Chapter 2

Line Search

2.1 Introduction

Line search, also called one-dimensional search, refers to an optimization procedure for univariable functions. It is the base of multivariable optimization. As stated before, in multivariable optimization algorithms, for given x_k, the iterative scheme is

$$x_{k+1} = x_k + \alpha_k d_k. \tag{2.1.1}$$

The key is to find the direction vector d_k and a suitable step size α_k. Let

$$\phi(\alpha) = f(x_k + \alpha d_k). \tag{2.1.2}$$

So, the problem that departs from x_k and finds a step size in the direction d_k such that

$$\phi(\alpha_k) < \phi(0)$$

is just line search about α.

If we find α_k such that the objective function in the direction d_k is minimized, i.e.,

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

or

$$\phi(\alpha_k) = \min_{\alpha>0} \phi(\alpha),$$

such a line search is called exact line search or optimal line search, and α_k is called optimal step size. If we choose α_k such that the objective function has acceptable descent amount, i.e., such that the descent $f(x_k) - f(x_k + \alpha_k d_k) >$

0 is acceptable by users, such a line search is called inexact line search, or approximate line search, or acceptable line search.

Since, in practical computation, theoretically exact optimal step size generally cannot be found, and it is also expensive to find almost exact step size, therefore the inexact line search with less computation load is highly popular.

The framework of line search is as follows. First, determine or give an initial search interval which contains the minimizer; then employ some section techniques or interpolations to reduce the interval iteratively until the length of the interval is less than some given tolerance.

Next, we give a notation about the search interval and a simple method to determine the initial search interval.

Definition 2.1.1 *Let* $\phi : R \to R, \alpha^* \in [0, +\infty)$, *and*

$$\phi(\alpha^*) = \min_{\alpha \geq 0} \phi(\alpha).$$

If there exists a closed interval $[a, b] \subset [0, +\infty)$ *such that* $\alpha^* \in [a, b]$, *then* $[a, b]$ *is called a search interval for one-dimensional minimization problem* $\min_{\alpha \geq 0} \phi(\alpha)$. *Since the exact location of the minimum of* ϕ *over* $[a, b]$ *is not known, this interval is also called the interval of uncertainty.*

A simple method to determine an initial interval is called the forward-backward method. The basic idea of this method is as follows. Given an initial point and an initial steplength, we attempt to determine three points at which their function values show "high–low–high" geometry. If it is not successful to go forward, we will go backward. Concretely, given an initial point α_0 and a steplength $h_0 > 0$. If

$$\phi(\alpha_0 + h_0) < \phi(\alpha_0),$$

then, next step, depart from $\alpha_0 + h_0$ and continue going forward with a larger step until the objective function increases. If

$$\phi(\alpha_0 + h_0) > \phi(\alpha_0),$$

then, next step, depart from α_0 and go backward until the objective function increases. So, we will obtain an initial interval which contains the minimum α^*.

Algorithm 2.1.2 *(Forward-Backward Method)*

Step 1. *Given* $\alpha_0 \in [0, \infty), h_0 > 0$*, the multiple coefficient* $t > 1$ *(Usually* $t = 2$*). Evaluate* $\phi(\alpha_0), k := 0$*.*

Step 2. *Compare the objective function values. Set* $\alpha_{k+1} = \alpha_k + h_k$ *and evaluate* $\phi_{k+1} = \phi(\alpha_{k+1})$*. If* $\phi_{k+1} < \phi_k$*, go to Step 3; otherwise, go to Step 4.*

Step 3. *Forward step. Set* $h_{k+1} := th_k, \alpha := \alpha_k, \alpha_k := \alpha_{k+1}, \phi_k := \phi_{k+1}, k := k + 1$*, go to Step 2.*

Step 4. *Backward step. If* $k = 0$*, invert the search direction. Set* $h_k := -h_k, \alpha_k := \alpha_{k+1}$*, go to Step 2; otherwise, set*

$$a = \min\{\alpha, \alpha_{k+1}\},\ b = \max\{\alpha, \alpha_{k+1}\},$$

output $[a, b]$ *and stop.* $\Box$

The methods of line search presented in this chapter use the unimodality of the function and interval. The following definitions and theorem introduce their concepts and properties.

Definition 2.1.3 *Let* $\phi : R \to R, [a, b] \subset R$*. If there is* $\alpha^* \in [a, b]$ *such that* $\phi(\alpha)$ *is strictly decreasing on* $[a, \alpha^*]$ *and strictly increasing on* $[\alpha^*, b]$*, then* $\phi(\alpha)$ *is called a unimodal function on* $[a, b]$*. Such an interval* $[a, b]$ *is called a unimodal interval related to* $\phi(\alpha)$*.*

The unimodal function can also be defined as follows.

Definition 2.1.4 *If there exists a unique* $\alpha^* \in [a, b]$*, such that for any* $\alpha_1, \alpha_2 \in [a, b], \alpha_1 < \alpha_2$*, the following statements hold:*

if $\alpha_2 < \alpha^*$*, then* $\phi(\alpha_1) > \phi(\alpha_2)$*;*

if $\alpha_1 > \alpha^*$*, then* $\phi(\alpha_1) < \phi(\alpha_2)$*;*

then $\phi(\alpha)$ *is the unimodal function on* $[a, b]$*.*

Note that, first, the unimodal function does not require the continuity and differentiability of the function; second, using the property of the unimodal function, we can exclude portions of the interval of uncertainty that do not contain the minimum, such that the interval of uncertainty is reduced. The following theorem shows that if the function ϕ is unimodal on $[a, b]$, then the interval of uncertainy could be reduced by comparing the function values of ϕ at two points within the interval.

Theorem 2.1.5 *Let $\phi : R \to R$ be unimodal on $[a, b]$. Let $\alpha_1, \alpha_2 \in [a, b]$, and $\alpha_1 < \alpha_2$. Then*

1. *if $\phi(\alpha_1) \le \phi(\alpha_2)$, then $[a, \alpha_2]$ is a unimodal interval related to ϕ;*

2. *if $\phi(\alpha_1) \ge \phi(\alpha_2)$, then $[\alpha_1, b]$ is a unimodal interval related to ϕ.*

Proof. From the Definition 2.1.3, there exists $\alpha^* \in [a, b]$ such that $\phi(\alpha)$ is strictly decreasing over $[a, \alpha^*]$ and strictly increasing over $[\alpha^*, b]$. Since $\phi(\alpha_1) \le \phi(\alpha_2)$, then $\alpha^* \in [a, \alpha_2]$ (see Figure 2.1.1). Since $\phi(\alpha)$ is unimodal on $[a, b]$, it is also unimodal on $[a, \alpha_2]$. Therefore $[a, \alpha_2]$ is a unimodal interval related to $\phi(\alpha)$ and the proof of the first part is complete.

The second part of the theorem can be proved similarly. □

This theorem indicates that, for reducing the interval of uncertainty, we must at least select two observations, evaluate and compare their function values.

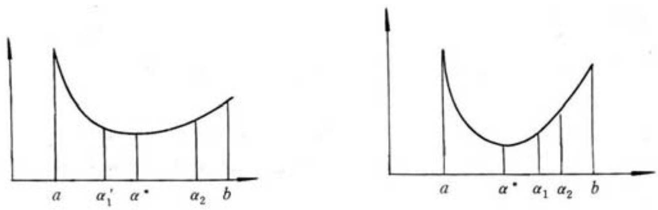


Figure 2.1.1 Properties of unimodal interval and unimodal function

2.2 Convergence Theory for Exact Line Search

The general form of an unconstrained optimization algorithm is as follows.

Algorithm 2.2.1 *(General Form of Unconstrained Optimization)*

Initial Step: Given $x_0 \in R^n, 0 \le \epsilon \ll 1$.

k-th Step: *Compute the descent direction d_k;*
Compute the step size α_k, such that

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k); \tag{2.2.1}$$

Set

$$x_{k+1} = x_k + \alpha_k d_k; \tag{2.2.2}$$

If $\|\nabla f(x_{k+1})\| \leq \epsilon$, stop; otherwise, repeat the above steps.
$\square$

Set

$$\phi(\alpha) = f(x_k + \alpha d_k), \tag{2.2.3}$$

obviously we have from the algorithm that

$$\phi(0) = f(x_k), \ \phi(\alpha) \leq \phi(0).$$

In fact, (2.2.1) is to find the global minimizer of $\phi(\alpha)$ which is rather difficult. Instead, we look for the first stationary point, i.e., take α_k such that

$$\alpha_k = \min\{\alpha \geq 0 \mid \nabla f(x_k + \alpha d_k)^T d_k = 0\}. \tag{2.2.4}$$

Since, by (2.2.1) and (2.2.4), we find the exact minimizer and the stationary point of $\phi(\alpha)$ respectively, we say that (2.2.1) and (2.2.4) are exact line searches.

Let $\langle d_k, -\nabla f(x_k)\rangle$ denote the angle between d_k and $-\nabla f(x_k)$, we have

$$\cos\langle d_k, -\nabla f(x_k)\rangle = -\frac{d_k^T \nabla f(x_k)}{\|d_k\| \|\nabla f(x_k)\|}. \tag{2.2.5}$$

The following theorem gives a bound of descent in function values for each iteration in exact line search.

Theorem 2.2.2 *Let $\alpha_k > 0$ be the solution of (2.2.1). Let $\|\nabla^2 f(x_k + \alpha d_k)\| \leq M$ $\forall \alpha > 0$, where M is some positive number. Then*

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \frac{1}{2M} \|\nabla f(x_k)\|^2 \cos^2\langle d_k, -\nabla f(x_k)\rangle. \tag{2.2.6}$$

Proof. From the assumptions we have that

$$f(x_k + \alpha d_k) \leq f(x_k) + \alpha d_k^T \nabla f(x_k) + \frac{\alpha^2}{2} M \|d_k\|^2, \ \forall \alpha > 0. \tag{2.2.7}$$

Set $\bar{\alpha} = -d_k^T \nabla f(x_k)/(M\|d_k\|^2)$; it follows from the assumptions, (2.2.7) and (2.2.5) that

$$\begin{aligned}
f(x_k) - f(x_k + \alpha_k d_k) &\geq f(x_k) - f(x_k + \bar{\alpha} d_k) \\
&\geq -\bar{\alpha} d_k^T \nabla f(x_k) - \frac{\bar{\alpha}^2}{2} M \|d_k\|^2 \\
&= \frac{1}{2} \frac{(d_k^T \nabla f(x_k))^2}{M\|d_k\|^2} \\
&= \frac{1}{2M} \|\nabla f(x_k)\|^2 \frac{(d_k^T \nabla f(x_k))^2}{\|d_k\|^2 \|\nabla f(x_k)\|^2} \\
&= \frac{1}{2M} \|\nabla f(x_k)\|^2 \cos^2 \langle d_k, -\nabla f(x_k) \rangle . \square
\end{aligned}$$

Now we are in position to state the convergence property of general unconstrained optimization algorithms with exact line search. The following two theorems state the convergence by different forms.

Theorem 2.2.3 *Let $f(x)$ be a continuously differentiable function on an open set $D \subset R^n$, assume that the sequence from Algorithm 2.2.1 satisfies $f(x_{k+1}) \leq f(x_k) \forall k$ and $\nabla f(x_k)^T d_k \leq 0$. Let $\bar{x} \in D$ be an accumulation point of $\{x_k\}$ and K_1 be an index set with $K_1 = \{k \mid \lim_{k\to\infty} x_k = \bar{x}\}$. Also assume that there exists $M > 0$ such that $\|d_k\| < M, \forall k \in K_1$. Then, if $\bar{d}$ is any accumulation point of $\{d_k\}$, we have*

$$\nabla f(\bar{x})^T \bar{d} = 0. \tag{2.2.8}$$

Furthermore, if $f(x)$ is twice continuously differentiable on D, then

$$\bar{d}^T \nabla^2 f(\bar{x}) \bar{d} \geq 0. \tag{2.2.9}$$

Proof. It is enough to prove (2.2.8) because the proof of (2.2.9) is similar.

Let $K_2 \subset K_1$ be an index set with $\bar{d} = \lim_{k\in K_2} d_k$. If $\bar{d} = 0$, (2.2.8) is trivial. Otherwise, we consider the following two cases.

(i) There exists an index set $K_3 \subset K_2$ such that $\lim_{k\in K_3} \alpha_k = 0$. Since α_k is an exact step size, then $\nabla f(x_k+\alpha_k d_k)^T d_k = 0$. Since $\|d_k\|$ is uniformly bounded above and $\alpha_k \to 0$, taking the limit yields

$$\nabla f(\bar{x})^T \bar{d} = 0.$$

(ii) Case of $\liminf_{k\in K_2} \alpha_k = \bar{\alpha} > 0$. Let $K_4 \subset K_2$ be an index set of k with $\alpha_k \geq \bar{\alpha}/2, \forall k \in K_4$. Now assume that the conclusion (2.2.8) is not true, then we have

$$\nabla f(\bar{x})^T \bar{d} < -\delta < 0.$$

So, there exist a neighborhood $N(\bar{x})$ of $\bar{x}$ and an index set $K_5 \subset K_4$ such that when $x \in N(\bar{x})$ and $k \in K_5$,

$$\nabla f(x)^T d_k \leq -\delta/2 < 0.$$

Let $\hat{\alpha}$ be a sufficiently small positive number, such that for all $0 \leq \alpha \leq \hat{\alpha}$ and all $k \in K_5$, $x_k + \alpha d_k \in N(\bar{x})$. Take $\alpha^* = \min(\bar{\alpha}/2, \hat{\alpha})$, then from the non-increasing property of the algorithm, exact line search and Taylor's expansion, we have

$$\begin{aligned}
f(\bar{x}) - f(x_0) &= \sum_{k=0}^{\infty}[f(x_{k+1}) - f(x_k)] \\
&\leq \sum_{k\in K_5} [f(x_{k+1}) - f(x_k)] \\
&\leq \sum_{k\in K_5} [f(x_k + \alpha^* d_k) - f(x_k)] && (2.2.10) \\
&= \sum_{k\in K_5} \nabla f(x_k + \tau_k d_k)^T \alpha^* d_k && (2.2.11) \\
&\leq \sum_{k\in K_5} -\left(\frac{\delta}{2}\right) \alpha^* \\
&= -\infty,
\end{aligned}$$

where $0 \leq \tau_k \leq \alpha^*$. The above contradiction shows that (2.2.8) also holds for case (ii).

The proof of (2.2.9) is similar. It is enough to note using the second-order form of the Taylor expansion instead of the first-order form in (2.2.11). In fact, from (2.2.10) we have

$$f(\bar{x}) - f(x_0)$$

$$\begin{aligned}
&\leq \sum_{k\in K_5}[f(x_k+\alpha^* d_k)-f(x_k)] \\
&= \sum_{k\in K_5}\left[\nabla f(x_k)^T(\alpha^* d_k)+\frac{(\alpha^*)^2}{2}d_k^T\nabla^2 f(x_k+\tau_k d_k)d_k\right] \text{ for } 0\leq\tau_k\leq\alpha^* \\
&\leq \sum_{k\in K_5}\frac{(\alpha^*)^2}{2}d_k^T\nabla^2 f(x_k+\tau_k d_k)d_k \text{ for } 0\leq\tau_k\leq\alpha^* \\
&\leq \sum_{k\in K_5}\left[-\frac{1}{2}\left(\frac{\delta}{2}\right)(\alpha^*)^2\right] \\
&= -\infty. \qquad (2.2.12)
\end{aligned}$$

We also get a contradiction which proves (2.2.9). □

Theorem 2.2.4 *Let $\nabla f(x)$ be uniformly continuous on the level set $L = \{x \in R^n \mid f(x) \leq f(x_0)\}$. Let also the angle θ_k between $-\nabla f(x_k)$ and the direction d_k generated by Algorithm 2.2.1 is uniformly bounded away from $90°$, i.e., satisfies*

$$\theta_k \leq \frac{\pi}{2}-\mu, \text{ for some } \mu>0. \qquad (2.2.13)$$

Then $\nabla f(x_k)=0$ for some k; or $f(x_k)\to-\infty$; or $\nabla f(x_k)\to 0$.

Proof. Assume that, for all k, $\nabla f(x_k)\neq 0$ and $f(x_k)$ is bounded below. Since $\{f(x_k)\}$ is monotonic descent, its limit exists. Therefore

$$f(x_k)-f(x_{k+1})\to 0. \qquad (2.2.14)$$

Assume, by contradiction, that $\nabla f(x_k)\to 0$ does not hold. Then there exists $\epsilon>0$ and a subset K, such that $\|\nabla f(x_k)\|\geq\epsilon\forall k\in K$. Therefore

$$-\nabla f(x_k)^T d_k/\|d_k\| = \|\nabla f(x_k)\|\cos\theta_k \geq \epsilon\sin\mu \triangleq \epsilon_1. \qquad (2.2.15)$$

Note that

$$\begin{aligned}
& f(x_k+\alpha d_k) \\
=\ & f(x_k)+\alpha\nabla f(\xi_k)^T d_k \\
=\ & f(x_k)+\alpha\nabla f(x_k)^T d_k+\alpha[\nabla f(\xi_k)-\nabla f(x_k)]^T d_k \\
\leq\ & f(x_k)+\alpha\|d_k\|\left(\frac{\nabla f(x_k)^T d_k}{\|d_k\|}+\|\nabla f(\xi_k)-\nabla f(x_k)\|\right), \qquad (2.2.16)
\end{aligned}$$

where ξ_k lies between x_k and $x_k + \alpha d_k$. Since $\nabla f(x)$ is uniformly continuous on the level set L, there exists $\bar{\alpha}$ such that when $0 \leq \alpha \|d_k\| \leq \bar{\alpha}$, we have

$$\|\nabla f(\xi_k) - \nabla f(x_k)\| \leq \frac{1}{2}\epsilon_1. \tag{2.2.17}$$

By (2.2.15)–(2.2.17), we get

$$\begin{aligned} f\left(x_k + \bar{\alpha}\frac{d_k}{\|d_k\|}\right) &\leq f(x_k) + \bar{\alpha}\left(\frac{\nabla f(x_k)^T d_k}{\|d_k\|} + \frac{1}{2}\epsilon_1\right) \\ &\leq f(x_k) - \frac{1}{2}\bar{\alpha}\epsilon_1. \end{aligned}$$

Therefore

$$f(x_{k+1}) \leq f\left(x_k + \bar{\alpha}\frac{d_k}{\|d_k\|}\right) \leq f(x_k) - \frac{1}{2}\bar{\alpha}\epsilon_1,$$

which contradicts (2.2.14). The contradiction shows that $\nabla f(x_k) \to 0$. We complete this proof. $\square$

In the remainder of this section, we discuss the convergence rate of minimization algorithms with exact line search. For convenience of the proof of the theorem, we first give some lemmas.

Lemma 2.2.5 *Let $\phi(\alpha)$ be twice continuously differentiable on the closed interval $[0, b]$ and $\phi'(0) < 0$. If the minimizer $\alpha^* \in (0, b)$ of $\phi(\alpha)$ on $[0, b]$, then*

$$\alpha^* \geq \tilde{\alpha} = -\phi'(0)/M, \tag{2.2.18}$$

where M is a positive number such that $\phi''(\alpha) \leq M, \forall \alpha \in [0, b]$.

Proof. Construct the auxiliary function

$$\psi(\alpha) = \phi'(0) + M\alpha,$$

which has the unique zero

$$\tilde{\alpha} = -\phi'(0)/M.$$

Noting that $\phi''(\alpha) \leq M$, it follows that

$$\phi'(\alpha) = \phi'(0) + \int_0^\alpha \phi''(\alpha)\mathrm{d}\,\alpha \leq \phi'(0) + \int_0^\alpha M\mathrm{d}\,\alpha = \psi(\alpha).$$

Setting $\alpha = \alpha^*$ in the above inequality and noting that $\phi'(\alpha^*) = 0$, we obtain

$$0 \le \psi(\alpha^*) = \phi'(0) + M\alpha^*$$

which is (2.2.18). $\square$

Lemma 2.2.6 *Let $f(x)$ be twice continuously differentiable on R^n. Then for any vector $x, d \in R^n$ and any number α, the equality*

$$f(x+\alpha d) = f(x) + \alpha \nabla f(x)^T d + \alpha^2 \int_0^1 (1-t)[d^T \nabla^2 f(x+t\alpha d)d]\,dt \quad (2.2.19)$$

holds.

Proof. From calculus, we have

$$\begin{aligned}
& f(x+\alpha d) - f(x) \\
= & \int_0^1 \mathrm{d} f(x+t\alpha d) \\
= & -\int_0^1 [\alpha \nabla f(x+t\alpha d)^T d] \mathrm{d}(1-t) \\
= & -[(1-t)\alpha \nabla f(x+t\alpha d)^T d]_0^1 + \int_0^1 (1-t)\mathrm{d}[\alpha \nabla f(x+t\alpha d)^T d] \\
= & \alpha \nabla f(x)^T d + \alpha^2 \int_0^1 [(1-t) d^T \nabla^2 f(x+t\alpha d)d]\mathrm{d}t. \quad \square
\end{aligned}$$

Lemma 2.2.7 *Let $f(x)$ be twice continuously differentiable in the neighborhood of the minimizer x^*. Assume that there exist $\epsilon > 0$ and $M > m > 0$, such that*

$$m\|y\|^2 \le y^T \nabla^2 f(x) y \le M\|y\|^2, \ \forall y \in R^n \quad (2.2.20)$$

holds when $\|x - x^*\| < \epsilon$. Then we have

$$\frac{1}{2} m\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{1}{2} M\|x - x^*\|^2 \quad (2.2.21)$$

and

$$\|\nabla f(x)\| \ge m\|x - x^*\|. \quad (2.2.22)$$

Proof. From Lemma 2.2.6 we have

$$\begin{aligned}
& f(x)-f(x^*) \\
= \;& \nabla f(x^*)^T(x-x^*)+\int_0^1(1-t)(x-x^*)^T\nabla^2 f(tx+(1-t)x^*)(x-x^*)\mathrm{d}t \\
= \;& \int_0^1(1-t)(x-x^*)^T\nabla^2 f(tx+(1-t)x^*)(x-x^*)\mathrm{d}t. \qquad (2.2.23)
\end{aligned}$$

Note that (2.2.20) and the integral mean-value theorem give

$$\begin{aligned}
& m\|x-x^*\|^2\int_0^1(1-t)\mathrm{d}t \\
\le \;& \int_0^1(1-t)(x-x^*)^T\nabla^2 f(tx+(1-t)x^*)(x-x^*)\mathrm{d}t \\
\le \;& M\|x-x^*\|^2\int_0^1(1-t)\mathrm{d}t. \qquad (2.2.24)
\end{aligned}$$

Then combining (2.2.23) and (2.2.24) yields (2.2.21).

Also, using Taylor expansion gives

$$\nabla f(x)=\nabla f(x)-\nabla f(x^*)=\int_0^1\nabla^2 f(tx+(1-t)x^*)(x-x^*)\mathrm{d}t.$$

Then

$$\begin{aligned}
\|\nabla f(x)\|\|x-x^*\| &\ge (x-x^*)^T\nabla f(x) \\
&= \int_0^1(x-x^*)^T\nabla^2 f(tx+(1-t)x^*)(x-x^*)\mathrm{d}t \\
&\ge m\|x-x^*\|^2
\end{aligned}$$

which proves (2.2.22). □

Now we are in the position to give the theorem about convergence rate which shows that the local convergence rate of Algorithm 2.2.1 with exact line search is at least linear.

Theorem 2.2.8 *Let the sequence $\{x_k\}$ generated by Algorithm 2.2.1 converge to the minimizer x^* of $f(x)$. Let $f(x)$ be twice continuously differentiable in a neighborhood of x^*. If there exist $\epsilon>0$ and $M>m>0$ such that when $\|x-x^*\|<\epsilon$,*

$$m\|y\|^2\le y^T\nabla^2 f(x)y\le M\|y\|^2,\ \forall y\in R^n \qquad (2.2.25)$$

holds, then the sequence $\{x_k\}$, at least, converges linearly to x^*.

Proof. Let $\lim_{k\to\infty} x_k = x^*$. We may assume that $\|x_k - x^*\| \le \epsilon$ for k sufficiently large. Since $\|x_{k+1} - x^*\| < \epsilon$, there exists $\delta > 0$ such that

$$\|x_k + (\alpha_k + \delta)d_k - x^*\| = \|x_{k+1} - x^* + \delta d_k\| < \epsilon. \tag{2.2.26}$$

Note that $\phi(\alpha) = f(x_k + \alpha d_k), \phi'(\alpha) = \nabla f(x_k + \alpha d_k)^T d_k, \phi'(0) = \nabla f(x_k)^T d_k$ and $|\phi'(0)| \le \|\nabla f(x_k)\| \|d_k\|$. We have $\phi'(0) < 0$,

$$\rho\|\nabla f(x_k)\|\|d_k\| \le -\phi'(0) \le \|\nabla f(x_k)\|\|d_k\|, \text{ for some } \rho \in (0,1) \tag{2.2.27}$$

and

$$\phi''(\alpha) = d_k^T \nabla^2 f(x_k + \alpha d_k) d_k \le M\|d_k\|^2.$$

Then, by Lemma 2.2.5, we know that the minimizer α_k of $\phi(\alpha)$ on $[0, \alpha_k + \delta]$ satisfies

$$\alpha_k \ge \tilde{\alpha}_k = \frac{-\phi'(0)}{M\|d_k\|^2} \ge \frac{\rho\|\nabla f(x_k)\|}{M\|d_k\|} \triangleq \bar{\alpha}_k. \tag{2.2.28}$$

Set $\bar{x}_k = x_k + \bar{\alpha}_k d_k$. Obviously, it follows from (2.2.26) that $\|\bar{x}_k - x^*\| < \epsilon$. Therefore,

$$\begin{aligned}
& f(x_k + \alpha_k d_k) - f(x_k) \\
\le\; & f(x_k + \bar{\alpha}_k d_k) - f(x_k) \\
=\; & \bar{\alpha}_k \nabla f(x_k)^T d_k + \bar{\alpha}_k^2 \int_0^1 (1-t) d_k^T \nabla^2 f(x_k + t\bar{\alpha}_k d_k) d_k \mathrm{d}t \text{ (from Lemma 2.2.6)} \\
\le\; & \bar{\alpha}_k(-\rho)\|\nabla f(x_k)\|\|d_k\| + \frac{1}{2} M \bar{\alpha}_k^2 \|d_k\|^2 \text{ (from (2.2.25) and (2.2.27))} \\
\le\; & -\frac{\rho^2}{2M}\|\nabla f(x_k)\|^2 \;\text{ (from (2.2.28))} \\
\le\; & -\frac{\rho^2}{2M} m^2 \|x_k - x^*\|^2 \;\text{ (from (2.2.22))} \\
\le\; & -\left(\frac{\rho m}{M}\right)^2 [f(x_k) - f(x^*)] \;\text{ (from (2.2.21))}.
\end{aligned}$$

The above inequalities give

$$\begin{aligned}
f(x_{k+1}) - f(x^*) &= [f(x_{k+1}) - f(x_k)] + [f(x_k) - f(x^*)] \\
&\le \left[1 - \left(\frac{\rho m}{M}\right)^2\right] [f(x_k) - f(x^*)].
\end{aligned} \tag{2.2.29}$$

Set

$$\theta = \left[1 - \left(\frac{\rho m}{M}\right)^2\right]^{\frac{1}{2}}. \tag{2.2.30}$$

Obviously $\theta \in (0,1)$. Therefore (2.2.29) can be written as

$$\begin{aligned} f(x_k) - f(x^*) &\leq \theta^2[f(x_{k-1}) - f(x^*)] \\ &\leq \cdots \\ &\leq \theta^{2k}[f(x_0) - f(x^*)]. \end{aligned} \tag{2.2.31}$$

Furthermore, by (2.2.21), we have

$$\begin{aligned} \|x_k - x^*\|^2 &\leq \frac{2}{m}[f(x_k) - f(x^*)] \\ &\leq \frac{2}{m}\theta^2[f(x_{k-1}) - f(x^*)] \\ &\leq \frac{2}{m}\theta^2\frac{M}{2}\|x_{k-1} - x^*\|^2 \end{aligned}$$

which implies that

$$\|x_k - x^*\| \leq \sqrt{\frac{M}{m}}\theta\|x_{k-1} - x^*\| \tag{2.2.32}$$

and that the sequence $\{x_k\}$, at least, converges linearly to x^*. □

In the end of this section, we give a theorem which describes a descent bound of the function value after each exact line search.

Theorem 2.2.9 *Let α_k be an exact step size. Assume that $f(x)$ satisfies*

$$(x - z)^T[\nabla f(x) - \nabla f(z)] \geq \eta\|x - z\|^2. \tag{2.2.33}$$

Then

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \frac{1}{2}\eta\|\alpha_k d_k\|^2. \tag{2.2.34}$$

Proof. Since α_k is an exact step size, then

$$\nabla f(x_k + \alpha_k d_k)^T d_k = 0. \tag{2.2.35}$$

Therefore, it follows from the mean-value theorem, (2.2.35) and (2.2.33) that

$$\begin{aligned} f(x_k) - f(x_k + \alpha_k d_k) &= \int_0^{\alpha_k} -d_k^T \nabla f(x_k + t d_k) \mathrm{d}t \\ &= \int_0^{\alpha_k} d_k^T [\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k + t d_k)] \mathrm{d}t \\ &\geq \int_0^{\alpha_k} \eta(\alpha_k - t) \mathrm{d}t \|d_k\|^2 \\ &= \frac{1}{2}\eta \|\alpha_k d_k\|^2. \end{aligned} \tag{2.2.36}$$

This completes the proof. □

2.3 The Golden Section Method and the Fibonacci Method

The golden section method and the Fibonacci method are section methods. Their basic idea for minimizing a unimodal function over $[a, b]$ is iteratively reducing the interval of uncertainty by comparing the function values of the observations. When the length of the interval of uncertainty is reduced to some desired degree, the points on the interval can be regarded as approximations of the minimizer. Such a class of methods only needs to evaluate the functions and has wide applications, especially it is suitable to nonsmooth problems and those problems with complicated derivative expressions.

2.3.1 The Golden Section Method

Let

$$\phi(\alpha) = f(x + \alpha d)$$

be a unimodal function on the interval $[a, b]$. At the iteration k of the golden section method, let the interval of uncertainty be $[a_k, b_k]$. Take two observations $\lambda_k, \mu_k \in [a_k, b_k]$ and $\lambda_k < \mu_k$. Evaluate $\phi(\lambda_k)$ and $\phi(\mu_k)$. By Theorem 2.1.5, we have

Case 1 if $\phi(\lambda_k) \leq \phi(\mu_k)$, then set $a_{k+1} = a_k, b_{k+1} = \mu_k$;

Case 2 if $\phi(\lambda_k) > \phi(\mu_k)$, then set $a_{k+1} = \lambda_k, b_{k+1} = b_k$.

How to choose the observations λ_k and μ_k? We require that λ_k and μ_k satisfy the following conditions:

1. The distances from λ_k and μ_k to the end points of the interval $[a_k, b_k]$ are equivalent, that is,
$$b_k - \lambda_k = \mu_k - a_k. \tag{2.3.1}$$

2. The reduction rate of the intervals of uncertainty for each iteration is the same, that is
$$b_{k+1} - a_{k+1} = \tau(b_k - a_k),\ \tau \in (0, 1). \tag{2.3.2}$$

3. Only one extra observation is needed for each new iteration.

Now we consider Case 1. Substituting the values of Case 1 into (2.3.2) and combining (2.3.1) yield

$$\begin{aligned}\mu_k - a_k &= \tau(b_k - a_k),\\ b_k - \lambda_k &= \mu_k - a_k.\end{aligned}$$

Arranging the above equations gives

$$\lambda_k = a_k + (1-\tau)(b_k - a_k), \tag{2.3.3}$$
$$\mu_k = a_k + \tau(b_k - a_k). \tag{2.3.4}$$

Note that, in this case, the new interval is $[a_{k+1}, b_{k+1}] = [a_k, \mu_k]$. For further reducing the interval of uncertainty, the observations λ_{k+1} and μ_{k+1} are selected. By (2.3.4),

$$\begin{aligned}\mu_{k+1} &= a_{k+1} + \tau(b_{k+1} - a_{k+1})\\ &= a_k + \tau(\mu_k - a_k)\\ &= a_k + \tau(a_k + \tau(b_k - a_k) - a_k)\\ &= a_k + \tau^2(b_k - a_k).\end{aligned} \tag{2.3.5}$$

If we set
$$\tau^2 = 1 - \tau, \tag{2.3.6}$$
then
$$\mu_{k+1} = a_k + (1-\tau)(b_k - a_k) = \lambda_k. \tag{2.3.7}$$

It means that the new observation μ_{k+1} does not need to compute, because μ_{k+1} coincides with λ_k.

Similarly, if we consider Case 2, the new observation λ_{k+1} coincides with μ_k. Therefore, for each new iteration, only one extra observation is needed, which is just required by the third condition.

What is the reduction rate of the interval of uncertainty for each iteration? By solving the equation (2.3.6), we immediately obtain

$$\tau = \frac{-1 \pm \sqrt{5}}{2}.$$

Since $\tau > 0$, then take

$$\tau = \frac{b_{k+1} - a_{k+1}}{b_k - a_k} = \frac{\sqrt{5} - 1}{2} \approx 0.618. \tag{2.3.8}$$

Then the formula (2.3.3)–(2.3.4) can be written as

$$\lambda_k = a_k + 0.382(b_k - a_k), \tag{2.3.9}$$
$$\mu_k = a_k + 0.618(b_k - a_k). \tag{2.3.10}$$

Therefore, the golden section method is also called the 0.618 method.

Obviously, comparing with the Fibonacci method below, the golden section method is more simple in performance and we need not know the number of observations in advance.

Since, for each iteration, the reduction rate of the interval of uncertainty is $\tau = 0.618$, then if the initial interval is $[a_1, b_1]$, the length of the interval after n-th iteration is $\tau^{n-1}(b_1 - a_1)$. Therefore the convergence rate of the golden section method is linear.

Algorithm 2.3.1 *(The Golden Section Method)*

Step 1. Initial step. Determine the initial interval $[a_1, b_1]$ and give the precision $\delta > 0$. Compute initial observations λ_1 and μ_1:

$$\lambda_1 = a_1 + 0.382(b_1 - a_1),$$
$$\mu_1 = a_1 + 0.618(b_1 - a_1),$$

evaluate $\phi(\lambda_1)$ and $\phi(\mu_1)$, set $k = 1$.

Step 2. Compare the function values. If $\phi(\lambda_k) > \phi(\mu_k)$, go to Step 3; if $\phi(\lambda_k) \leq \phi(\mu_k)$, go to Step 4.

Step 3. *(Case 2) If $b_k - \lambda_k \le \delta$, stop and output μ_k; otherwise, set*

$$a_{k+1} := \lambda_k, b_{k+1} := b_k, \lambda_{k+1} := \mu_k,$$
$$\phi(\lambda_{k+1}) := \phi(\mu_k), \mu_{k+1} := a_{k+1} + 0.618(b_{k+1} - a_{k+1}).$$

Evaluate $\phi(\mu_{k+1})$ and go to Step5.

Step 4. *(Case 1) If $\mu_k - a_k \le \delta$, stop and output λ_k; otherwise set*

$$a_{k+1} := a_k, b_{k+1} := \mu_k, \mu_{k+1} := \lambda_k,$$
$$\phi(\mu_{k+1}) := \phi(\lambda_k), \lambda_{k+1} := a_{k+1} + 0.382(b_{k+1} - a_{k+1}).$$

Evaluate $\phi(\lambda_{k+1})$ and go to Step 5.

Step 5. *$k := k + 1$, go to Step 2.* □

2.3.2 The Fibonacci Method

Another section method which is similar to the golden section method is the Fibonacci method. Their main difference is in that the reduction rate of the interval of uncertainty for the Fibonacci method does not use the golden section number $\tau \approx 0.618$, but uses the Fibonacci number. Therefore the reduction of the interval of uncertainty varies from one iteration to another.

The Fibonacci sequence $\{F_k\}$ is defined as follows:

$$F_0 = F_1 = 1, \tag{2.3.11}$$
$$F_{k+1} = F_k + F_{k-1},\ k = 1, 2, \cdots. \tag{2.3.12}$$

If we use F_{n-k}/F_{n-k+1} instead of τ in (2.3.3)–(2.3.4), we immediately obtain the formula

$$\begin{aligned} \lambda_k &= a_k + \left(1 - \frac{F_{n-k}}{F_{n-k+1}}\right)(b_k - a_k) &(2.3.13)\\ &= a_k + \frac{F_{n-k-1}}{F_{n-k+1}}(b_k - a_k),\ k = 1, \cdots, n-1, \\ \mu_k &= a_k + \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k),\ k = 1, \cdots, n-1, &(2.3.14) \end{aligned}$$

which is called the Fibonacci formula.

As stated in the last section, in Case 1, if $\phi(\lambda_k) \leq \phi(\mu_k)$, the new interval of uncertainty is $[a_{k+1}, b_{k+1}] = [a_k, \mu_k]$. So, by using (2.3.14), we get

$$b_{k+1} - a_{k+1} = \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k) \tag{2.3.15}$$

which gives a reduction in each iteration. This equality is also true for Case 2.

Assume that we ask for the length of the final interval no more than δ, i.e.,

$$b_n - a_n \leq \delta.$$

Since

$$\begin{aligned} b_n - a_n &= \frac{F_1}{F_2}(b_{n-1} - a_{n-1}) \\ &= \frac{F_1}{F_2}\frac{F_2}{F_3}\cdots\frac{F_{n-1}}{F_n}(b_1 - a_1) \\ &= \frac{1}{F_n}(b_1 - a_1), \end{aligned} \tag{2.3.16}$$

then

$$F_n \geq \frac{b_1 - a_1}{\delta}. \tag{2.3.17}$$

Therefore, given initial interval $[a_1, b_1]$ and the upper bound δ of the length of the final interval, we can find the Fibonacci number F_n and further n from (2.3.17). Our search proceeds until the n-th observation. The procedure of the Fibonacci method is similar to Algorithm 2.3.1. We leave it as an exercise.

Letting $F_k = r^k$ and substituting in (2.3.11)-(2.3.12), we get

$$r^2 - r - 1 = 0. \tag{2.3.18}$$

Solving (2.3.18) gives

$$r_1 = \frac{1 + \sqrt{5}}{2}, \; r_2 = \frac{1 - \sqrt{5}}{2}. \tag{2.3.19}$$

Then, the general solution of the difference equation $F_{k+1} = F_k + F_{k-1}$ is

$$F_k = Ar_1^k + Br_2^k. \tag{2.3.20}$$

Using the initial condition $F_0 = F_1 = 1$, we get

$$A = \frac{r_1}{\sqrt{5}}, \ B = -\frac{r_2}{\sqrt{5}}.$$

Substituting in (2.3.20) gives

$$F_k = \frac{1}{\sqrt{5}} \left\{ \left(\frac{1+\sqrt{5}}{2} \right)^{k+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \right\}. \tag{2.3.21}$$

Hence

$$\lim_{k\to\infty} \frac{F_{k-1}}{F_k} = \frac{\sqrt{5}-1}{2} = \tau. \tag{2.3.22}$$

This shows that, when $k \to \infty$, the Fibonacci method and the golden section method have the same reduction rate of the interval of uncertainty. Therefore the Fibonacci method converges with convergence ratio τ. It is worth mentioning that the Fibonacci method is the optimal sectioning method for one-dimensional optimization and it requires the smallest observations for a given final length δ, and that the golden section method is approximately optimal. However, since the procedure of the golden section method is very simple, it is more popular.

2.4 Interpolation Method

Interpolation Methods are the other approach of line search. This class of methods approximates $\phi(\alpha) = f(x + \alpha d)$ by fitting a quadratic or cubic polynomial in α to known data, and choosing a new α-value which minimizes the polynomial. Then we reduce the bracketing interval by comparing the new α-value and the known points. In general, when the function has good analytical properties, for example, it is easy to get the derivatives, the interpolation methods are superior to the golden section method and the Fibonacci method discussed in the last subsection.

2.4.1 Quadratic Interpolation Methods

1. Quadratic Interpolation Method with Two Points (I).
Given two points α_1, α_2, and their function values $\phi(\alpha_1)$ and $\phi(\alpha_2)$, and the derivative $\phi'(\alpha_1)$ (or $\phi'(\alpha_2)$). Construct the quadratic interpolation function

$q(\alpha) = a\alpha^2 + b\alpha + c$ with the interpolation conditions:

$$\begin{aligned} q(\alpha_1) &= a\alpha_1^2 + b\alpha_1 + c = \phi(\alpha_1), \\ q(\alpha_2) &= a\alpha_2^2 + b\alpha_2 + c = \phi(\alpha_2), \\ q'(\alpha_1) &= 2a\alpha_1 + b = \phi'(\alpha_1). \end{aligned} \tag{2.4.1}$$

Write $\phi_1 = \phi(\alpha_1), \phi_2 = \phi(\alpha_2), {\phi'}_1 = \phi'(\alpha_1)$, and ${\phi'}_2 = \phi'(\alpha_2)$. Solving (2.4.1) gives

$$\begin{aligned} a &= \frac{\phi_1 - \phi_2 - {\phi'}_1(\alpha_1 - \alpha_2)}{-(\alpha_1 - \alpha_2)^2}, \\ b &= {\phi'}_1 + 2\frac{\phi_1 - \phi_2 - {\phi'}_1(\alpha_1 - \alpha_2)}{(\alpha_1 - \alpha_2)^2}\alpha_1. \end{aligned}$$

Hence

$$\begin{aligned} \bar{\alpha} &= -\frac{b}{2a} \\ &= \alpha_1 + \frac{1}{2}\frac{{\phi'}_1(\alpha_1 - \alpha_2)^2}{\alpha_1 - \alpha_2 - {\phi'}_1(\alpha_1 - \alpha_2)} \\ &= \alpha_1 - \frac{1}{2}\frac{(\alpha_1 - \alpha_2){\phi'}_1}{{\phi'}_1 - \frac{\phi_1 - \phi_2}{\alpha_1 - \alpha_2}}. \end{aligned} \tag{2.4.2}$$

Then we get the following iteration formula:

$$\alpha_{k+1} = \alpha_k - \frac{1}{2}\frac{(\alpha_k - \alpha_{k-1}){\phi'}_k}{{\phi'}_k - \frac{\phi_k - \phi_{k-1}}{\alpha_k - \alpha_{k-1}}}. \tag{2.4.3}$$

where $\phi_k = \phi(\alpha_k), \phi_{k-1} = \phi(\alpha_{k-1})$, and ${\phi'}_k = \phi'(\alpha_k)$.

After finding the new α_{k+1}, we compare α_{k+1} with α_k and α_{k-1}, and reduce the bracketing interval. The procedure will continue until the length of the interval is less than a prescribed tolerance.

2. Quadratic Interpolation Method with Two Points (II).
Given two points α_1, α_2, and one function value $\phi(\alpha_1)$ (or $\phi(\alpha_2)$), and two derivative values $\phi'(\alpha_1)$ and $\phi'(\alpha_2)$. Construct the quadratic interpolation function with the following conditions:

$$\begin{aligned} q(\alpha_1) &= a\alpha_1^2 + b\alpha_1 + c = \phi(\alpha_1), \\ q'(\alpha_1) &= 2a\alpha_1 + b = \phi'(\alpha_1), \\ q'(\alpha_2) &= 2a\alpha_2 + b = \phi'(\alpha_2). \end{aligned} \tag{2.4.4}$$

Precisely, with the same discussion as above we obtain

$$\bar{\alpha} = -\frac{b}{2a} = \alpha_1 - \frac{\alpha_1 - \alpha_2}{\phi'_1 - \phi'_2}\phi'_1. \tag{2.4.5}$$

Therefore the iteration scheme is

$$\alpha_{k+1} = \alpha_k - \frac{\alpha_k - \alpha_{k-1}}{\phi'_k - \phi'_{k-1}}\phi'_k \tag{2.4.6}$$

which is also called the secant formula. The formula (2.4.5) can also be got by setting $L(\alpha) = 0$ where $L(\alpha)$ is the Lagrange interpolation formula

$$L(\alpha) = \frac{(\alpha - \alpha_1)\phi'_2 - (\alpha - \alpha_2)\phi'_1}{\alpha_2 - \alpha_1} \tag{2.4.7}$$

which interpolates the derivative values of $\phi'(\alpha)$ at two points α_1 and α_2.

In the following discussion, for convenience, we call the quadratic interpolating method (I) the quadratic interpolation formula, and the quadratic interpolation method (II) the secant formula. Next, we turn to the convergence of the quadratic interpolating method with two points.

Theorem 2.4.1 *Let $\phi : R \to R$ be three times continuously differentiable. Let α^* be such that $\phi'(\alpha^*) = 0$ and $\phi''(\alpha^*) \neq 0$. Then the sequence $\{\alpha_k\}$ generated from (2.4.6) converges to α^* with the order $\frac{1+\sqrt{5}}{2} \approx 1.618$ of convergence rate.*

Proof. By the representation of residual term of the Lagrange interpolation formula

$$R_2(\alpha) = \phi'(\alpha) - L(\alpha) = \frac{1}{2}\phi'''(\xi)(\alpha - \alpha_k)(\alpha - \alpha_{k-1}),\ \xi \in (\alpha, \alpha_{k-1}, \alpha_k). \tag{2.4.8}$$

Setting $\alpha = \alpha_{k+1}$ and noting that $L(\alpha_{k+1}) = 0$, we have

$$\phi'(\alpha_{k+1}) = \frac{1}{2}\phi'''(\xi)(\alpha_{k+1} - \alpha_k)(\alpha_{k+1} - \alpha_{k-1}),\ \xi \in (\alpha_{k-1}, \alpha_k, \alpha_{k+1}), \tag{2.4.9}$$

Substituting (2.4.6) into (2.4.9) yields

$$\phi'(\alpha_{k+1}) = \frac{1}{2}\phi'''(\xi)\phi'_k\phi'_{k-1}\frac{(\alpha_k - \alpha_{k-1})^2}{(\phi'_k - \phi'_{k-1})^2},\ \xi \in (\alpha_{k-1}, \alpha_k, \alpha_{k+1}). \tag{2.4.10}$$

We know from the mean-value theorem of differentiation that

$$\frac{\phi'_k - \phi'_{k-1}}{\alpha_k - \alpha_{k-1}} = \phi''(\xi_0),\ \xi_0 \in (\alpha_{k-1}, \alpha_k), \tag{2.4.11}$$

$$\phi'_i = \phi'_i - \phi'(\alpha^*) = (\alpha_i - \alpha^*)\phi''(\xi_i), \tag{2.4.12}$$

where $\xi_i \in (\alpha_i, \alpha^*), i = k-1, k, k+1$. Therefore it follows from (2.4.10)-(2.4.12) that

$$\alpha_{k+1} - \alpha^* = \frac{1}{2}\frac{\phi'''(\xi)\phi''(\xi_k)\phi''(\xi_{k-1})}{\phi''(\xi_{k+1})[\phi''(\xi_0)]^2}(\alpha_k - \alpha^*)(\alpha_{k-1} - \alpha^*). \tag{2.4.13}$$

Let $e_i = |\alpha_i - \alpha^*|, (i = k-1, k, k+1)$. In the intervals considered, let

$$0 < m_2 \le |\phi'''(\alpha)| \le M_2,\ 0 < m_1 \le |\phi''(\alpha)| \le M_1,$$

$$K_1 = m_2 m_1^2/(2M_1^3),\ K = M_2 M_1^2/(2m_1^3).$$

Then

$$K_1|\alpha_k - \alpha^*||\alpha_{k-1} - \alpha^*| \le |\alpha_{k+1} - \alpha^*| \le K|\alpha_k - \alpha^*||\alpha_{k-1} - \alpha^*|. \tag{2.4.14}$$

Noting that ϕ'' and ϕ''' are continuous at α^*, we get

$$\frac{\alpha_{k+1} - \alpha^*}{(\alpha_k - \alpha^*)(\alpha_{k-1} - \alpha^*)} \to \frac{1}{2}\frac{\phi'''(\alpha^*)}{\phi''(\alpha^*)}. \tag{2.4.15}$$

Therefore

$$e_{k+1} = \left|\frac{\phi'''(\eta_1)}{2\phi''(\eta_2)}\right| e_k e_{k-1} \triangleq M e_k e_{k-1}, \tag{2.4.16}$$

where $\eta_1 \in (\alpha_{k-1}, \alpha_k, \alpha^*), \eta_2 \in (\alpha_{k-1}, \alpha_k), M = |\phi'''(\eta_1)/2\phi''(\eta_2)|$. The above relations indicate that there exists $\delta > 0$ such that, when the initial points $\alpha_0, \alpha_1 \in (\alpha^* - \delta, \alpha^* + \delta)$ and $\alpha_0 \neq \alpha_1$, the sequence $\{\alpha_k\} \to \alpha^*$.

Next, we consider the convergence rate. Set $\epsilon_i = Me_i, y_i = \ln \epsilon_i, i = k-1, k, k+1$, then

$$\epsilon_{k+1} = \epsilon_k \epsilon_{k-1}, \tag{2.4.17}$$

$$y_{k+1} = y_k + y_{k-1}. \tag{2.4.18}$$

Obviously, (2.4.18) is the equation that the Fibonacci sequence satisfies, and its characteristic equation is

$$t^2 - t - 1 = 0 \tag{2.4.19}$$

whose solutions are

$$t_1 = \frac{1+\sqrt{5}}{2},\ t_2 = \frac{1-\sqrt{5}}{2}. \tag{2.4.20}$$

Therefore the Fibonacci sequence $\{y_k\}$ can be written as

$$y_k = At_1^k + Bt_2^k,\ k = 0, 1, \cdots, \tag{2.4.21}$$

where A and B are coefficients to be determined. Obviously, when $k \to \infty$,

$$\ln \epsilon_k = y_k \approx At_1^k. \tag{2.4.22}$$

Since

$$\frac{\epsilon_{k+1}}{\epsilon_k^{t_1}} \approx \frac{\exp(At_1^{k+1})}{[\exp(At_1^k)]^{t_1}} = 1,$$

then

$$\frac{e_{k+1}}{e_k^{t_1}} \approx M^{t_1-1} \tag{2.4.23}$$

which implies that the convergence rate is $t_1 = \frac{1+\sqrt{5}}{2} \approx 1.618$. $\square$

This theorem tells us that the secant method has superlinear convergence.

3. Quadratic Interpolation Method with Three Points.

Given three distinct points α_1, α_2 and α_3, and their function values. The required interpolation conditions are

$$q(\alpha_i) = a\alpha_i^2 + b\alpha_i + c = \phi(\alpha_i),\ i = 1, 2, 3. \tag{2.4.24}$$

By solving the above equations, we obtain

$$a = -\frac{(\alpha_2-\alpha_3)\phi_1 + (\alpha_3-\alpha_1)\phi_2 + (\alpha_1-\alpha_2)\phi_3}{(\alpha_1-\alpha_2)(\alpha_2-\alpha_3)(\alpha_3-\alpha_1)},$$

$$b = \frac{(\alpha_2^2-\alpha_3^2)\phi_1 + (\alpha_3^2-\alpha_1^2)\phi_2 + (\alpha_1^2-\alpha_2^2)\phi_3}{(\alpha_1-\alpha_2)(\alpha_2-\alpha_3)(\alpha_3-\alpha_1)}.$$

Then

$$\begin{aligned}\bar{\alpha} &= -\frac{b}{2a} \\ &= \frac{1}{2}\frac{(\alpha_2^2-\alpha_3^2)\phi_1 + (\alpha_3^2-\alpha_1^2)\phi_2 + (\alpha_1^2-\alpha_2^2)\phi_3}{(\alpha_2-\alpha_3)\phi_1 + (\alpha_3-\alpha_1)\phi_2 + (\alpha_1-\alpha_2)\phi_3} && (2.4.25) \\ &= \frac{1}{2}(\alpha_1+\alpha_2) + \frac{1}{2}\frac{(\phi_1-\phi_2)(\alpha_2-\alpha_3)(\alpha_3-\alpha_1)}{(\alpha_2-\alpha_3)\phi_1 + (\alpha_3-\alpha_1)\phi_2 + (\alpha_1-\alpha_2)\phi_3}. && (2.4.26)\end{aligned}$$

Equations (2.4.25) and (2.4.26) are called the quadratic interpolation formula with three points. The above formula can also be obtained from considering the Lagrange interpolation formula

$$L(\alpha) = \frac{(\alpha-\alpha_2)(\alpha-\alpha_3)}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)}\phi_1 + \frac{(\alpha-\alpha_1)(\alpha-\alpha_3)}{(\alpha_2-\alpha_1)(\alpha_2-\alpha_3)}\phi_2 + \frac{(\alpha-\alpha_1)(\alpha-\alpha_2)}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}\phi_3, \tag{2.4.27}$$

and setting $L'(\alpha) = 0$.

Algorithm 2.4.2 *(Line Search Employing Quadratic Interpolation with Three Points)*

Step 0. Given tolerance ϵ. Find an initial bracket $\{\alpha_1, \alpha_2, \alpha_3\}$ containing α^*; Compute $\phi(\alpha_i), i = 1, 2, 3$.

Step 1. Use the formula (2.4.25) to produce $\bar{\alpha}$;

Step 2. If $(\bar{\alpha}-\alpha_1)(\bar{\alpha}-\alpha_3) \geq 0$ go to Step 3; otherwise go to Step 4;

Step 3. Construct new bracket $\{\alpha_1, \alpha_2, \alpha_3\}$ from $\alpha_1, \alpha_2, \alpha_3$ and $\bar{\alpha}$. Go to Step 1.

Step 4. If $|\bar{\alpha}-\alpha_2| < \varepsilon$, stop; otherwise go to Step 3. □

Figure 2.4.1 is a diagram for the quadratic interpolation line search with three points.

The following theorem shows that the above algorithm has convergence rate with order 1.32.

Theorem 2.4.3 *Let $\phi(\alpha)$ have continuous fourth-order derivatives. Let α^* satisfy $\phi'(\alpha^*) = 0$ and $\phi''(\alpha^*) \neq 0$. Then the sequence $\{\alpha_k\}$ generated from the formula (2.4.25) has convergence rate with order 1.32.*

Proof. By Lagrange interpolation formula (2.4.27), we have

$$\phi(\alpha) = L(\alpha) + R_3(\alpha), \tag{2.4.28}$$

where

$$R_3(\alpha) = \frac{1}{6}\phi'''(\xi(\alpha))(\alpha-\alpha_1)(\alpha-\alpha_2)(\alpha-\alpha_3). \tag{2.4.29}$$

Since $0 = \phi'(\alpha^*) = L'(\alpha^*) + R_3'(\alpha^*)$, we get

$$\begin{aligned} &\phi_1 \frac{2\alpha^* - (\alpha_2 + \alpha_3)}{(\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)} + \phi_2 \frac{2\alpha^* - (\alpha_3 + \alpha_1)}{(\alpha_2 - \alpha_3)(\alpha_2 - \alpha_1)} \\ &+\phi_3 \frac{2\alpha^* - (\alpha_1 + \alpha_2)}{(\alpha_3 - \alpha_1)(\alpha_3 - \alpha_2)} + R_3'(\alpha^*) = 0. \end{aligned} \tag{2.4.30}$$

Noting that (2.4.25) can be rewritten as

$$\alpha_4 = \frac{1}{2} \frac{\frac{\phi_1(\alpha_2+\alpha_3)}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} + \frac{\phi_2(\alpha_3+\alpha_1)}{(\alpha_2-\alpha_3)(\alpha_2-\alpha_1)} + \frac{\phi_3(\alpha_1+\alpha_2)}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}}{\frac{\phi_1}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} + \frac{\phi_2}{(\alpha_2-\alpha_3)(\alpha_2-\alpha_1)} + \frac{\phi_3}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}}, \tag{2.4.31}$$

it follows from (2.4.30) and (2.4.31) that

$$\alpha^* - \alpha_4 = \frac{1}{2} \frac{R_3'(\alpha^*)}{\frac{\phi_1}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)} + \frac{\phi_2}{(\alpha_2-\alpha_3)(\alpha_2-\alpha_1)} + \frac{\phi_3}{(\alpha_3-\alpha_1)(\alpha_3-\alpha_2)}}. \tag{2.4.32}$$

Let $e_i = \alpha^* - \alpha_i, i = 1, 2, 3, 4$. It follows from (2.4.32) that

$$\begin{aligned} & e_4[-\phi_1(e_2 - e_3) - \phi_2(e_3 - e_1) - \phi_3(e_1 - e_2)] \\ = \quad & -\frac{1}{2} R_3'(\alpha^*)(e_1 - e_2)(e_2 - e_3)(e_3 - e_1). \end{aligned} \tag{2.4.33}$$

Noting that $\phi'(\alpha^*) = 0$, it follows from Taylor expansion that

$$\phi_i = \phi(\alpha^*) + \frac{1}{2} e_i^2 \phi''(\alpha^*) + O(e_i^3). \tag{2.4.34}$$

Neglecting the third-order term and substituting (2.4.34) into (2.4.33) give

$$e_4 = \frac{1}{\phi''(\alpha^*)} R_3'(\alpha^*). \tag{2.4.35}$$

Also, by the Lagrange interpolation formula, we have

$$\begin{aligned} R_3'(\alpha) = \quad & \frac{1}{6}\phi'''(\xi(\alpha))[(\alpha - \alpha_2)(\alpha - \alpha_3) + (\alpha - \alpha_1)(\alpha - \alpha_3) \\ & +(\alpha - \alpha_1)(\alpha - \alpha_2)] + \frac{1}{24}\phi^{(4)}(\eta)(\alpha - \alpha_1)(\alpha - \alpha_2)(\alpha - \alpha_3), \end{aligned}$$

which implies

$$R_3'(\alpha^*) = \frac{1}{6}\phi'''(\xi(\alpha^*))(e_1 e_2 + e_2 e_3 + e_3 e_1) + \frac{1}{24}\phi^{(4)}(\eta) e_1 e_2 e_3. \tag{2.4.36}$$

Neglecting the fourth-order derivative term, it follows from (2.4.35) and (2.4.36) that

$$e_4 = \frac{\phi'''(\xi(\alpha^*))}{6\phi''(\alpha^*)}(e_1e_2 + e_2e_3 + e_3e_1) = M(e_1e_2 + s_2e_3 + e_3e_1),$$

where M is some constant. In general, we have

$$e_{k+2} = M(e_{k-1}e_k + e_ke_{k+1} + e_{k+1}e_{k-1}). \tag{2.4.37}$$

Since $e_{k+1} = O(e_k) = O(e_{k-1})$ when $e_k \to 0$, there exists $\bar{M} > 0$ such that

$$|e_{k+2}| \leq \bar{M}|e_{k-1}||e_k|,$$

i.e.,

$$\bar{M}|e_{k+2}| \leq \bar{M}|e_{k-1}|\bar{M}|e_k|.$$

When $|e_i|, (i = 1, 2, 3)$ are sufficiently small such that

$$\delta = \max\{\bar{M}|e_1|, \bar{M}|e_2|, \bar{M}|e_3|\} < 1,$$

one has

$$\bar{M}|e_4| \leq \bar{M}|e_1|\bar{M}|e_2| \leq \delta^2.$$

Set

$$\bar{M}|e_k| \leq \delta^{q_k}, \tag{2.4.38}$$

then

$$\bar{M}|e_{k+2}| \leq \bar{M}|e_k|\bar{M}|e_{k-1}| \leq \delta^{q_k}\delta^{q_{k-1}} \triangleq \delta^{q_{k+2}},$$

hence

$$q_{k+2} = q_k + q_{k-1}, \ (k \geq 2) \tag{2.4.39}$$

where $q_1 = q_2 = q_3 = 1$. Obviously, the characteristic equation of (2.4.39) is

$$t^3 - t - 1 = 0 \tag{2.4.40}$$

with one root $t_1 \approx 1.32$ and other two conjugate complex roots, $|t_2| = |t_3| < 1$. The general solution of (2.4.39) has form

$$q_k = At_1^k + Bt_2^k + Ct_3^k, \tag{2.4.41}$$

where A, B and C are coefficients to be determined. Clearly, when $k \to \infty$,

$$q_{k+1} - t_1q_k = Bt_2^k(t_2 - t_1) + Ct_3^k(t_3 - t_1) \to 0.$$

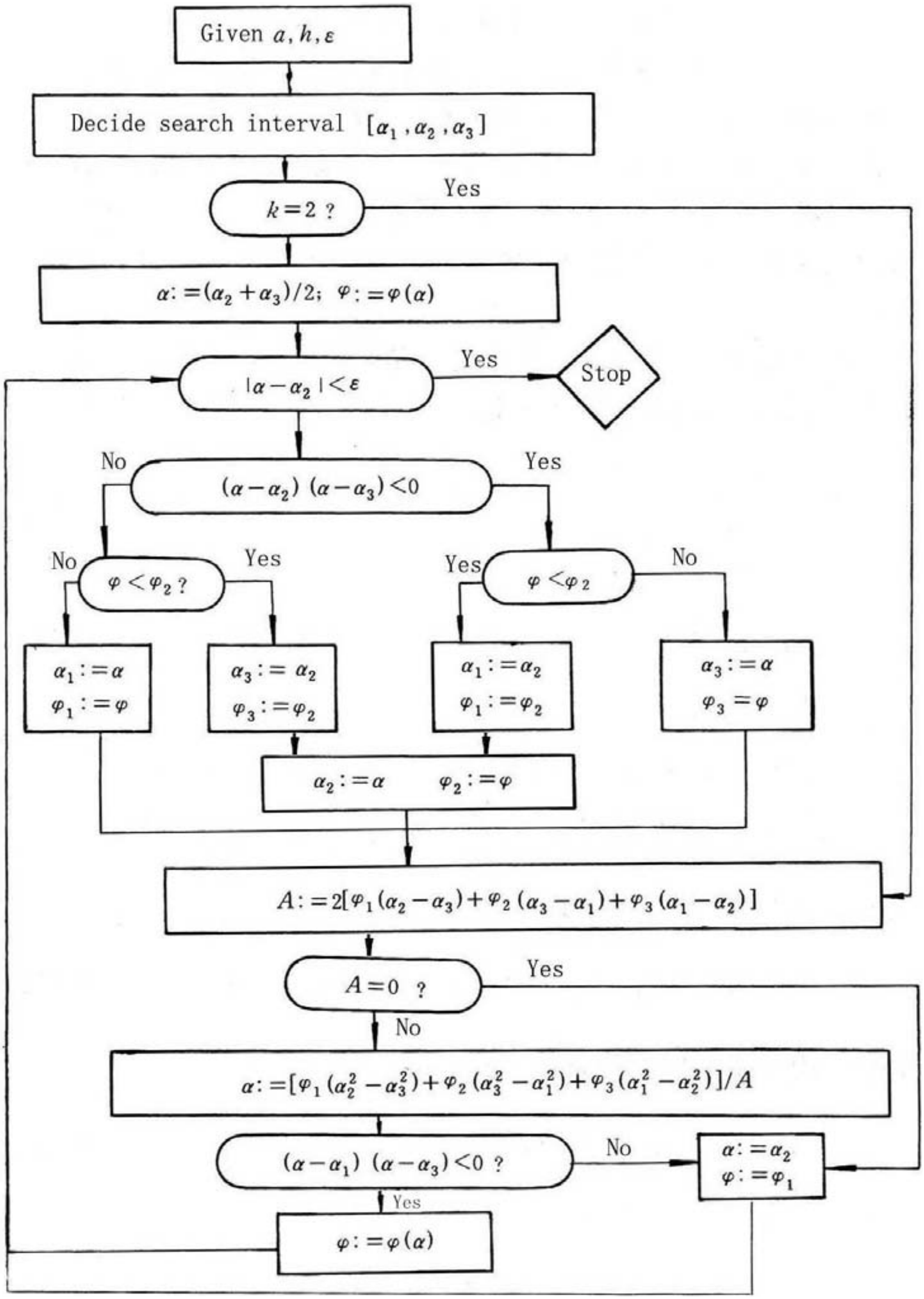


Figure 2.4.1 Flow chart for quadratic interpolation method with three points

So, when k is sufficiently large, we have

$$q_{k+1} - t_1 q_k \geq -0.1. \tag{2.4.42}$$

Note from (2.4.38) that $|e_k| \leq (1/\bar{M})\delta^{q_k} \triangleq B_k$, $(k \geq 1)$. Then, by (2.4.42), when k is sufficiently large,

$$\frac{B_{k+1}}{B_k} = \frac{\delta^{q_{k+1}}/\bar{M}}{\delta^{t_1 q_k}/(\bar{M})^{t_1}} = \bar{M}^{t_1-1}\delta^{q_{k+1}-t_1 q_k} \leq \delta^{-0.1}\bar{M}^{t_1-1},$$

which indicates that the convergence order $t_1 \approx 1.32$. $\square$

2.4.2 Cubic Interpolation Method

The cubic interpolation method approximates the objective function $\phi(\alpha)$ by a cubic polynomial. To construct the cubic polynomial $p(\alpha)$, four interpolation conditions are required. For example, we may use function values at four points, or function values at three points and a derivative value at one point, or function values and derivative values at two points. Note that, in general, the cubic interpolation has better convergence than the quadratic interpolation, but that it needs computing of derivatives and more expensive computation. Hence it is often used for smooth functions. In the following, we discuss the cubic interpolation method with two points.

We are given two points a and b, the function values $\phi(a)$ and $\phi(b)$, and the derivative values $\phi'(a)$ and $\phi'(b)$ to construct a cubic polynomial of the form

$$p(\alpha) = c_1(\alpha - a)^3 + c_2(\alpha - a)^2 + c_3(\alpha - a) + c_4 \tag{2.4.43}$$

where c_i are the coefficients of the polynomial which are chosen such that

$$\begin{aligned}
&p(a) = c_4 = \phi(a),\\
&p'(a) = c_3 = \phi'(a),\\
&p(b) = c_1(b-a)^3 + c_2(b-a)^2 + c_3(b-a) + c_4 = \phi(b),\\
&p'(b) = 3c_1(b-a)^2 + 2c_2(b-a) + c_3 = \phi'(b).
\end{aligned} \tag{2.4.44}$$

From the sufficient condition of the minimizer, we have

$$p'(\alpha) = 3c_1(\alpha - a)^2 + 2c_2(\alpha - a) + c_3 = 0 \tag{2.4.45}$$

and

$$p''(\alpha) = 6c_1(\alpha - a) + 2c_2 > 0. \tag{2.4.46}$$

Solving (2.4.45) yields

$$\begin{aligned} \alpha &= a + \frac{-c_2 \pm \sqrt{c_2^2 - 3c_1c_3}}{3c_1}, \text{ if } c_1 \neq 0, && (2.4.47) \\ \alpha &= a - \frac{c_3}{2c_2}, \text{ if } c_1 = 0. && (2.4.48) \end{aligned}$$

In order to guarantee the condition (2.4.46) holding, we only take the positive in (2.4.47). So we combine (2.4.47) with (2.4.48), and get

$$\alpha - a = \frac{-c_2 + \sqrt{c_2^2 - 3c_1c_3}}{3c_1} = \frac{-c_3}{c_2 + \sqrt{c_2^2 - 3c_1c_3}}. \tag{2.4.49}$$

When $c_1 = 0$, (2.4.49) is just (2.4.48). Then the minimizer of $p(\alpha)$ is

$$\bar{\alpha} = a - \frac{c_3}{c_2 + \sqrt{c_2^2 - 3c_1c_3}}. \tag{2.4.50}$$

The minimizer in (2.4.50) is represented by c_1, c_2 and c_3. We hope to represent $\bar{\alpha}$ by $\phi(a), \phi(b), \phi'(a)$ and $\phi'(b)$ directly.

Let

$$\begin{aligned} & s = 3\frac{\phi(b) - \phi(a)}{b-a}, \ z = s - \phi'(a) - \phi'(b), \\ & w^2 = z^2 - \phi'(a)\phi'(b). \end{aligned} \tag{2.4.51}$$

By use of (2.4.44), we have

$$\begin{aligned} s &= 3\frac{\phi(b) - \phi(a)}{b-a} = 3[c_1(b-a)^2 + c_2(b-a) + c_3], \\ z &= s - \phi'(a) - \phi'(b) = c_2(b-a) + c_3, \\ w^2 &= z^2 - \phi'(a)\phi'(b) = (b-a)^2(c_2^2 - 3c_1c_3). \end{aligned}$$

Then

$$(b-a)c_2 = z - c_3, \ \sqrt{c_2^2 - 3c_1c_3} = \frac{w}{b-a},$$

and so

$$c_2 + \sqrt{c_2^2 - 3c_1c_3} = \frac{z + w - c_3}{b-a}. \tag{2.4.52}$$

Using $c_3 = \phi'(a)$ and substituting (2.4.52) into (2.4.50), we get

$$\bar{\alpha} - a = \frac{-(b-a)\phi'(a)}{z + w - \phi'(a)}, \tag{2.4.53}$$

which is

$$\begin{aligned} \bar{\alpha} - a &= \frac{-(b-a)\phi'(a)\phi'(b)}{(z + w - \phi'(a))\phi'(b)} = \frac{-(b-a)(z^2 - w^2)}{\phi'(b)(z+w) - (z^2 - w^2)} \\ &= \frac{(b-a)(w-z)}{\phi'(b) - z + w}. \end{aligned} \tag{2.4.54}$$

Unfortunately, the formula (2.4.54) is not adequate for calculating $\bar{\alpha}$, because its denominator is possibly zero or merely very small. Fortunately, it can be overcome by use of (2.4.53) and (2.4.54), and we have

$$\begin{aligned} \bar{\alpha} - a &= \frac{-(b-a)\phi'(a)}{z + w - \phi'(a)} = \frac{(b-a)(w-z)}{\phi'(b) - z + w} \\ &= \frac{(b-a)(-\phi'(a) + w - z)}{\phi'(b) - \phi'(a) + 2w} \\ &= (b-a)\left(1 - \frac{\phi'(b) + z + w}{\phi'(b) - \phi'(a) + 2w}\right), \end{aligned} \tag{2.4.55}$$

or

$$\bar{\alpha} = a + (b-a)\frac{w - \phi'(a) - z}{\phi'(b) - \phi'(a) + 2w}. \tag{2.4.56}$$

In (2.4.55) and (2.4.56), the denominator $\phi'(b) - \phi'(a) + 2w \neq 0$. In fact, since $\phi'(a) < 0$ and $\phi'(b) > 0$, then $w^2 = z^2 - \phi'(a)\phi'(b) > 0$. Taking $w > 0$, it follows that the denominator $\phi'(b) - \phi'(a) + 2w > 0$.

In the same way as we did in the last subsection, we can discuss the convergence rate of the cubic interpolation method. Similar to (2.4.16), we can obtain

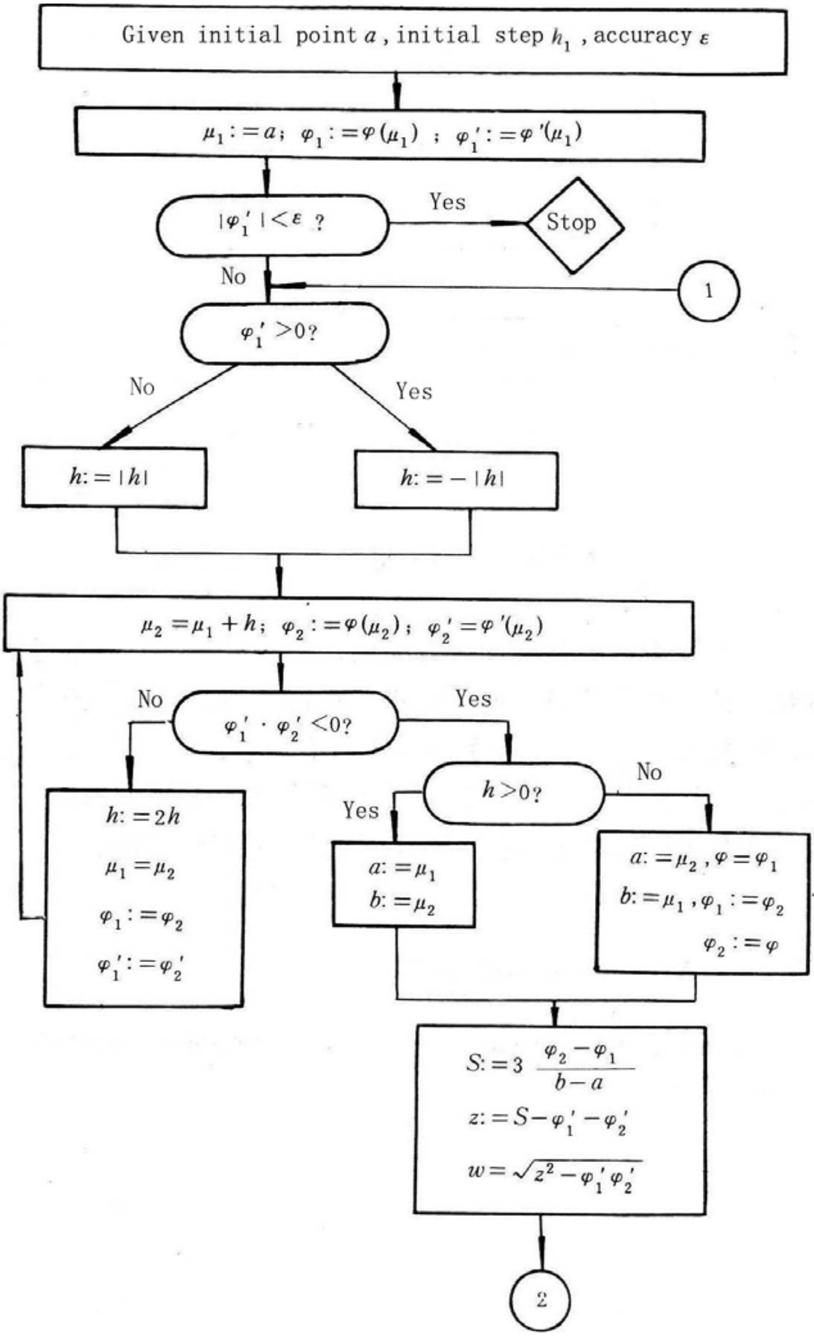
$$e_{k+1} = M(e_k e_{k-1}^2 + e_k^2 e_{k-1}),$$

where M is some constant. We can show that the characteristic equation is

$$t^2 - t - 2 = 0,$$

which solution is $t = 2$. Therefore the cubic interpolation method with two points has convergence rate with order 2.

Finally, we give a flow diagram of the method in Figure 2.4.2.


Given initial point a, initial step h_1, accuracy ε
$\mu_1 := a$; $\varphi_1 := \varphi(\mu_1)$; $\varphi_1' := \varphi'(\mu_1)$
$|\varphi_1'| < \varepsilon$?
Yes
Stop
No
1
$\varphi_1' > 0$?
No
Yes
$h := |h|$
$h := -|h|$
$\mu_2 = \mu_1 + h$; $\varphi_2 := \varphi(\mu_2)$; $\varphi_2' = \varphi'(\mu_2)$
No
$\varphi_1' \cdot \varphi_2' < 0$?
Yes
$h > 0$?
No
Yes
$h := 2h$
$\mu_1 = \mu_2$
$\varphi_1 := \varphi_2$
$\varphi_1' := \varphi_2'$
$a := \mu_1$
$b := \mu_2$
$a := \mu_2$, $\varphi = \varphi_1$
$b := \mu_1$, $\varphi_1 := \varphi_2$
$\varphi_2 := \varphi$
$S := 3\,\dfrac{\varphi_2 - \varphi_1}{b - a}$
$z := S - \varphi_1' - \varphi_2'$
$w = \sqrt{z^2 - \varphi_1'\varphi_2'}$
2

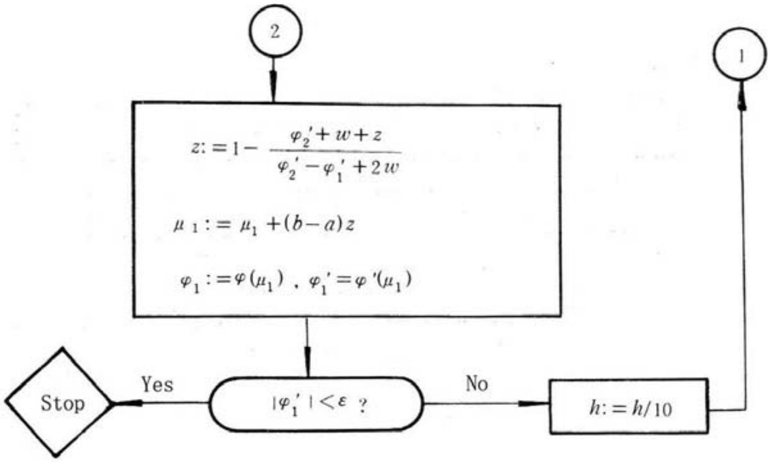


Figure 2.4.2 Flow chart for cubic interpolation method with two points

2.5 Inexact Line Search Techniques

Line search is a basic part of optimization methods. In the last sections we have discussed some exact line search techniques which find α_k such that

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k),$$

or

$$\alpha_k = \min\{\alpha |\ \nabla f(x_k + \alpha d_k)^T d_k = 0,\ \alpha \geq 0\}.$$

However, commonly, the exact line search is expensive. Especially, when an iterate is far from the solution of the problem, it is not effective to solve exactly a one-dimension subproblem. Also, in practice, for many optimization methods, for example, Newton method and quasi-Newton method, their convergence rate does not depend on the exact line search. Therefore, as long as there is an acceptable steplength rule which ensures that the objective function has sufficient descent, the exact line search can be avoided and the computing efforts will be decreased greatly. In the following, we define $g_k = \nabla f(x_k)$ without special indication.

2.5.1 Armijo and Goldstein Rule

Armijo rule [4] is as follows: Given $\beta \in (0,1), \rho \in (0,\frac{1}{2}), \tau > 0$, there exists the least nonnegative integer m_k such that

$$f(x_k) - f(x_k + \beta^m \tau d_k) \geq -\rho \beta^m \tau g_k^T d_k. \tag{2.5.1}$$

Goldstein (1965) [157] presented the following rule. Let

$$J = \{\alpha > 0 \mid f(x_k + \alpha d_k) < f(x_k)\} \tag{2.5.2}$$

be an interval. In Figure 2.5.1 $J = (0, a)$. In order to guarantee the objective function decreases sufficiently, we want to choose α such that it is away from the two end points of the interval J. The two reasonable conditions are

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha g_k^T d_k \tag{2.5.3}$$

and

$$f(x_k + \alpha d_k) \geq f(x_k) + (1-\rho)\alpha g_k^T d_k, \tag{2.5.4}$$

which exclude those points near the right end-point and the left end-point, where $0 < \rho < \frac{1}{2}$, All α satisfying (2.5.3)-(2.5.4) constitute the interval $J_2 = [b, c]$. We call (2.5.3)-(2.5.4) Goldstein inexact line search rule, in brief, Goldstein rule. When a step-length factor α satisfies (2.5.3)-(2.5.4), it is called an acceptable step-length factor, and the obtained interval $J_2 = [b, c]$ is called an acceptable interval.

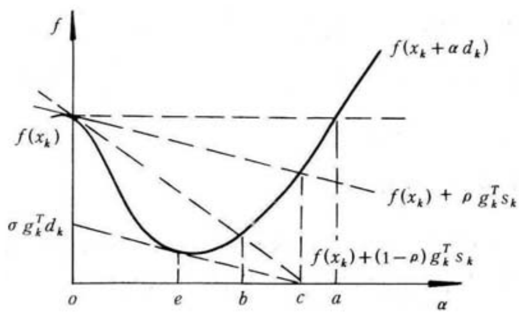


Figure 2.5.1 Inexact line search

As before, let $\phi(\alpha) = f(x_k + \alpha d_k)$. Then (2.5.3) and (2.5.4) can be rewritten respectively

$$\phi(\alpha_k) \leq \phi(0) + \rho \alpha_k \phi'(0), \tag{2.5.5}$$

$$\phi(\alpha_k) \geq \phi(0) + (1-\rho)\alpha_k \phi'(0). \tag{2.5.6}$$

Note that the restriction $\rho < \frac{1}{2}$ is necessary. In fact, if $\phi(\alpha)$ is a quadratic function satisfying $\phi'(0) < 0$ and $\phi''(0) > 0$, then the global minimizer α^* of ϕ satisfies

$$\phi(\alpha^*) = \phi(0) + \frac{1}{2}\alpha^*\phi'(0).$$

Hence α^* satisfies (2.5.5) if and only if $\rho < \frac{1}{2}$. The restriction $\rho < \frac{1}{2}$ will also finally permit $\alpha = 1$ for Newton method and quasi-Newton method. Therefore, without the restriction $\rho < \frac{1}{2}$, the superlinear convergence of the methods will not be guaranteed.

2.5.2 Wolfe-Powell Rule

As shown in Figure 2.5.1, it is possible that the rule (2.5.4) excludes the minimizing value of α outside the acceptable interval. Instead, the Wolfe-Powell rule gives another rule to replace (2.5.4):

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k, \ \sigma \in (\rho, 1), \tag{2.5.7}$$

which implies that

$$\begin{aligned} \phi'(\alpha_k) &= [\nabla f(x_k + \alpha_k d_k)]^T d_k \geq \sigma \nabla f(x_k)^T d_k \\ &= \sigma\phi'(0) > \phi'(0). \end{aligned} \tag{2.5.8}$$

It shows that the geometric interpretation of (2.5.7) is that the slope $\phi'(\alpha_k)$ at the acceptable point must be greater than or equal to some multiple $\sigma \in (0, 1)$ of the initial slope. The rule (2.5.3) and (2.5.7) is called the Wolfe-Powell inexact line search rule, in brief, the Wolfe-Powell rule, which gives the acceptable interval $J_3 = [e, c]$ that includes the minimizing values of α.

In fact, the rule (2.5.7) can be obtained from the mean-value theorem and (2.5.4). Let α_k satisfy (2.5.4). Then

$$\begin{aligned} \alpha_k[\nabla f(x_k + \theta_k\alpha_k d_k)]^T d_k &= f(x_k + \alpha_k d_k) - f(x_k) \\ &\geq (1 - \rho)\alpha_k \nabla f(x_k)^T d_k \end{aligned}$$

which shows (2.5.7). Now we show the existence of α_k satisfying (2.5.3) and (2.5.7). Let $\hat{\alpha}_k$ satisfy the equality in (2.5.3). By the mean-value theorem and (2.5.3), we have

$$\begin{aligned} \hat{\alpha}_k[\nabla f(x_k + \theta_k\hat{\alpha}_k d_k)]^T d_k &= f(x_k + \hat{\alpha}_k d_k) - f(x_k) \\ &= \rho\hat{\alpha}_k \nabla f(x_k)^T d_k, \end{aligned}$$

where $\theta_k \in (0,1)$. Let $\rho < \sigma < 1$, and note that $\nabla f(x_k)^T d_k < 0$, we have

$$[\nabla f(x_k + \theta_k \hat{\alpha}_k d_k)]^T d_k = \rho \nabla f(x_k)^T d_k > \sigma \nabla f(x_k)^T d_k$$

which is just (2.5.7) if we set $\alpha_k = \theta_k \hat{\alpha}_k$. The discussion above also shows that the requirement $\rho < \sigma < 1$ is necessary, such that there exists steplength factor α_k satisfying the Wolfe-Powell rule.

It should point out that the inequality requirement (2.5.7) is an approximation of the orthogonal condition

$$g_{k+1}^T d_k = 0$$

which is satisfied by exact line search. However, unfortunately, one possible disadvantage of (2.5.7) is that it does not reduce to an exact line search in the limit $\sigma \to 0$. In addition, a steplength may satisfy the Wolfe-Powell rule (2.5.3) and (2.5.7) without being close to a minimizer of ϕ. Luckily, if we replace (2.5.7) by using the rule

$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k, \tag{2.5.9}$$

the exact line search is obtained in the limit $\sigma \to 0$, and the points that are far from a stationary point of ϕ will be excluded. Therefore the rule (2.5.3) and (2.5.9) is also a successful pair of inexact line search rules which is called the strong Wolfe-Powell rule. Furthermore, we often employ the following form of the strong Wolfe-Powell rule:

$$|g_{k+1}^T d_k| \leq \sigma |g_k^T d_k| \tag{2.5.10}$$

or

$$|\phi'(\alpha_k)| \leq \sigma |\phi'(0)|. \tag{2.5.11}$$

In general, the smaller the value σ, the more exact the line search. Normally, taking $\sigma = 0.1$ gives a fairly accurate line search, whereas the value $\sigma = 0.9$ gives a weak line search. However, taking too small σ may be unwise, because the smaller the value σ, the more expensive the computing effort. Usually, $\rho = 0.1$ and $\sigma = 0.4$ are suitable, and it depends on the specific problem.

2.5.3 Goldstein Algorithm and Wolfe-Powell Algorithm

Although it is possible that the minimizing value of α may be excluded by the rule (2.5.4), it seldom occurs in practice. Therefore, Goldstein rule (2.5.3)-(2.5.4) is a frequently used rule in practice. The overall structure is illustrated in Figure 2.5.2 and the details of the algorithm are described in Algorithm 2.5.1.

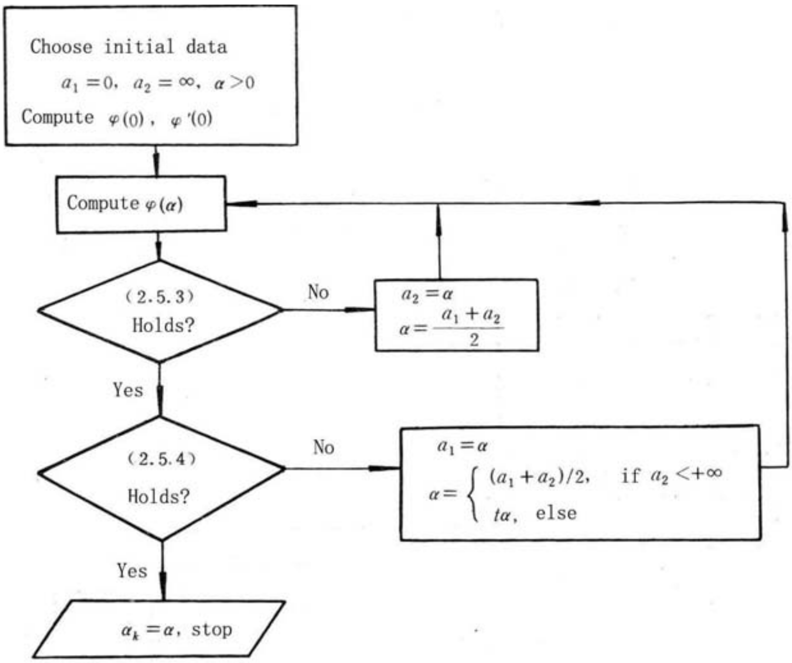


Figure 2.5.2 Flow chart for Goldstein inexact line search

Algorithm 2.5.1 *(Inexact Line Search with Goldstein Rule)*

Step 1. Choose initial data. Take initial point α_0 *in* $[0,+\infty)$ *(or* $[0,\alpha_{max}]$*). Compute* $\phi(0),\phi'(0)$*. Given* $\rho\in(0,\frac{1}{2}),t>1$*. Set* $a_0:=0,b_0:=+\infty$ *(or* α_{max}*),* $k:=0$*.*

Step 2. Check the rule (2.5.3). Compute $\phi(\alpha_k)$*. If*

$$\phi(\alpha_k)\leq\phi(0)+\rho\alpha_k\phi'(0),$$

go to Step 3; otherwise, set $a_{k+1} := a_k, b_{k+1} := \alpha_k$*, go to Step 4.*

Step 3. Check the rule (2.5.4). If

$$\phi(\alpha_k) \geq \phi(0) + (1-\rho)\alpha_k \phi'(0),$$

stop, and output α_k*; otherwise, set* $a_{k+1} := \alpha_k, b_{k+1} := b_k$*. If* $b_{k+1} < +\infty$*, go to Step 4; otherwise set* $\alpha_{k+1} := t\alpha_k, k := k+1$*, go to Step 2.*

Step 4. Choose a new point. Set

$$\alpha_{k+1} := \frac{a_{k+1} + b_{k+1}}{2},$$

and $k := k+1$*, go to Step 2.* □

Similarly, we give in Figure 2.5.3 the diagram of the Wolfe-Powell algorithm.

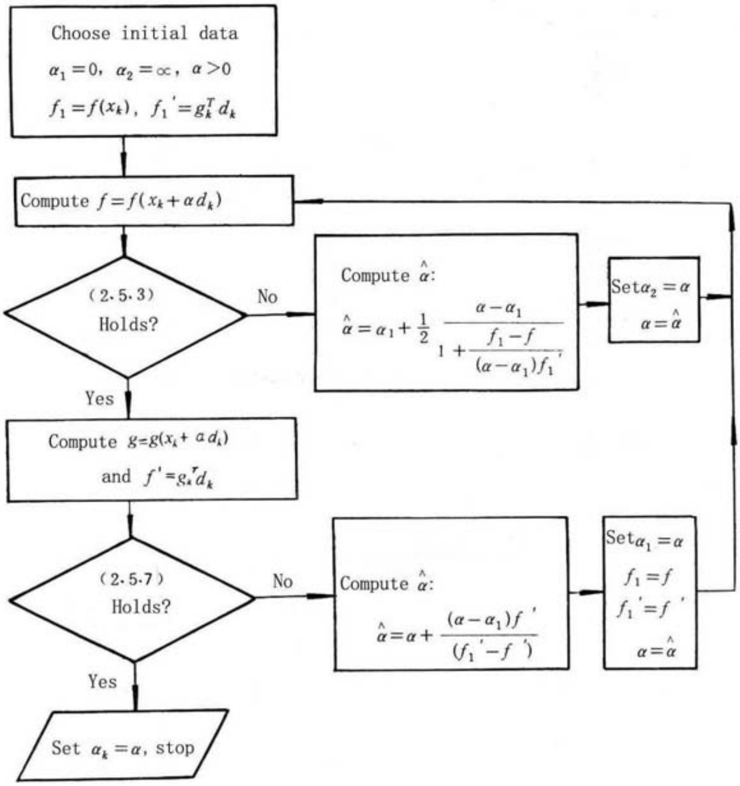


Figure 2.5.3 Flow chart for Wolfe-Powell inexact line search

2.5.4 Backtracking Line Search

In practice, frequently, we also use only the condition (2.5.3) if we choose an appropriate α which is not too small. This method is called backtracking line search. The idea of backtracking is, at the beginning, to set $\alpha = 1$. If $x_k + \alpha d_k$ is not acceptable, we reduce α until $x_k + \alpha d_k$ satisfies (2.5.3).

Algorithm 2.5.2

Step 1. Given $\rho \in (0, \frac{1}{2}), 0 < l < u < 1$, *set* $\alpha = 1$.

Step 2. Test

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha g_k^T d_k;$$

Step 3. If (2.5.3) is not satisfied, set $\alpha := \omega\alpha, \omega \in [l, u]$, *and go to Step 2; otherwise, set* $\alpha_k := \alpha$ *and* $x_{k+1} := x_k + \alpha_k d_k$. □

In Step 3 of the above algorithm, the quadratic interpolation can be used to reduce α. Let

$$\phi(\alpha) = f(x_k + \alpha d_k). \tag{2.5.12}$$

At the beginning, we have

$$\phi(0) = f(x_k),\ \phi'(0) = \nabla f(x_k)^T d_k. \tag{2.5.13}$$

After computing $f(x_k + d_k)$, we have

$$\phi(1) = f(x_k + d_k). \tag{2.5.14}$$

If $f(x_k + d_k)$ does not satisfy (2.5.3), the following quadratic model can be used to approximate $\phi(\alpha)$:

$$m(\alpha) = [\phi(1) - \phi(0) - \phi'(0)]\alpha^2 + \phi'(0)\alpha + \phi(0), \tag{2.5.15}$$

which obeys the three conditions in (2.5.13)-(2.5.14). Setting $m'(\alpha) = 0$ gives

$$\hat{\alpha} = -\frac{\phi'(0)}{2[\phi(1) - \phi(0) - \phi'(0)]}, \tag{2.5.16}$$

which can be taken as the next value of α.

In order to prevent α from being too small and not terminating, some safeguards are needed. For example, given the least step minstep, if (2.5.3) is not satisfied but $\|\alpha d_k\| <$ minstep, the line search stops.

In summary, in this section we introduced three kind of inexact line search rules:

1. Goldstein rule: (2.5.3)-(2.5.4).
2. Wolfe-Powell rule: (2.5.3) and (2.5.7); Strong Wolfe-Powell rule: (2.5.3) and (2.5.9).
3. Backtracking rule (also called Armijo rule): (2.5.3) or (2.5.1).

The above three inexact line search rules are frequently used in optimization methods below.

2.5.5 Convergence Theorems of Inexact Line Search

In the final subsection we establish convergence theorems of inexact line search methods. To prove the descent property of the methods, we try to avoid the case in which the search directions $s_k = \alpha_k d_k$ are nearly orthogonal to the negative gradient $-g_k$, that is, the angle θ_k between s_k and $-g_k$ is uniformly bounded away from 90^o,

$$\theta_k \leq \frac{\pi}{2} - \mu, \ \forall k \tag{2.5.17}$$

where $\mu > 0, \theta_k \in [0, \frac{\pi}{2}]$ is defined by

$$\cos\theta_k = -g_k^T s_k / (\|g_k\| \|s_k\|), \tag{2.5.18}$$

because, otherwise, $g_k^T s_k$ will approach zero and so s_k is almost not a descent direction.

A general descent algorithm with inexact line search is as follows:

Algorithm 2.5.3

Step 1. Given $x_0 \in R^n, 0 \leq \varepsilon < 1, k := 0$.

Step 2 If $\|\nabla f(x_k)\| \leq \varepsilon$, stop; otherwise, find a descent direction d_k such that $d_k^T \nabla f(x_k) < 0$.

Step 3 *Find the steplength factor* α_k *by use of Goldstein rule (2.5.3)-(2.5.4) or Wolfe-Powell rule (2.5.3) and (2.5.7).*

Step 4 *Set* $x_{k+1} = x_k + \alpha_k d_k; k := k + 1$, *go to Step 2.* □

In Algorithm 2.5.3, d_k is a general descent direction provided it satisfies $d_k^T \nabla f(x_k) < 0$, and α_k is a general inexact line-search factor provided some inexact line search rule is satisfied. So, this algorithm is a very general algorithm, that is, it contains a great class of methods.

Now, we establish the global convergence of the general descent algorithm with inexact line search.

Theorem 2.5.4 *Let* α_k *in Algorithm 2.5.3 be defined by Goldstein rule (2.5.3)-(2.5.4) or Wolfe-Powell rule (2.5.3) and (2.5.7). Let also* s_k *satisfy (2.5.17). If* ∇f *exists and is uniformly continuous on the level set* $\{x|\ f(x) \leq f(x_0)\}$, *then either* $\nabla f(x_k) = 0$ *for some* k, *or* $f(x_k) \to -\infty$,*or* $\nabla f(x_k) \to 0$.

Proof. Let α_k be defined by (2.5.3)-(2.5.4). Assume that, for all k, $g_k = \nabla f(x_k) \neq 0$ (whence $s_k = \alpha_k d_k \neq 0$) and $f(x_k)$ is bounded below, it follows that $f(x_k) - f(x_{k+1}) \to 0$, hence $-g_k^T s_k \to 0$ from (2.5.3).

Now assume that $g_k \to 0$ does not hold. Then there exist $\varepsilon > 0$ and a subsequence such that $\|g_k\| \geq \varepsilon$ and $\|s_k\| \to 0$. Since $\theta_k \leq \frac{\pi}{2} - \mu$, we get

$$\cos\theta_k \geq \cos(\frac{\pi}{2} - \mu) = \sin\mu,$$

hence

$$-g_k^T s_k \geq \sin\mu \|g_k\| \|s_k\| \geq \varepsilon \sin\mu \|s_k\|.$$

But the Taylor series gives

$$f(x_{k+1}) = f(x_k) + g(\xi_k)^T s_k,$$

where ξ_k is on the line segment (x_k, x_{k+1}). By uniform continuity, we have $g(\xi_k) \to g_k$ when $s_k \to 0$. So

$$f(x_{k+1}) = f(x_k) + g_k^T s_k + o(\|s_k\|).$$

Therefore we obtain

$$\frac{f(x_k) - f(x_{k+1})}{-g_k^T s_k} \to 1,$$

which contradicts (2.5.4). Hence, $g_k \to 0$, and the proof is complete.

Similarly, instead of (2.5.4), if we use (2.5.7), we can get global convergence of the Wolfe-Powell algorithm. The proof is essentially the same as above. We need only to note that, by uniform continuity of $g(x)$, it follows that

$$g_{k+1}^T s_k = g_k^T s_k + o(\|s_k\|),$$

such that

$$\frac{g_{k+1}^T s_k}{g_k^T s_k} \to 1.$$

This contradicts $g_{k+1}^T s_k / g_k^T s_k \leq \sigma < 1$ given by (2.5.7). Hence $g_k \to 0$. Therefore, the global convergence theorem also holds when α_k is defined by Wolfe-Powell rule (2.5.3) and (2.5.7). $\square$

Next, we give the convergence theorems with the Wolfe-Powell rule.

Theorem 2.5.5 *Let $f : R^n \to R$ be continuously differentiable and bounded below, and let ∇f be uniformly continuous on the level set $\Omega = \{x \mid f(x) \leq f(x_0)\}$. Assume that α_k is defined by Wolfe-Powell rule (2.5.3) and (2.5.7). Then the sequence generated by Algorithm 2.5.3 satisfies*

$$\lim_{k \to +\infty} \frac{\nabla f(x_k)^T s_k}{\|s_k\|} = 0, \tag{2.5.19}$$

which means

$$\|\nabla f(x_k)\| \cos\theta_k \to 0. \tag{2.5.20}$$

Proof. Since $\nabla f(x_k)^T s_k < 0$ and f is bounded below, then the sequence $\{x_k\}$ is well-defined and $\{x_k\} \subset \Omega$. Also, since $\{f(x_k)\}$ is a descent sequence, hence it is convergent.

We now prove (2.5.19) by contradiction. Assume that (2.5.19) does not hold. Then there exist $\varepsilon > 0$ and a subsequence with index set K, such that

$$-\frac{\nabla f(x_k)^T s_k}{\|s_k\|} \geq \varepsilon, \ k \in K.$$

From (2.5.3), one has

$$f(x_k) - f(x_{k+1}) \geq \rho \|s_k\| \left(-\frac{\nabla f(x_k)^T s_k}{\|s_k\|} \right) \geq \rho \|s_k\| \varepsilon.$$

Since also $\{f(x_k)\}$ is a convergent sequence, then $\{s_k : k \in K\}$ converges to zero. Also by (2.5.7), we have

$$(1-\sigma)(-\nabla f(x_k)^T s_k) \leq (\nabla f(x_k + s_k) - \nabla f(x_k))^T s_k,\ k \geq 0.$$

Therefore

$$\varepsilon \leq -\frac{\nabla f(x_k)^T s_k}{\|s_k\|} \leq \frac{1}{1-\sigma}\|\nabla f(x_k + s_k) - \nabla f(x_k)\|,\ k \in K. \tag{2.5.21}$$

However, since we have proved $\{s_k \,|\, k \in K\} \to 0$, then the right-hand side of (2.5.21) goes to zero by the uniform continuity of ∇f on the level set Ω. Hence there is a contradiction which completes the proof. $\Box$

Note that (2.5.19) implies

$$\|\nabla f(x_k)\| \cos\theta_k \to 0,$$

which is called the Zoutendijk condition, where θ_k is the angle between $-\nabla f(x_k)$ and s_k. If $\cos\theta_k \geq \delta > 0$, we have $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$. Also, if the assumption of uniform continuity is replaced by Lipschitz continuity, the theorem is also true. In the theorem below, we prove this case. We first prove a lemma which gives a bound of descent for a single step.

Lemma 2.5.6 *Let $f : D \subset R^n \to R$ be continuously differentiable, also let $\nabla f(x)$ satisfy Lipschitz condition*

$$\|\nabla f(y) - \nabla f(z)\| \leq M\|y - z\|,$$

where $M > 0$ is a constant. If $f(x_k + \alpha d_k)$ is bounded below and $\alpha > 0$, then for all $\alpha_k > 0$ satisfying (2.5.3) and (2.5.7), we have

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \beta\|\nabla f(x_k)\|^2 \cos^2\langle d_k, -\nabla f(x_k)\rangle, \tag{2.5.22}$$

where $\beta > 0$ is a constant.

Proof. From Lipschitz condition of ∇f and (2.5.7) we have

$$\alpha_k M\|d_k\|^2 \geq d_k^T[\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k)] \geq -(1-\sigma)d_k^T \nabla f(x_k),$$

that is

$$\begin{aligned} \alpha_k\|d_k\| &\geq \frac{1-\sigma}{M\|d_k\|}\|d_k\|\|\nabla f(x_k)\| \cos\langle d_k, -\nabla f(x_k)\rangle \\ &= \frac{1-\sigma}{M}\|\nabla f(x_k)\| \cos\langle d_k, -\nabla f(x_k)\rangle. \end{aligned}$$

Using (2.5.3) yields

$$\begin{array}{rl} & f(x_k) - f(x_k + \alpha_k d_k) \geq -\alpha_k \rho d_k^T \nabla f(x_k) \\ = & \alpha_k \rho \|d_k\| \|\nabla f(x_k)\| \cos\langle d_k, -\nabla f(x_k)\rangle \\ \geq & \rho \|\nabla f(x_k)\| \cos\langle d_k, -\nabla f(x_k)\rangle \dfrac{1-\sigma}{M} \|\nabla f(x_k)\| \cos\langle d_k, -\nabla f(x_k)\rangle \\ = & \dfrac{\rho(1-\sigma)}{M} \|\nabla f(x_k)\|^2 \cos^2\langle d_k, -\nabla f(x_k)\rangle, \end{array}$$

which is (2.5.22) in which $\beta = \rho(1-\sigma)/M$. □

Theorem 2.5.7 *Let $f(x)$ be continuously differentiable on R^n, and let $\nabla f(x)$ satisfy Lipschitz condition*

$$\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|. \tag{2.5.23}$$

Also let α_k in Algorithm 2.5.3 be defined by Wolfe-Powell rule (2.5.3) and (2.5.7). If the condition (2.5.17) is satisfied, then, for the sequence $\{x_k\}$ generated by Algorithm 2.5.3, either $\nabla f(x_k) = 0$ for some k, or $f(x_k) \to -\infty$, or $\nabla f(x_k) \to 0$.

Proof. Assume that $\nabla f(x_k) \neq 0, \forall k$. By Lemma 2.5.6, we have

$$f(x_k) - f(x_{k+1}) \geq \beta \cos^2 \theta_k \|\nabla f(x_k)\|^2, \tag{2.5.24}$$

where $\beta = \rho(1-\sigma)/M$ is a positive constant being independent of k. Then, for all $k > 0$, we have

$$\begin{aligned} f(x_0) - f(x_k) &= \sum_{i=0}^{k-1} [f(x_i) - f(x_{i+1})] \\ &\geq \beta \min_{0 \leq i \leq k} \|\nabla f(x_i)\|^2 \sum_{i=0}^{k-1} \cos^2 \theta_i. \end{aligned} \tag{2.5.25}$$

Since θ_k satisfies (2.5.17), this means that

$$\sum_{k=0}^{\infty} \cos^2 \theta_k = +\infty. \tag{2.5.26}$$

Then it follows from (2.5.25) that either $\nabla f(x_k) \to 0$ or $f(x_k) \to -\infty$. This completes the proof. □

In fact, Theorem 2.5.7 is a direct result coming from (2.5.20) and the angle condition (2.5.17).

Finally, we derive an estimate of descent amount of $f(x)$ under inexact line search.

Theorem 2.5.8 *Let α_k satisfy (2.5.3). If $f(x)$ is a uniformly convex function, i.e., there exists a constant $\eta > 0$ such that*

$$(y-z)^T[\nabla f(y) - \nabla f(z)] \geq \eta \|y-z\|^2, \tag{2.5.27}$$

or there exist positive constants m and M $(m < M)$, such that

$$m\|y\|^2 \leq y^T \nabla^2 f(x) y \leq M\|y\|^2. \tag{2.5.28}$$

Then

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \frac{\rho\eta}{1+\sqrt{M/m}} \|\alpha_k d_k\|^2, \tag{2.5.29}$$

where ρ is defined in (2.5.3).

Proof. We divide into two cases.

First, assume that $d_k^T \nabla f(x_k + \alpha_k d_k) \leq 0$. In this case we have

$$\begin{aligned} f(x_k) - f(x_k + \alpha_k d_k) &= \int_0^{\alpha_k} -d_k^T \nabla f(x_k + t d_k) \mathrm{d}t \\ &= \int_0^{\alpha_k} d_k^T [\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k + t d_k)] \mathrm{d}t \\ &\geq \int_0^{\alpha_k} \eta(\alpha_k - t) \mathrm{d}t \|d_k\|^2 \\ &= \frac{1}{2}\eta \|\alpha_k d_k\|^2 \\ &\geq \frac{\rho\eta}{1+\sqrt{M/m}} \|\alpha_k d_k\|^2. \end{aligned} \tag{2.5.30}$$

Second, assume that $d_k^T \nabla f(x_k + \alpha_k d_k) > 0$. Then there exists $0 < \alpha^* < \alpha_k$, such that $d_k^T \nabla f(x_k + \alpha^* d_k) = 0$. So, it follows from (2.5.28) that

$$f(x_k) - f(x_k + \alpha^* d_k) \leq \frac{1}{2} M \|\alpha^* d_k\|^2, \tag{2.5.31}$$

and

$$f(x_k + \alpha_k d_k) - f(x_k + \alpha^* d_k) \geq \frac{1}{2} m \|(\alpha_k - \alpha^*) d_k\|^2. \tag{2.5.32}$$

Since $f(x_k + \alpha_k d_k) < f(x_k)$, it follows from (2.5.31) and (2.5.32) that

$$\alpha_k \leq \left(1 + \sqrt{\frac{M}{m}}\right) \alpha^*. \tag{2.5.33}$$

Hence

$$\begin{aligned} f(x_k) - f(x_k + \alpha_k d_k) &\geq -\alpha_k \rho d_k^T \nabla f(x_k) \\ &\geq \alpha_k \rho d_k^T [\nabla f(x_k + \alpha^* d_k) - \nabla f(x_k)] \\ &\geq \eta \rho \alpha_k \alpha^* \|d_k\|^2 \\ &\geq \frac{\rho\eta}{1 + \sqrt{M/m}} \|\alpha_k d_k\|^2. \end{aligned} \tag{2.5.34}$$

Hence (2.5.29) holds in both cases. This completes the proof. $\Box$

In this chapter we have discussed exact and inexact line search techniques which guarantee monotonic decrease of the objective function. On the other hand it is found that enforcing monotonicity of the function values may considerably slow the rate of convergence, especially in the presence of narrow curved valleys. Therefore, it is reasonable to present a nonmonotonic line search technique for optimization which allows an increase in function value at each step, while retaining global convergence. Grippo etc. [164] generalized the Armijo rule to the nonmonotone case and relaxed the condition of monotonic decrease. Several papers also deal with these techniques. Here we only state the basic result of nonmonotonic line search as follows.

Theorem 2.5.9 *Let $\{x_k\}$ be a sequence defined by*

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

Let $\tau > 0, \sigma \in (0,1), \gamma \in (0,1)$ and let M be a nonnegative integer. Assume that

(i) the level set $\Omega = \{x \mid f(x) \leq f(x_0)\}$ is compact;

(ii) there exist positive numbers c_1, c_2 such that

$$\nabla f(x_k)^T d_k \leq -c_1 \|\nabla f(x_k)\|^2, \tag{2.5.35}$$

$$\|d_k\| \leq c_2 \|\nabla f(x_k)\|; \tag{2.5.36}$$

(iii) $\alpha_k = \sigma^{h_k}\tau$, where h_k is the first nonnegative integer h, such that

$$f(x_k + \sigma^h \tau d_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \gamma \sigma^h \tau \nabla f(x_k)^T d_k, \tag{2.5.37}$$

where $m(0) = 0$ *and* $0 \le m(k) \le \min[m(k-1)+1, M], k \ge 1$.

Then the sequence $\{x_k\}$ *remains in* Ω *and every accumulation point* $\bar{x}$ *satisfies* $\nabla f(\bar{x}) = 0$.

Proof. See Grippo etc. [164]. $\square$

Exercises

1. Let $f(x) = (\sin x)^6 \tan(1-x) e^{30x}$. Find the maximum of $f(x)$ in $[0,1]$ by use of the 0.618 method, quadratic interpolation method, and Goldstein line search, respectively.

2. Write the Fibonacci algorithm and its program in MATLAB (or FORTRAN, C).

3. Let $\phi(t) = e^{-t} + e^t$. Let the initial interval be $[-1, 1]$.
(1) Minimize $\phi(t)$ by 0.618 method.
(2) Minimize $\phi(t)$ by Fibonacci method.
(3) Minimize $\phi(t)$ by Armijo line search.

4. Let $\phi(t) = 1 - te^{-t^2}$. Let the initial interval be $[0, 1]$. Try to minimize $\phi(t)$ by quadratic interpolation method.

5. Let $\phi(t) = -2t^3 + 21t^2 - 60t + 50$.
(1) Minimize $\phi(t)$ by Armijo rule if $t_0 = 0.5$ and $\rho = 0.1$.
(2) Minimize $\phi(t)$ by Goldstein rule if $t_0 = 0.5$ and $\rho = 0.1$.
(3) Minimize $\phi(t)$ by Wolfe rule if $t_0 = 0.5, \rho = 0.1$, and $\sigma = 0.8$.

6. Let $f(x) = x_1^4 + x_1^2 + x_2^2$. Given current point $x_k = (1,1)^T$ and $d_k = (-3,-1)^T$. Let $\rho = 0.1, \sigma = 0.5$.
(1) Try using the Wolfe rule to find a new point x_{k+1}.
(2) Set $\alpha = 1, \alpha = 0.5, \alpha = 0.1$ respectively, describe that for which α satisfies the Wolfe rule and for which α does not satisfy the Wolfe rule.

7. Show that if $0 < \sigma < \rho < 1$, then there may be no steplengths that satisfy the Wolfe rule.

8. Describe the outline of Theorem 2.5.4.

9. Prove the other form of Theorem 2.5.5: Let $f : R^n \to R$ be continuously differentiable and bounded below, and let ∇f be Lipschitz continuous on the level set $\Omega = \{x \mid f(x) \leq f(x_0)\}$. Assume that α_k is defined by Wolfe-Powell rule (2.5.3) and (2.5.7). Then the sequence generated by Algorithm 2.5.3 satisfies

$$\lim_{k\to+\infty} \frac{\nabla f(x_k)^T s_k}{\|s_k\|} = 0,$$

which means

$$\|\nabla f(x_k)\| \cos\theta_k \to 0.$$