Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Applied Mathematics and Computation 157 (2004) 393–405
APPLIED MATHEMATICS AND COMPUTATION
www.elsevier.com/locate/amc

Convergence of line search methods for unconstrained optimization ☆

Zhen-Jun Shi *

College of Operations Research and Management, Qufu Normal University, Rizhao Campus, Rizhao, Shandong 276826, PR China
Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100080, China

Abstract

Line search methods are traditional and successful methods for solving unconstrained optimization problems. Its convergence has attracted more attention in recent years. In this paper we analyze the general results on convergence of line search methods with seven line search rules. It is clarified that the search direction plays a main role in these methods and that step-size guarantees the global convergence in some cases. It is also proved that many line search methods have same convergence property. These convergence results can enable us to design powerful, effective, and stable algorithms in practice. Finally, a class of special line search methods is investigated.

Keywords: Unconstrained minimization; Line search method; Convergence

1. Introduction

Unconstrained minimization problem is stated as

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

where R^n is an n-dimensional Euclidean space, $f : R^n \to R^1$ a continuously differentiable function. Line search methods for solving (1) take the form

☆ The work is supported by Natural Science Foundation of China (grant: 10171054).
* Address: College of Operations Research and Management, Qufu Normal University, Rizhao Campus, Rizhao, Shandong 276826, PR China.
E-mail address: zjshi@qrnu.edu.cn (Z.-J. Shi).

0096-3003/$ - see front matter
doi:10.1016/j.amc.2003.08.058

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

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

Line search methods are traditional and efficient methods for solving unconstrained minimization problems. Its convergence has attracted more attention in recent years (e.g., [2,8,14,17,19,25,26], etc.). Some new line search methods have been proposed year by year (e.g., [5,9,13,14,27], etc.).

The search direction d_k is generally required to satisfy the descent condition

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

which guarantees that d_k is a descent direction of $f(x)$ at x_k (e.g., [12,15,24], etc.). It is well known that

(1) $d_k = -g_k$ corresponds to steepest descent method or Cauchy method (e.g., [4,24], etc.);
(2) $d_k = (\nabla^2 f(x_k))^{-1} g_k$ corresponds to Newton method with $(\nabla^2 f(x_k))^{-1}$ being available (e.g., [3,7,16], etc.);
(3) $d_k = -H_k g_k$ corresponds to quasi-Newton or variable metric method with H_k an $n \times n$ matrix satisfying some quasi-Newton conditions (e.g., [4,16]);
(4)
$$d_k = \begin{cases} -g_k, & \text{if } k = 1; \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geqslant 2, \end{cases} \tag{4}$$
corresponds to conjugate gradient method with β_k a parameter (e.g., [6,10,11,16,23], etc.).

There are several line search rules for choosing step-size α_k ([1,4,13,14], etc.), for example, exact minimization rule, Armijo rule, Goldstein rule, Wolfe rule, etc.

In this paper we analyze the general results on convergence of line search methods with seven line search rules. It is clarified that search direction plays a main role in the algorithm and that step-size guarantees the global convergence in some cases. These convergence results can make us design powerful, effective, and stable algorithms in practice. Finally, a class of line search methods is investigated.

The remainder of the paper is organized as follows. In Section 2, we describe diverse line search methods and its properties. In Section 3, we analyze the global convergence properties. In Section 4, we analyze the convergence rate of line search methods, and in Section 5, we study further the convergence property of special line search methods.

2. Line search methods

2.1. Line search rules

We first assume that

(H_1): The function f has lower bound on $L_0 = \{x \in R^n \mid f(x) \leqslant f(x_0)\}$, where $x_0 \in R^n$ is available.

(H_2): The gradient g is Lipschitz continuous in an open convex set B which contains L_0, i.e., there exists $L > 0$ such that

$$\|g(x) - g(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in B. \tag{5}$$

There are a lot of rules for choosing step-size α_k [4], for example:

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

$$f(x_k + \alpha_k d_k) = \min_{\alpha>0} f(x_k + \alpha d_k). \tag{6}$$

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

$$\alpha_k = \min\{\alpha \mid g(x_k + \alpha d_k)^{\mathrm{T}} d_k = 0, \alpha > 0\}. \tag{7}$$

(c) *Armijo rule*. Set scalars s_k, β, and σ with $s_k = -\frac{g_k^{\mathrm{T}} d_k}{\|d_k\|^2}$, $\beta \in (0, 1)$, and $\sigma \in (0, \frac{1}{2})$, and we set $\alpha_k = \beta^{m_k} s_k$, where m_k is the first nonnegative integer m for which

$$f_k - f(x_k + \beta^m s_k d_k) \geqslant -\sigma\beta^m s_k g_k^{\mathrm{T}} d_k, \tag{8}$$

i.e., $m = 0, 1, \ldots$, are tried successively until the inequality above is satisfied for $m = m_k$.

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

$$f(x_k + \alpha_k d_k) = \min_{\alpha\in[0,s_k]} f(x_k + \alpha d_k). \tag{9}$$

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

$$\sigma \leqslant \frac{f(x_k + \alpha_k d_k) - f_k}{\alpha_k g_k^{\mathrm{T}} d_k} \leqslant 1 - \sigma. \tag{10}$$

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

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

$$f_k - f(x_k + \alpha_k d_k) \geqslant -\sigma\alpha_k g_k^{\mathrm{T}} d_k \tag{11}$$

and

$$| g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k | \leqslant -\beta g_k^{\mathrm{T}} d_k, \tag{12}$$

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

(g) *Wolfe rule*. α_k is chosen to satisfy (11) and

$$g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k \geqslant \beta g_k^{\mathrm{T}} d_k. \tag{13}$$

Lemma 2.1 [3]. *Assume that there is a scalar M such that $f(x) \geqslant M$ for all $x \in R^n$, let $\sigma \in (0, \frac{1}{2})$ and $\beta \in (\sigma, 1)$, and assume that $g_k^{\mathrm{T}} d_k < 0$. There exists an interval $[c_1, c_2]$ with $0 < c_1 < c_2$, such that every $\alpha \in [c_1, c_2]$ satisfies* (11) *and* (12) (*and hence also* (11) *and* (13)).

2.2. Line search methods

In line search methods, if the current point x_k is not a stationary point then we must choose a descent direction d_k at first and then find a new point x_{k+1} along the ray $\{x_k + \alpha d_k \mid \alpha > 0\}$.

Algorithm. Choose some parameters and select an initial point x_1, set $k := 0$;

Step 1. If $\|g_k\| = 0$ then stop! else goto step 2;
Step 2. $x_{k+1} = x_k + \alpha_k d_k$, where d_k is a descent direction of $f(x)$ at x_k, α_k is defined by line search rule (a), or (b), or (c), or (d), or (e), or (f), or (g);
Step 3. $k := k + 1$, goto step 1.

For the above seven line search rules, we denote the corresponding algorithms by Algorithm (a)–(g), respectively.

We will clarify that the search direction d_k plays a main role in line search methods and step-size α_k seems to guarantee the global convergence in some cases.

3. Convergence property

If (H_1), (H_2) and (3) hold, then for any line search rule the following convergence property holds:

$$\lim_{k\to\infty} \left(\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|} \right)^2 = 0. \tag{14}$$

Theorem 3.1. *If* (H_1), (H_2), (3) *hold*, *Algorithm* (a) *generates an infinite sequence* $\{x_k\}$, *then* (14) *holds*.

Proof. By line search rule (a), mean value theorem, Cauchy–Schwartz inequality, (H_2), set $\alpha = -\frac{g_k^T d_k}{L\|d_k\|^2}$, we have

$$\begin{aligned} f_k - f_{k+1} \geqslant f_k - f(x_k + \alpha d_k) &= -\alpha \int_0^1 g(x_k + \alpha t d_k)^T d_k \, dt \\ &= -\alpha g_k^T d_k - \alpha \int_0^1 (g(x_k + \alpha t d_k) - g_k)^T d_k \, dt \\ &\geqslant -\alpha g_k^T d_k - \alpha \int_0^1 \|g(x_k + \alpha t d_k) - g_k\| \cdot \|d_k\| \, dt \\ &\geqslant -\alpha g_k^T d_k - \frac{\alpha^2}{2} L\|d_k\|^2 = \frac{1}{2L}\left(\frac{-g_k^T d_k}{\|d_k|}\right)^2. \end{aligned}$$

By (H_1) and (3), it follows that $\{f_k\}$ is a monotone decreasing number sequence and has a bound below, thus $\{f_k\}$ has a limit, and therefore (14) holds. □

Theorem 3.2. *If* (H_1), (H_2), (3) *hold*, *Algorithm* (b) *generates an infinite sequence* $\{x_k\}$, *then* (14) *holds*.

Proof. By line search rule (b), mean value theorem, Cauchy–Schwartz inequality, (H_2), whenever $\alpha \leqslant \alpha_k$, we have

$$\begin{aligned} f_k - f_{k+1} &\geqslant f_k - f(x_k + \alpha d_k) \\ &= -\alpha \int_0^1 g(x_k + \alpha t d_k)^T d_k \, dt \\ &= -\alpha g_k^T d_k - \alpha \int_0^1 (g(x_k + \alpha t d_k) - g_k)^T d_k \, dt \\ &\geqslant -\alpha g_k^T d_k - \alpha \int_0^1 \|g(x_k + \alpha t d_k) - g_k\| \cdot \|d_k\| \, dt \\ &\geqslant -\alpha g_k^T d_k - \frac{\alpha^2}{2} L\|d_k\|^2, \end{aligned}$$

thus

$$f_k - f_{k+1} \geqslant -\alpha g_k^T d_k - \frac{\alpha^2}{2} L\|d_k\|^2, \quad \alpha \leqslant \alpha_k. \tag{15}$$

By (H_2), Cauchy–Schwartz inequality, line search rule (b), it holds that

$$\alpha_k L\|d_k\|^2 \geqslant \|g_{k+1} - g_k\| \cdot \|d_k\| \geqslant (g_{k+1} - g_k)^T d_k = -g_k^T d_k,$$

therefore

$$\alpha_k \geqslant -\frac{g_k^{\mathrm{T}} d_k}{L\|d_k\|^2}. \tag{16}$$

Set $\alpha = -\frac{g_k^{\mathrm{T}} d_k}{L\|d_k\|^2}$, by (16), $\alpha \leqslant \alpha_k$. Substituting α into (15), we have

$$f_k - f_{k+1} \geqslant \frac{1}{2L} \cdot \left(\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|} \right)^2.$$

By (H_1) and (3), we obtain that (14) holds. □

Theorem 3.3. *If* (H_1), (H_2), (3) *hold*, *Algorithm* (c) *or* (d) *generates an infinite sequence* $\{x_k\}$, *then* (14) *holds.*

Proof. For Algorithm (c), set $K_1 = \{k \mid \alpha_k = s_k\}$, $K_2 = \{k \mid \alpha_k < s_k\}$, by (8), we have

$$f_k - f_{k+1} \geqslant -s_k \sigma g_k^{\mathrm{T}} d_k, \quad \forall k \in K_1, \tag{17}$$

$$f_k - f_{k+1} \geqslant -\alpha_k \sigma g_k^{\mathrm{T}} d_k, \quad \forall k \in K_2. \tag{18}$$

By line search rule (c), since $\alpha_k/\beta \leqslant s_k$, $\forall k \in K_2$, we have

$$f_k - f(x_k + \alpha_k/\beta d_k) < -\alpha_k \sigma g_k^{\mathrm{T}} d_k/\beta, \quad \forall k \in K_2.$$

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

$$-\alpha_k g(x_k + \alpha_k \theta_k d_k/\beta)^{\mathrm{T}} d_k/\beta < -\alpha_k \sigma g_k^{\mathrm{T}} d_k/\beta, \quad \forall k \in K_2,$$

therefore

$$g(x_k + \theta_k \alpha_k d_k/\beta) > \sigma g_k^{\mathrm{T}} d_k, \quad k \in K_2. \tag{19}$$

By (H_2), Cauchy–Schwartz inequality, and (19), we have

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

thus

$$\alpha_k \geqslant -\frac{\beta(1-\sigma) g_k^{\mathrm{T}} d_k}{L\|d_k\|^2}, \quad k \in K_2. \tag{20}$$

It follows from (18) and (20) that

$$f_k - f_{k+1} > \frac{\beta\sigma(1-\sigma)}{L} \cdot \left(\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|} \right)^2, \quad k \in K_2. \tag{21}$$

By (17) and the definition of s_k, we have

$$f_k - f_{k+1} \geqslant \sigma\left(\frac{-g_k^T d_k}{\|d_k\|}\right)^2, \quad k \in K_1. \tag{22}$$

Set $\eta = \min\{\sigma, \beta\sigma(1-\sigma)/L\}$, by (21) and (22), we have

$$f_k - f_{k+1} \geqslant \eta\left(\frac{-g_k^T d_k}{|d_k\|}\right)^2. \tag{23}$$

By (H_1) and (3), it follows that (14) holds.

For Algorithm (d), it follows from line search rule (d) that

$$f_k - f_{k+1} \geqslant f_k - f(x_k + \alpha d_k), \quad \forall \alpha \in [0, s_k].$$

Assume that α_k' is the step-size in Algorithm (c), then $\alpha_k' \in [0, s_k]$, by the above inequality and (23), we have

$$f_k - f_{k+1} \geqslant f_k - f(x_k + \alpha_k' d_k) \geqslant \eta\left(\frac{-g_k^T d_k}{\|d_k\|}\right)^2,$$

thus (14) also holds similarly to the case in Algorithm (c). □

Theorem 3.4. *If* (H_1), (H_2), (3) *hold*, *Algorithm* (e) *or* (f) *or* (g) *generates an infinite sequence* $\{x_k\}$, *then* (14) *holds.*

Proof. It suffices to prove Algorithms (e) and (g) have the property (14). For Algorithm (e), by the right-hand side of (10), we have

$$f_{k+1} - f_k \geqslant \alpha_k(1-\sigma) g_k^T d_k.$$

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

$$\alpha_k g(x_k + \alpha_k\theta_k d_k)^T d_k \geqslant \alpha_k(1-\sigma) g_k^T d_k,$$

therefore

$$g(x_k + \alpha_k\theta_k d_k)^T d_k \geqslant (1-\sigma) g_k^T d_k. \tag{24}$$

By (H_2), Cauchy–Schwartz inequality, and (24), we have

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

thus

$$\alpha_k \geqslant \sigma \frac{-g_k^T d_k}{L|d_k\|^2}. \tag{25}$$

By left-hand side of (10) and (25), we obtain

$$f_k - f_{k+1} \geqslant \frac{\sigma^2}{L}\left(\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|}\right)^2. \tag{26}$$

It follows from (H_1) and (3) that (14) holds.

For Algorithm (g), by (H_2), Cauchy–Schwartz inequality and (13), we have

$$\alpha_k L\|d_k\|^2 \geqslant \|g_{k+1} - g_k\| \cdot \|d_k\| \geqslant (g_{k+1} - g_k)^{\mathrm{T}} d_k \geqslant -(1-\beta) g_k^{\mathrm{T}} d_k,$$

thus

$$\alpha_k \geqslant -\frac{(1-\beta) g_k^{\mathrm{T}} d_k}{L\|d_k\|^2}.$$

By (11) and the above inequality, we obtain

$$f_k - f_{k+1} \geqslant \frac{\sigma(1-\beta)}{L} \cdot \left(\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|}\right)^2. \tag{27}$$

By (H_1) and (3), it follows that (14) holds. □

4. Convergence rate

In Section 3, we have proved that if (H_1), (H_2) and (3) hold then (14) holds for all seven algorithms and there exists a positive number η_0 such that

$$f_k - f_{k+1} \geqslant \eta_0 \left(\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|}\right)^2, \quad \forall k. \tag{28}$$

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

We assume that

(H_3): f is twice continuously differentiable and uniformly convex on R^n.

Lemma 4.1. *Assume that* (H_3) *holds, then* (H_1), (H_2) *hold*, $f(x)$ *has a unique minimal point* x^*, *and there exist* $0 < m \leqslant M$ *such that*

$$m\|y\|^2 \leqslant y^{\mathrm{T}} \nabla^2 f(x) y \leqslant M\|y\|^2, \quad \forall x, y \in R^n; \tag{29}$$

$$\tfrac{1}{2} m\|x - x^*\|^2 \leqslant f(x) - f(x^*) \leqslant \tfrac{1}{2} M\|x - x^*\|^2, \quad \forall x \in R^n; \tag{30}$$

$$M\|x - y\|^2 \geqslant (g(x) - g(y))^{\mathrm{T}}(x - y) \geqslant m\|x - y\|^2, \quad \forall x, y \in R^n; \tag{31}$$

and thus

$$M\|x - x^*\|^2 \geqslant g(x)^{\mathrm{T}}(x - x^*) \geqslant m\|x - x^*\|^2, \quad \forall x \in R^n. \tag{32}$$

By (32), we can also obtain from Cauchy–Schwartz inequality and mean value theorem that

$$M\|x - x^*\| \geqslant \|g(x)\| \geqslant m\|x - x^*\|, \quad \forall x \in R^n. \tag{33}$$

Its proof can be seen from (e.g., [18,20], etc.).

If descent condition (3) is replaced by

$$-\frac{g_k^{\mathrm{T}} d_k}{\|g_k\| \cdot \|d_k\|} \geqslant \tau, \tag{34}$$

where $0 < \tau \leqslant 1$, then we can obtain the following conclusion.

Theorem 4.1. *Suppose that* (H_3) *and* (34) *hold, any of seven algorithms generates an infinite sequence* $\{x_k\}$, *then*

$$\lim_{k\to\infty} \|g_k\| = 0, \tag{35}$$

and $\{x_k\}$ *converges to* x^* *at least linearly.*

Proof. By (28), (34) and (33), we have

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

thus (35) holds and $\{x_k\}$ converges to x^*. By the above inequality and (30), set $\rho = \sqrt{2\frac{\eta_0 \tau^2 m^2}{M}}$ and $\theta = \sqrt{1 - \rho^2}$, we can prove that

$$\rho \leqslant 1. \tag{36}$$

In fact, it follows from (31) that

$$m \leqslant L. \tag{37}$$

For Algorithms (a) and (b), $\eta_0 = 1/(2L)$, it follows from (37) that

$$\rho^2 = 2\frac{\eta_0 \tau^2 m^2}{M} = \frac{\tau^2 m^2}{LM} \leqslant \frac{\tau^2 m}{M} \leqslant \tau^2 \leqslant 1.$$

For Algorithms (c) and (d), since $\eta_0 = \min\{\sigma, \beta\sigma(1-\sigma)/L\} \leqslant \beta\sigma(1-\sigma)/L$, by (37) we have

$$\rho^2 = 2\frac{\eta_0\tau^2 m^2}{M} = \frac{2\sigma\beta(1-\sigma)\tau^2 m^2}{LM} \leqslant \frac{\beta(1-\sigma)\tau^2 m}{M} \leqslant \tau^2\beta(1-\sigma) < 1.$$

For Algorithm (e), $\eta_0 = \sigma^2/L$, by (37) we obtain

$$\rho^2 = 2\frac{\eta_0\tau^2 m^2}{M} = \frac{2\sigma^2\tau^2 m^2}{LM} \leqslant \frac{\sigma\tau^2 m}{M} \leqslant \tau^2\sigma < 1.$$

For Algorithms (f) and (g), because $\eta_0 = \sigma(1-\beta)/L$, we obtain from (37) that

$$\rho^2 = 2\frac{\eta_0\tau^2 m^2}{M} = \frac{2\sigma(1-\beta)\tau^2 m^2}{LM} \leqslant \frac{\tau^2(1-\beta)m}{M} \leqslant \tau^2(1-\beta) < 1.$$

This shows that (36) holds, we have

$$f_k - f(x^*) \leqslant (1-\rho^2)(f_{k-1} - f(x^*) = \theta^2(f_{k-1} - f(x^*)) \leqslant \cdots \leqslant \theta^{2k}(f_0 - f(x^*)),$$

thus by (30) and set $\omega = \sqrt{\frac{2(f_0 - f(x^*))}{m}}$, it follows that

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

thus

$$\|x_k - x^*\| \leqslant \omega\theta^k$$

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

5. Further discussion

Definition 5.1 [3]. Let $\{x_k\}$ be a sequence generated by a gradient method $x_{k+1} = x_k + \alpha_k d_k$. We say that the sequence $\{d_k\}$ is uniformly gradient related to $\{x_k\}$ if for every convergent subsequence $\{x_k\}_K$ for which

$$\lim_{k\in K, k\to\infty} g_k \neq 0 \tag{38}$$

there holds

$$0 < \liminf_{k\in K, k\to\infty} |g_k^{\mathrm{T}} d_k|, \qquad \limsup_{k\in K, k\to\infty} |d_k| < \infty. \tag{39}$$

Lemma 5.1 [3]. *Assume that $f(x)$ is bounded below. Let $\{x_k\}$ be a sequence generated by a gradient method $x_{k+1} = x_k + \alpha_k d_k$ and assume that $\{d_k\}$ is uniformly gradient related and α_k is chosen by the line rule* (a), *or* (b), *or* (c), *or* (d), *or* (e), *or* (f), *or* (g). *Then*

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

In fact, it is easy to obtain this property from (14). Moreover, if d_k satisfies

$$-g_k^{\mathrm{T}} d_k \geqslant c_1 \|g_k\|^2, \tag{41}$$

and

$$\|d_k\| \leqslant c_2 \max_{0\leqslant j\leqslant k} \|g_j\|, \tag{42}$$

where $c_1 > 0$, $c_2 > 0$, then we can prove that

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

Some useful line search methods satisfy the condition (41) and (42) (see, [21,22]).

Theorem 5.1. *Assume that* (H_1), (H_2), (41), *and* (42) *hold, any one of the seven algorithms generates an infinite sequence* $\{x_k\}$, *then*

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

Proof. It suffices to prove that $\{\|g_k\|\}$ has a bound. By (42) $\{\|d_k\|\}$ has a bound, noting (41) we can obtain that $\{d_k\}$ is uniformly gradient related to $\{x_k\}$. By Lemma 5.1, the proof is completed.

In what follows we will prove that $\{\|g_k\|\}$ has a bound. Let

$$\delta_k = \max_{1\leqslant j\leqslant k} \{\|g_j\|\},$$

if $\{\|g_k\|\}$ has no bound, then there exists an infinite subset N of $\{0, 1, 2, \ldots\}$ such that

$$\|g_k\| = \delta_k, \tag{43}$$

and

$$\delta_k \to +\infty \quad (k \in N, k \to +\infty). \tag{44}$$

By (28) and (H_1) we have

$$\sum_{k=0}^{\infty} \left(\frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|} \right)^2 < +\infty. \tag{45}$$

By (41), (42), and (45), we obtain

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\delta_k^2} < +\infty. \tag{46}$$

By (43) and (46) we have

$$\sum_{k\in N}\delta_k^2=\sum_{k\in N}\frac{\delta_k^4}{\delta_k^2}=\sum_{k\in N}\frac{\|g_k\|^4}{\delta_k^2}\leqslant\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\delta_k^2}<+\infty.$$

This is a contradiction with (44). The contradiction shows that $\{\|g_k\|\}$ has a bound. □

6. Conclusion

In this paper we analyze the general results on convergence of line search methods with seven line search rules. It is proved that search direction plays a main rule in line search methods and that step-size guarantees the global convergence in some cases. These results can make us design powerful, effective, and stable algorithms in practice. A class of special line search methods is investigated. For further research, we should study the implementation of diverse line search methods and probe the numerical performance for practical problems.

References

[1] L. Armijo, Minimization of function having Lipschitz continuous first partial derivatives, Pac. J. Math. 16 (1966) 1–3.
[2] D.P. Bertsekas, J.N. Tsitsiklis, Gradient convergence in gradient methods with errors, SIAM J. Optim. 10 (3) (2000) 627–642.
[3] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, 1982.
[4] A.I. Cohen, Stepsize analysis for descent methods, J. Optim. Theory Appl. 33 (2) (1981) 187–205.
[5] J.P. Dussault, Convergence of implementable descent algorithms for unconstrained optimization, J. Optim. Theory Appl. 104 (3) (2000) 749–750.
[6] Y.H. Dai, J. Han, G. Liu, D. Sun, H. Yin, Y.-X. Yuan, Convergence properties of nonlinear conjugate gradient methods, SIAM J. Optim. 10 (2) (1999) 345–358.
[7] Y.G. Evtushenko, Numerical Optimization Techniques, Optimization Software, New York, 1985.
[8] M. Gaviano, D. Lera, On linear convergence of gradient-type minimization algorithms, J. Optim. Theory Appl. 98 (2) (1998) 475–487.
[9] T.N. Grapsa, M.N. Vrahatis, A dimensional-reducing method for unconstrained optimization, J. Comput. Appl. Math. 66 (1996) 239–253.
[10] M.R. Hestenes, Conjugate Direction Methods in Optimization, Springer-Verlag, New York, 1980.
[11] G.H. Liu, L.L. Jing, L.X. Han, D. Han, A class of nonmonotone conjuagte gradient methods for unconstrained optimization, J. Optim. Theory Appl. 101 (1) (1999) 127–140.
[12] C. Lemarechal, R. Mifflin, Global and superlinear convergence of an algorithm for one-dimensional minimization of convex functions, Math. Program. 24 (1982) 241–256.

[13] M.J.D. Powell, Direct search algorithms for optimization calculations, Acta Numer. 7 (1998) 287–336.
[14] J.M. Moguerza, F.J. Prieto, Combining search directions using gradient flows, Math. Program. 96 (3) (2003) 529–559.
[15] R.H. Mladineo, Convergence rate of a global optimization algorithm, Math. Program. 54 (1992) 223–232.
[16] J. Nocedal, J.W. Stephen, Numerical Optimization, Springer-Verlag, New York, 1999.
[17] J. Nocedal, Theory of algorithms for unconstrained optimization, Acta Numer. 1 (1992) 199–242.
[18] J.M. Ortega, W.C. Rheinboldt, Iterative Solution Of Nonlinear Equation in Several Variables, Academic Press, 1970.
[19] F.A. Potra, Y. Shi, Efficient line search algorithm for unconstrained optimization, J. Optim. Theory Appl. 85 (3) (1995) 677–704.
[20] R.T. Rockafellar, Convex Analysis, Princeton University Press, 1970.
[21] Z.J. Shi, A new memory gradient method under exact line search, Asian Pac. J. Oper. Res. 20 (2) (2003).
[22] Z.J. Shi, The supermemory gradient method with inexact line searches, J. Appl. Sci. 21 (3) (2003).
[23] S. Sanmatlas, E. Vercher, A generalized conjugate gradient algorithm, J. Optim. Theory Appl. 98 (2) (1998) 489–502.
[24] M.J. Todd, On convergence properties of algorithms for unconstrained minimization, IMA J. Numer. Anal. 9 (1989) 435–441.
[25] Z. Wei, L. Qi, H. Jiang, Some convergence properties of descent methods, J. Optim. Theory Appl. 95 (1) (1997) 177–188.
[26] M.N. Vrahatis, G.S. Androulakis, J.N. Lambrino, G.D. Magoulas, A class of gradient unconstrained minimization algorithms with adaptive stepsize, J. Comput. Appl. Math. 114 (2000) 367–386.
[27] M.N. Vrahatis, G.S. Androulakis, G.E. Manoussakis, A new unconstrained optimization method for imprecise function and gradient values, J. Math. Anal. Appl. 197 (1996) 586–607.