \in K)$, since $\alpha_k \leq s_k = -\frac{g_k^T d_k}{d_k^T B_k d_k}$, we have

$$\begin{aligned} f_k - f_{k+1} &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2}\alpha_k d_k^T B_k d_k] \\ &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2} s_k d_k^T B_k d_k] \\ &= -\frac{1}{2}\alpha_k \mu g_k^T d_k \\ &\geq \frac{\mu\epsilon}{2}\alpha_k \|d_k\|, \ \forall k \in K. \end{aligned}$$

This and $(H1)$ imply that

$$\alpha_k d_k \leq \alpha_k \|d_k\| \to 0 \ (k \in K, \ k \to \infty). \tag{12}$$

Also by the new Armijo search, $\|B_k\| \leq \beta$ implies that

$$\|\hat{B}_k\| \leq 2\beta + 1, \forall k. \tag{13}$$

Let

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

we can prove that K_1 is a finite subset. In fact, if K_1 is an infinite subset, (12) and (13) imply that

$$-\frac{g_k^T d_k \|d_k\|}{(2\beta+1)\|d_k\|^2} \leq -\frac{g_k^T d_k}{d_k^T \hat{B}_k d_k}\|d_k\| = \alpha_k \|d_k\| \to 0 \ (k \in K_1, \ k \to \infty),$$

which contradicts (10). Thus K_2 must be an infinite subset and $\alpha_k/\rho \leq s_k, \ \forall k \in K_2$. By the new Armijo search, $\alpha = \alpha_k/\rho$ will make the inequality in the new Armijo search fail to hold, i.e.,

$$f_k - f(x_k + (\alpha_k/\rho)d_k) < -(\alpha_k/\rho)\mu[g_k^T d_k + \frac{1}{2}(\alpha_k/\rho)d_k^T B_k d_k], \ k \in K_2.$$

Therefore

$$\begin{aligned} f(x_k + (\alpha_k/\rho)d_k) - f_k &\geq (\alpha_k/\rho)\mu[g_k^T d_k + \frac{1}{2}(\alpha_k/\rho)d_k^T B_k d_k] \\ &\geq (\alpha_k/\rho)\mu[g_k^T d_k - \frac{1}{2}(\alpha_k/\rho)d_k^T \hat{B}_k d_k] \\ &\geq (\alpha_k/\rho)\mu[g_k^T d_k - \frac{1}{2}s_k d_k^T \hat{B}_k d_k] \\ &= \frac{3}{2}(\alpha_k/\rho)\mu g_k^T d_k, \ k \in K_2. \end{aligned}$$

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

$$(\alpha_k/\rho)g(x_k + \theta_k(\alpha_k/\rho)d_k)^T d_k \geq \frac{3}{2}(\alpha_k/\rho)\mu g_k^T d_k, \ k \in K_2.$$

Hence,

$$g(x_k + \theta_k(\alpha_k/\rho)d_k)^T d_k \geq \frac{3}{2}\mu g_k^T d_k, \ k \in K_2. \tag{14}$$

By Cauchy-Schiwartz inequality, (13), (12) and $(H2)$ we have

$$\begin{aligned} -(1 - \frac{3}{2}\mu)\frac{g_k^T d_k}{\|d_k\|} &\leq \frac{[g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k]^T d_k}{\|d_k\|} \\ &\leq \frac{\|g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k\| \cdot \|d_k\|}{\|d_k\|} \\ &= \|g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k\| \to 0 \ (k \in K_2, k \to \infty), \end{aligned}$$

which also contradicts (10) and thus the conclusion holds.

□

Theorem 3.2. If $(H1)$ and $(H2)'$ hold, d_k satisfies (3) and α_k is defined by the new Armijo search. Algorithm (A) generates an infinite sequence $\{x_k\}$ and $\{B_k\}$ is uniformly bounded, that is, there is a $\beta > 0$ such that $\|B_k\| \le \beta$, $\forall k$. Then

$$\sum_{k=0}^{\infty}(-\frac{g_k^T d_k}{\|d_k\|})^2 < +\infty. \tag{15}$$

Proof: Since $(H2)'$ implies $(H2)$, the conclusion in Theorem 3.1 holds. Let

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

we obtain by (13) that

$$\begin{aligned} f_k - f_{k+1} &\ge -\alpha_k\mu[g_k^T d_k + \frac{1}{2}\alpha_k d_k B_k d_k] \\ &\ge -\alpha_k\mu[g_k^T d_k + \frac{1}{2}\alpha_k d_k \hat{B}_k d_k] \\ &= -\frac{1}{2}\alpha_k\mu g_k^T d_k \\ &= \frac{1}{2}\mu\frac{(g_k^T d_k)^2}{d_k^T \hat{B}_k d_k} \\ &\ge \frac{\mu}{2(2\beta+1)}(-\frac{g_k^T d_k}{\|d_k\|})^2,\ k \in K_1. \end{aligned}$$

Thus,

$$f_k - f_{k+1} \ge \frac{\mu}{2(2\beta+1)}(-\frac{g_k^T d_k}{\|d_k\|})^2,\ k \in K_1. \tag{16}$$

In the case of $k \in K_2$ we have $\alpha_k/\rho \le s_k$, we can prove similarly as (14) that

$$g(x_k + \theta_k(\alpha_k/\rho)d_k)^T d_k \ge \frac{3}{2}\mu g_k^T d_k,\ k \in K_2, \tag{17}$$

and thus $(H2)'$ implies that

$$-(1-\frac{3}{2}\mu)\frac{g_k^T d_k}{\|d_k\|} \le \|g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k\| \le M'\alpha_k/\rho\|d_k\|,\ k \in K_2,$$

i.e.,

$$\alpha_k \ge -\rho(1-\frac{3}{2}\mu)M'^{-1}\frac{g_k^T d_k}{\|d_k\|^2}, k \in K_2. \tag{18}$$

Therefore,

$$\begin{aligned} f_k - f_{k+1} &\geq -\alpha_k\mu[g_k^T d_k + \frac{1}{2}\alpha_k d_k^T B_k d_k] \\ &\geq -\alpha_k\mu[g_k^T d_k + \frac{1}{2}s_k d_k^T \hat{B}_k d_k] \\ &\geq \frac{\mu\rho(1-\frac{3}{2}\mu)}{2M'}(-\frac{g_k^T d_k}{\|d_k\|})^2, \ k \in K_2. \end{aligned}$$

By the above inequality, (16) and letting

$$\eta' = \frac{\mu}{2}\min(\frac{1}{2\beta+1}, \frac{\rho(1-\frac{3}{2}\mu)}{M'}),$$

we have

$$f_k - f_{k+1} \geq \eta'(-\frac{g_k^T d_k}{\|d_k\|})^2, \ \forall k. \tag{19}$$

This and $(H1)$ imply that (15) holds.

□

Corollary 3.1. If $(H1)$ and $(H2)$ hold, d_k satisfies (5) and α_k is defined by the new Armijo search. Algorithm (A) generates an infinite sequence $\{x_k\}$ and $\|B_k\| \leq \beta, \forall k$. Then

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{20}$$

Proof. By Theorem 3.1, we have

$$\tau\|g_k\| \leq -\frac{g_k^T d_k}{\|g_k\| \cdot \|d_k\|}\|g_k\| = -\frac{g_k^T d_k}{\|d_k\|} \to 0(k\to\infty).$$

The proof is finished.

□

4 Convergence Rate

In order to analyze the convergence rate, we further assume that

$(H3)$. $x_k \to x^*$ as $k \to \infty$, $\nabla^2 f(x^*) \succ 0$ and $f(x)$ is twice continuously differentiable on $N(x^*, \epsilon_0) = \{x | \ \|x - x^*\| < \epsilon_0\}$.

Lemma 4.1. Assume that $(H3)$ holds. Then there exist $0 < m' \leq M'$ and $\epsilon \leq \epsilon_0$ such that

$$m'\|y\|^2 \leq y^T \nabla^2 f(x) y \leq M'\|y\|^2, \ \forall x, y \in N(x^*, \epsilon); \tag{21}$$

$$\frac{1}{2}m'\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2}M'\|x - x^*\|^2, \ \forall x \in N(x^*, \epsilon); \tag{22}$$

$$M'\|x-y\|^2 \geq (g(x)-g(y))^T(x-y) \geq m'\|x-y\|^2,\ \forall x,y \in N(x^*,\epsilon); \tag{23}$$

and thus

$$M'\|x-x^*\|^2 \geq g(x)^T(x-x^*) \geq m'\|x-x^*\|^2,\ \forall x \in N(x^*,\epsilon). \tag{24}$$

By (24) and (23) we can also obtain, from the Cauchy-Schwartz inequality, that

$$M'\|x-x^*\| \geq \|g(x)\| \geq m'\|x-x^*\|,\ \forall x \in N(x^*,\epsilon), \tag{25}$$

and

$$\|g(x)-g(y)\| \leq M'\|x-y\|, \forall x,y \in N(x^*,\epsilon). \tag{26}$$

Its proof can be found in the literature (e.g.[5]).

4.1 Linear Convergence Rate

Theorem 4.1. Assume that $(H3)$ holds, d_k satisfies (5) and α_k is defined by the new Armijo search and that $\|B_k\| \leq \beta, \forall k$. If Algorithm (A) generates an infinite sequence $\{x_k\}$, then $\{x_k\}$ converges to x^* at least R-linearly.

Proof. If $(H3)$ holds then there exists k' such that $x_k \in N(x^*,\epsilon_0), \forall k \geq k'$ and $(H1)$ and $(H2)'$ hold if $x_0 \in N(x^*,\epsilon_0)$. By Theorem 3.2 and (5) we have

$$f_k - f_{k+1} \geq \eta'(-\frac{g_k^T d_k}{\|d_k\|})^2 \geq \eta'\tau^2\|g_k\|^2, k \geq k'.$$

By the above inequality and Lemma 4.1, letting $\eta = \eta'\tau^2$, we obtain

$$\begin{aligned} f_k - f_{k+1} &\geq \eta\|g_k\|^2 \\ &\geq \eta m'^2\|x_k - x^*\|^2 \\ &\geq \frac{2\eta m'^2}{M'}(f_k - f^*). \end{aligned}$$

Set

$$\theta = m'\sqrt{\frac{2\eta}{M'}},$$

we can prove that $\theta < 1$. In fact, by the definition of η and η' in the proof of Theorem 3.2, we obtain

$$\begin{aligned} \theta^2 &= \frac{2m'^2\eta}{M'} \leq \frac{2m'^2\tau^2\eta'}{M'} \\ &\leq \frac{2m'^2\tau^2}{M'} \cdot \frac{\mu\rho(1-\frac{3}{2}\mu)}{2M'} \\ &\leq \tau^2\mu\rho(1-\frac{3}{2}\mu) \\ &\leq \mu\rho(1-\frac{3}{2}\mu) < 1. \end{aligned}$$

By setting

$$\omega = \sqrt{1-\theta^2},$$

(obviously $\omega < 1$), we obtain from the above inequalities that

$$\begin{aligned} f_{k+1} - f^* &\leq (1-\theta^2)(f_k - f^*) \\ &= \omega^2(f_k - f^*) \\ &\leq \ldots \\ &\leq \omega^{2(k-k')}(f_{k'+1} - f^*). \end{aligned}$$

By Lemma 4.1 and the above inequalities we have

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \frac{2}{m'}(f_{k+1} - f^*) \\ &\leq \omega^{2(k-k')}\frac{2(f_{k'+1} - f^*)}{m'}, \end{aligned}$$

thus

$$\|x_{k+1} - x^*\| \leq \omega^{k-k'}\sqrt{\frac{2(f_{k'+1} - f^*)}{m'}},$$

i.e.,

$$\|x_k - x^*\| \leq \omega^k \sqrt{\frac{2(f_{k'+1} - f^*)}{m'\omega^{2(k'+1)}}}.$$

We finally have

$$R_1\{x_k\} = \lim_{k\to\infty} \|x_k - x^*\|^{1/k} \leq \omega < 1,$$

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

□

4.2 Superlinear Convergence Rate

We further assume that

(H4). $\{B_k\}$ is a sequence of positive definite matrices and $\|B_k\| \leq \beta, \forall k$. Algorithm (A) with $d_k = -B_k^{-1}g_k$ satisfies the following condition

$$\lim_{k\to\infty} \frac{\|[B_k - \nabla^2 f(x^*)]d_k\|}{\|d_k\|} = 0. \tag{27}$$

Lemma 4.2. *If* $(H3)$ *and* $(H4)$ *hold. Algorithm (A) generates an infinite sequence* $\{x_k\}$*. Then there exists* k' *such that*

$$\alpha_k = 1, \ \forall k \geq k'. \tag{28}$$

Proof. By Corollary 3.1 and $(H3)$ we have

$$\lim_{k\to\infty} x_k = x^*, \quad \lim_{k\to\infty} \|d_k\| = 0, \tag{29}$$

and thus

$$\lim_{k\to\infty}(x_k + td_k - x^*) = 0, \tag{30}$$

where $t \in [0, 1]$. Assumption $(H4)$ implies that

$$d_k^T[B_k - \nabla^2 f(x^*)]d_k = o(\|d_k\|^2). \tag{31}$$

By the mean value theorem, $(H3)$, (29), (30) and (31), for sufficiently large k, we have

$$\begin{aligned} f(x_k + d_k) - f_k &= g_k^T d_k + \int_0^1 (1-t)d_k^T \nabla^2 f(x_k + td_k) d_k dt \\ &= [g_k^T d_k + \frac{1}{2} d_k B_k d_k] \\ &\quad + \int_0^1 (1-t) d_k^T [\nabla^2 f(x_k + td_k) - \nabla^2 f(x^*)] d_k dt \\ &\quad + \frac{1}{2} d_k^T [\nabla^2 f(x^*) - B_k] d_k \\ &= [g_k^T d_k + \frac{1}{2} d_k B_k d_k] + o(\|d_k\|^2) \\ &\leq \mu [g_k^T d_k + \frac{1}{2} d_k B_k d_k]. \end{aligned}$$

This implies that there exists k' making (28) valid.

□

Theorem 4.2. *If $(H3)$ and $(H4)$ hold. Algorithm (A) generates an infinite sequence $\{x_k\}$. Then $\{x_k\}$ converges to x^* superlinearly.*

Proof. By Corollary 3.1 and Lemma 4.1 we know that $\{x_k\} \to x^*$. By Lemma 4.2, there exists k' such that (28) holds and we have

$$x_{k+1} = x_k + d_k, \; k \geq k',$$

where $d_k = B_k^{-1} g_k$. By the mean value theorem, Lemma 4.1 and (30), it follows that

$$\begin{aligned} g_{k+1} - g_k &= \int_0^1 \nabla^2 f(x_k + t(x_{k+1} - x_k))(x_{k+1} - x_k) dt \\ &= \int_0^1 \nabla^2 f(x_k + td_k) d_k dt \\ &= \nabla^2 f(x^*) d_k + \int_0^1 [\nabla^2 f(x_k + td_k) - \nabla^2 f(x^*)] d_k dt \\ &= \nabla^2 f(x^*) d_k + o(\|d_k\|), \end{aligned}$$

thus

$$\begin{aligned} g_{k+1} &= g_k + \nabla^2 f(x^*) d_k + o(\|d_k\|) \\ &= -B_k d_k + \nabla^2 f(x^*) d_k + o(\|d_k\|) \\ &= -[B_k - \nabla^2 f(x^*)] d_k + o(\|d_k\|). \end{aligned}$$

By (27) and the above equality we have

$$\lim_{k\to\infty} \frac{\|g_{k+1}\|}{\|d_k\|} = 0. \tag{32}$$

From (25) and (32) it follows that

$$\begin{aligned}
\frac{\|g_{k+1}\|}{\|d_k\|} &\geq \frac{m'\|x_{k+1}-x^*\|}{\|d_k\|} \\
&= \frac{m'\|x_{k+1}-x^*\|}{\|x_{k+1}-x_k\|} \\
&\geq \frac{m'\|x_{k+1}-x^*\|}{\|x_{k+1}-x^*\|+\|x_k-x^*\|} \\
&= m'\frac{\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}}{1+\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}},
\end{aligned}$$

and thus

$$\lim_{k\to\infty} \frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|} = 0$$

which implies that $\{x_k\}$ converges to x^* superlinearly.

□

4.3 Quadratic Convergence Rate

If we take $B_k = \nabla^2 f(x_k)$ in the algorithm (A), then $(H4)$ holds. we have the following result.

Theorem 4.3. *If $(H3)$ holds, $B_k = \nabla^2 f(x_k)$ for sufficiently large k. Algorithm (A) generates an infinite sequence $\{x_k\}$. Then $\{x_k\}$ converges to x^* at least superlinearly.*

Proof. In this case, $(H4)$ holds automatically, thus the results in Theorem 4.2 hold.

Theorem 4.4. *If $(H3)$ holds, $B_k = \nabla^2 f(x_k)$ for sufficiently large k. Moreover, there exists a neighborhood $N(x^*, \epsilon) = \{x \in R^n | \ \|x - x^*\| < \epsilon\}$ of x^* with $\epsilon < \epsilon_0$ such that $\nabla^2 f(x)$ is Lipschitz continuous on $N(x^*, \epsilon)$, i.e., there exists $L(\epsilon)$ such that*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L(\epsilon)\|x - y\|, \ \forall x, y \in N(x^*, \epsilon). \tag{33}$$

Algorithm (A) generates an infinite sequence $\{x_k\}$. Then $\{x_k\}$ converges to x^* quadratically.

Proof. By Corollary 3.1, Lemmas 4.1 and 4.2, it follows that $\{x_k\}$ converges to x^* and there exists k' such that for all $k \geq k'$, $x_k \in N(x^*, \epsilon)$, $B_k = \nabla^2 f(x_k)$, and $\alpha_k = 1$. Let $\epsilon_k = x_k - x^*$. By the mean value theorem we have

$$\begin{aligned}
\epsilon_{k+1} &= x_{k+1} - x^* \\
&= x_k - x^* + d_k \\
&= \epsilon_k - \nabla^2 f(x_k)^{-1} g_k \\
&= \epsilon_k - \nabla^2 f(x_k)^{-1}(g_k - g^*) \\
&= \epsilon_k - \nabla^2 f(x_k)^{-1} \int_0^1 \nabla^2 f(x^* + t\epsilon_k)\epsilon_k dt \\
&= \nabla^2 f(x_k)^{-1}[\nabla^2 f(x_k)\epsilon_k - \int_0^1 \nabla^2 f(x^* + t\epsilon_k)\epsilon_k dt] \\
&= \nabla^2 f(x_k)^{-1} \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + t\epsilon_k)]\epsilon_k dt,
\end{aligned}$$

which and (33) imply that

$$\begin{aligned}
\|\epsilon_{k+1}\| &= \|\nabla^2 f(x_k)^{-1} \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + t\epsilon_k)]dt\epsilon_k\| \\
&\le \|\nabla^2 f(x_k)^{-1}\| \int_0^1 \|\nabla^2 f(x_k) - \nabla^2 f(x_k + t\epsilon_k)\|dt\|\epsilon_k\| \\
&\le \|\nabla^2 f(x_k)^{-1}\| \cdot L(\epsilon)\|\epsilon_k\|^2 \int_0^1 (1-t)dt \\
&= \frac{1}{2}\|\nabla^2 f(x_k)^{-1}\| \cdot L(\epsilon)\|\epsilon_k\|^2.
\end{aligned}$$

Therefore,

$$\lim_{k\to\infty} \frac{\|\epsilon_{k+1}\|}{\|\epsilon_k\|^2} \le \frac{1}{2} \lim_{k\to\infty} L(\epsilon)\|\nabla^2 f(x_k)^{-1}\| = \frac{1}{2} L(\epsilon)\|\nabla^2 f(x^*)^{-1}\|$$

which implies that $\{x_k\}$ converges to x^* quadratically.

□

5 Relationship with the Trust Region Method

The relationship between the new line search method and trust region method will be revealed in this section.

In trust region method, we need to seek a solution to the subproblem

$$\min_{p\in R^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p, \quad s.t. \ \|p\| \le \Delta_k, \tag{34}$$

where Δ_k is a trust region radius. We define $\|\cdot\|$ to be the Euclidean norm, so that the solution p_k^* of (34) is the minimizer of $m_k(p)$ in the ball with the radius Δ_k. Thus, the trust region method requires us to solve a sequence of subproblems (34) in which the objective function and constraint (which can be written as $p^T p \le \Delta_k^2$) are both quadratic.

The first issue arising in defining a trust region method is the strategy for choosing the trust region radius Δ_k at each iteration. Based this choice on the agreement between the model m_k and the objective function f at the previous iterations, we define the ratio

$$r_k = \frac{f_k - f(x_k + p_k)}{m_k(0) - m_k(p_k)}, \tag{35}$$

the numerator is called the actual reduction and the denominator is the predicted reduction. Since the step p_k is obtained by minimizing the model m_k over a region that includes the step $p = 0$, the predicted reduction will always be nonnegative. Thus, if r_k is negative, then the new objective value $f(x_k + p_k)$ is greater than the current value f_k, so the step must be rejected.

On the other hand, if r_k is close to 1, there is a good agreement between the model m_k and the function f over this step, so it is safe to expand the trust region for the next iteration. If r_k is positive but not close to 1, we do not alter the trust region, but if it is close to zero or negative, we shrink the trust region. The following algorithm describes the process.

Algorithm 5.1 (Trust Region)
Given $\overline{\Delta} > 0$, $\Delta_0 \in (0, \overline{\Delta})$, and $\mu \in [0, \frac{1}{4})$;
For $k = 0, 1, 2, ...$
  Obtain p_k by (approximately) solving (34);
  Evaluate r_k from (35);
  if $r_k < \frac{1}{4}$
    $\Delta_{k+1} = \frac{1}{4}\|p_k\|$
  else
    if $r_k > \frac{3}{4}$ and $\|p_k\| = \Delta_k$
      $\Delta_{k+1} = \min(2\Delta_k, \overline{\Delta})$
    else
      $\Delta_{k+1} = \Delta_k$;
  if $r_k > \mu$
    $x_{k+1} = x_k + p_k$
  else
    $x_{k+1} = x_k$;
end(for).

Sometimes, we need not to solve (34) exactly, we may find p_k satisfying

$$m_k(0) - m_k(p_k) \ge c_1 \|g_k\| \min(\Delta_k, \frac{\|g_k\|}{\|B_k\|}), \tag{36}$$

and

$$\|p_k\| \le \gamma \Delta_k, \tag{37}$$

for $\gamma \ge 1$ and $c_1 \in (0, 1]$.

Indeed, the exact solution p_k^* of (34) satisfies (36) and (37) ([3]).

Lemma 5.1([3]). Let $\mu = 0$ in Algorithm 5.1. Suppose that $\|B_k\| \le \beta$ for some constant β, that f is continuously differentiable and bounded below on the level set

$$L_0 = \{x \in R^n | f(x) \le f(x_1)\},$$

and that all approximate solutions of (34) satisfy the inequalities (36) and (37) for positive constants c_1 and γ. We then have

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{38}$$

Lemma 5.2([3]). Let $\mu \in (0, \frac{1}{4})$ in Algorithm 2.1. Suppose that $\|B_k\| \le \beta$ for some constant β, that f is Lipschitz continuously differentiable and bounded below on the level set L_0 and that all approximate solutions of (34) satisfy the inequalities (36) and (37) for some positive constants c_1 and γ. We then have

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{39}$$

In the new Armijo search, if we set $p_k = \alpha_k d_k$ then

$$r_k = \frac{f_k - f_{k+1}}{m_k(0) - m_k(p_k)} \ge \mu, \tag{40}$$

and $x_{k+1} = x_k + p_k$ is just an accepted iterative point in trust region method. It is obvious that (40) coincides with the accepted condition in (8).

This r_k is just the ratio of the actual reduction and the predicted reduction. If the above relation holds then the new point $\overline{x}_k = x_k + p_k$ is accepted both by the new line search method and by the trust region method. Otherwise, we should adjust step size in the new line search method or adjust the trust region radius in the trust region method. In fact, if the new point $\overline{x}_k = x_k + p_k$ is rejected, we must reduce the step size in the new line search method or reduce the trust region radius in the trust region method. From this point of view, the trust region method and line search method can be unified in a general form.

6 Conclusion Remarks

In this paper we use the Armijo line search rule in a novel way and propose a new line search method for unconstrained optimization problems. The global convergence and convergence rate of the new method are analyzed under mild conditions. Furthermore, each iterate generated by the new Armijo-type line search is shown to be an approximate solution of the subproblem of a corresponding trust region method, which reveals the relationship between line search method and trust region method in some sense.

To put it in detail, if we let $p_k = \alpha_k d_k$ in the proposed line search method then we have the accepted condition

$$r_k = -\frac{f_k - f(x_k + p_k)}{g_k^T p_k + \frac{1}{2} p_k^T B_k p_k} \ge \mu$$

which is the same condition in the trust region method; vice versa, if $x_k + p_k$ is an accepted point in the trust region method then $p_k = \alpha_k d_k$ with $\alpha_k = 1$ must satisfy the new Armijo search with $\mu \in (0, 1/4)$. This implies that the new line search method possess the advantage of the trust region method in some sense.

References

[1] L. Armijo, Minimization of fuctions having Lipschits continuous partial derivatives, Pacific Journal of Mathematics, 16(1966) 1-3.

[2] J. Nocedal, Theory of algorithms for unconstrained optimization, Acta Numer. 1 (1992) 199-242.

[3] J. Nocedal and J. S. Wright, Numerical Optimization, Springer-Verlag New York, Inc.(1999).

[4] F.A. Potra, Y. Shi, Efficient line search algorithm for unconstrained optimization, J. Optim. Theory Appl. 85 (3) (1995) 677-704.

[5] M.J.D. Powell, Direct search algorithms for optimization calculations, Acta Numer. 7 (1998) 287-336.

[6] G.A. Schultz, R.B. Schnabel and R.H. Byrd, A family of trust-region-based algorithms for unconstrained minimization with strong global convergence, SIAM J. Namer. Anal. 22(1985) 47-67.

[7] Z.J. Shi, Convergence of line search methods for unconstrained optimization, Applied Mathematics and Computation 157 (2004) 393-405.

[8] Z.J. Shi and J. Shen, A gradient-related algorithm with inexact line searches, Journal of Computational and Applied Mathematics 170 (2004) 349-370.

[9] Y. Yuan, On the convergence of trust region algorithms, Mathematics Numerica Sinica 16(1996) 333-346.

[10] Y. Yuan and W. Sun, Optimization Theory and Methods, Science Publish House of China, (1997).

[11] Y. Yuan, Numerical Methods for Nonlinear Programming, Shanghai Scientific & Technical Publishers, 1993.

[12] X.S. Zhang, J.L. Zhang and L.Z. Liao, An adaptive trust region method and its convergence, Science in China 45(2002), 620-631.