PRACTICAL OPTIMIZATION
Algorithms and Engineering Applications

Andreas Antoniou
Wu-Sheng Lu

Department of Electrical and Computer Engineering
University of Victoria, Canada

Springer

as decreasing values of x. Constant ε in Step 1 determines the precision of the solution. If ε is very small, say, less than 10^{-6}, then as the solution is approached, we have

$$f_{n-2,k} \approx f_{n-1,k} \approx f_{m,k} \approx f_{n,k}$$

Consequently, the distinct possibility of dividing by zero may arise in the evaluation of $x_{0,k+1}$. However, this problem can be easily prevented by using appropriate checks in Steps 6 and 7.

An alternative form of the above algorithm can be obtained by replacing the quadratic interpolation formula for equally-spaced points by the general formula of Eq. (4.30). If this is done, the mid-interval function evaluation of Step 5 is unnecessary. Consequently, if the additional computation required by Eq. (4.31) is less than one complete evaluation of $f(x)$, then the modified algorithm is likely to be more efficient.

Another possible modification is to use the cubic interpolation of Sec. 4.6 instead of quadratic interpolation. Such an algorithm is likely to reduce the number of function evaluations. However, the amount of computation could increase owing to the more complex formulation in the cubic interpolation.

4.8 Inexact Line Searches

In the multidimensional algorithms to be studied, most of the computational effort is spent in performing function and gradient evaluations in the execution of line searches. Consequently, the amount of computation required tends to depend on the efficiency and precision of the line searches used. If high precision line searches are necessary, the amount of computation will be large and if inexact line searches do not affect the convergence of an algorithm, a small amount of computation might be sufficient.

Many optimization methods have been found to be quite tolerant to line-search imprecision and, for this reason, inexact line searches are usually used in these methods.

Let us assume that

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{d}_k$$

where $\mathbf{d}_k$ is a given direction vector and α is an independent search parameter, and that function $f(\mathbf{x}_{k+1})$ has a unique minimum for some positive value of α. The linear approximation of the Taylor series in Eq. (2.4d) gives

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k) + \mathbf{g}_k^T \mathbf{d}_k \alpha \tag{4.52}$$

where

$$\mathbf{g}_k^T \mathbf{d}_k = \left. \frac{df(\mathbf{x}_k + \alpha \mathbf{d}_k)}{d\alpha} \right|_{\alpha=0}$$

Eq. (4.52) represents line A shown in Fig. 4.14a. The equation

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k) + \rho \mathbf{g}_k^T \mathbf{d}_k \alpha \tag{4.53}$$

where $0 \le \rho < \frac{1}{2}$ represents line B in Fig. 4.14a whose slope ranges from 0 to $\frac{1}{2}\mathbf{g}_k^T \mathbf{d}_k$ depending on the value of ρ, as depicted by shaded area B in Fig. 4.14a. On the other hand, the equation

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k) + (1-\rho)\mathbf{g}_k^T \mathbf{d}_k \alpha \tag{4.54}$$

represents line C in Fig. 4.14a whose slope ranges from $\mathbf{g}_k^T \mathbf{d}_k$ to $\frac{1}{2}\mathbf{g}_k^T \mathbf{d}_k$ as depicted by shaded area C in Fig. 4.14a. The angle between lines C and B, designated as θ, is given by

$$\theta = \tan^{-1}\left[\frac{-(1-2\rho)\mathbf{g}_k^T \mathbf{d}_k}{1+\rho(1-\rho)(\mathbf{g}_k^T \mathbf{d}_k)^2}\right]$$

as illustrated in Fig. 4.14b. Evidently by adjusting ρ in the range 0 to $\frac{1}{2}$, the slope of θ can be varied in the range $-\mathbf{g}_k^T \mathbf{d}_k$ to 0. By fixing ρ at some value in the permissible range, two values of α are defined by the intercepts of the lines in Eqs. (4.53) and (4.54) and the curve for $f(\mathbf{x}_{k+1})$, say, α_1 and α_2, as depicted in Fig. 4.14b.

Let α_0 be an estimate of the value of α that minimizes $f(\mathbf{x}_k + \alpha \mathbf{d}_k)$. If $f(\mathbf{x}_{k+1})$ for $\alpha = \alpha_0$ is equal to or less than the corresponding value of $f(\mathbf{x}_{k+1})$ given by Eq. (4.53), and is equal to or greater than the corresponding value of $f(\mathbf{x}_{k+1})$ given by Eq. (4.54), that is, if

$$f(\mathbf{x}_{k+1}) \le f(\mathbf{x}_k) + \rho \mathbf{g}_k^T \mathbf{d}_k \alpha_0 \tag{4.55}$$

and

$$f(\mathbf{x}_{k+1}) \ge f(\mathbf{x}_k) + (1-\rho)\mathbf{g}_k^T \mathbf{d}_k \alpha_0 \tag{4.56}$$

then α_0 may be deemed to be an acceptable estimate of α^* in that it will yield a sufficient reduction in $f(\mathbf{x})$. Under these circumstances, we have $\alpha_1 \le \alpha_0 \le \alpha_2$, as depicted in Fig. 4.14b, i.e., α_1 and α_2 constitute a bracket of the estimated minimizer α_0. Eqs. (4.55) and (4.56), which are often referred to as the *Goldstein conditions*, form the basis of a class of *inexact line searches*. In these methods, an estimate α_0 is generated by some means, based on available information, and the conditions in Eqs. (4.55) and (4.56) are checked. If both conditions are satisfied, then the reduction in $f(\mathbf{x}_{k+1})$ is deemed to be acceptable, and the procedure is terminated. On the other hand, if either Eq. (4.55) or Eq. (4.56) is violated, the reduction in $f(\mathbf{x}_{k+1})$ is deemed to be insufficient and an improved estimate of α^*, say, $\breve{\alpha}_0$, can be obtained. If Eq. (4.55) is violated, then $\alpha_0 > \alpha_2$ as depicted in Fig. 4.15a and since $\alpha_L < \alpha^* < \alpha_0$, the new

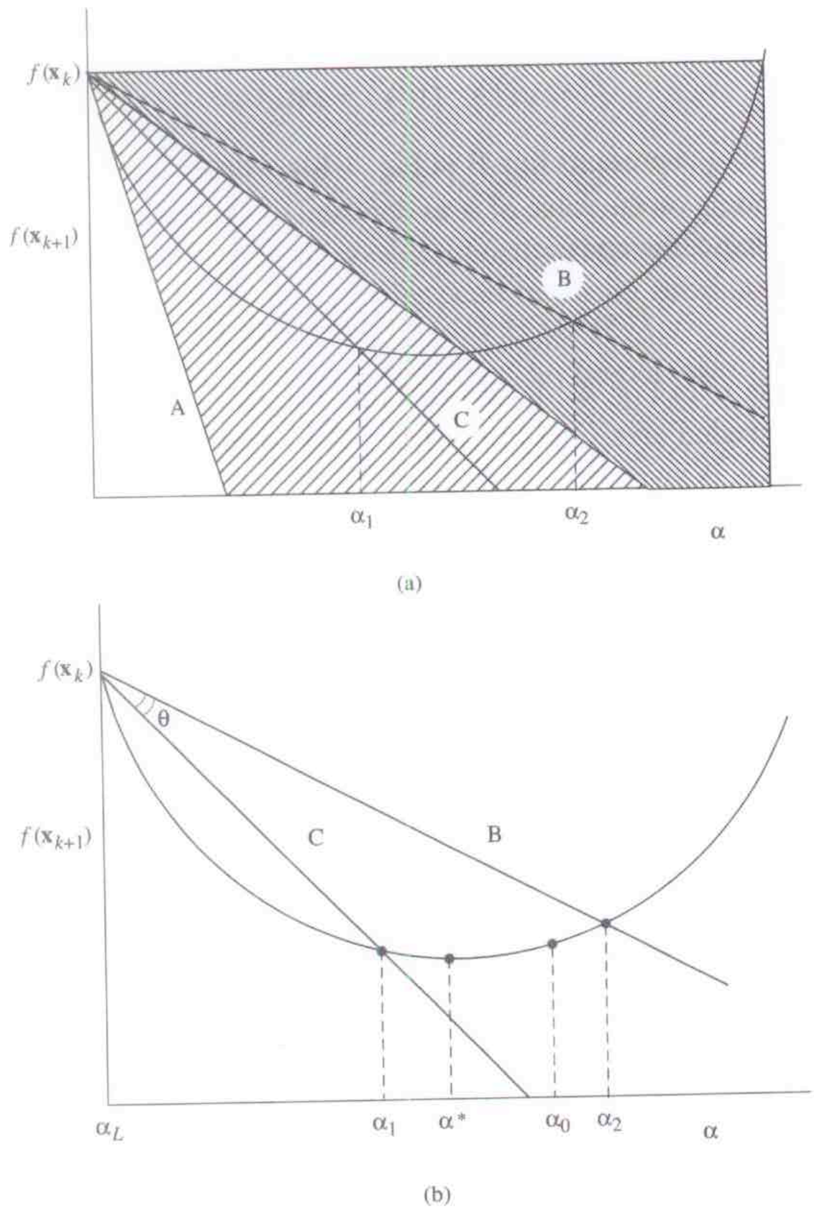

Figure 4.14. (a) The Goldstein tests. (b) Goldstein tests satisfied.

estimate $\breve{\alpha}_0$ can be determined by using interpolation. On the other hand, if Eq. (4.56) is violated, $\alpha_0 < \alpha_1$ as depicted in Fig. 4.15b, and since α_0 is likely to be in the range $\alpha_L < \alpha_0 < \alpha^*$, $\breve{\alpha}_0$ can be determined by using extrapolation.

If the value of $f(\mathbf{x}_k + \alpha\mathbf{d}_k)$ and its derivative with respect to α are known for $\alpha = \alpha_L$ and $\alpha = \alpha_0$, then for $\alpha_0 > \alpha_2$ a good estimate for $\breve{\alpha}_0$ can be deduced by using the interpolation formula

$$\breve{\alpha}_0 = \alpha_L + \frac{(\alpha_0 - \alpha_L)^2 f'_L}{2[f_L - f_0 + (\alpha_0 - \alpha_L)f'_L]} \tag{4.57}$$

and for

can be u

f_L

f_0

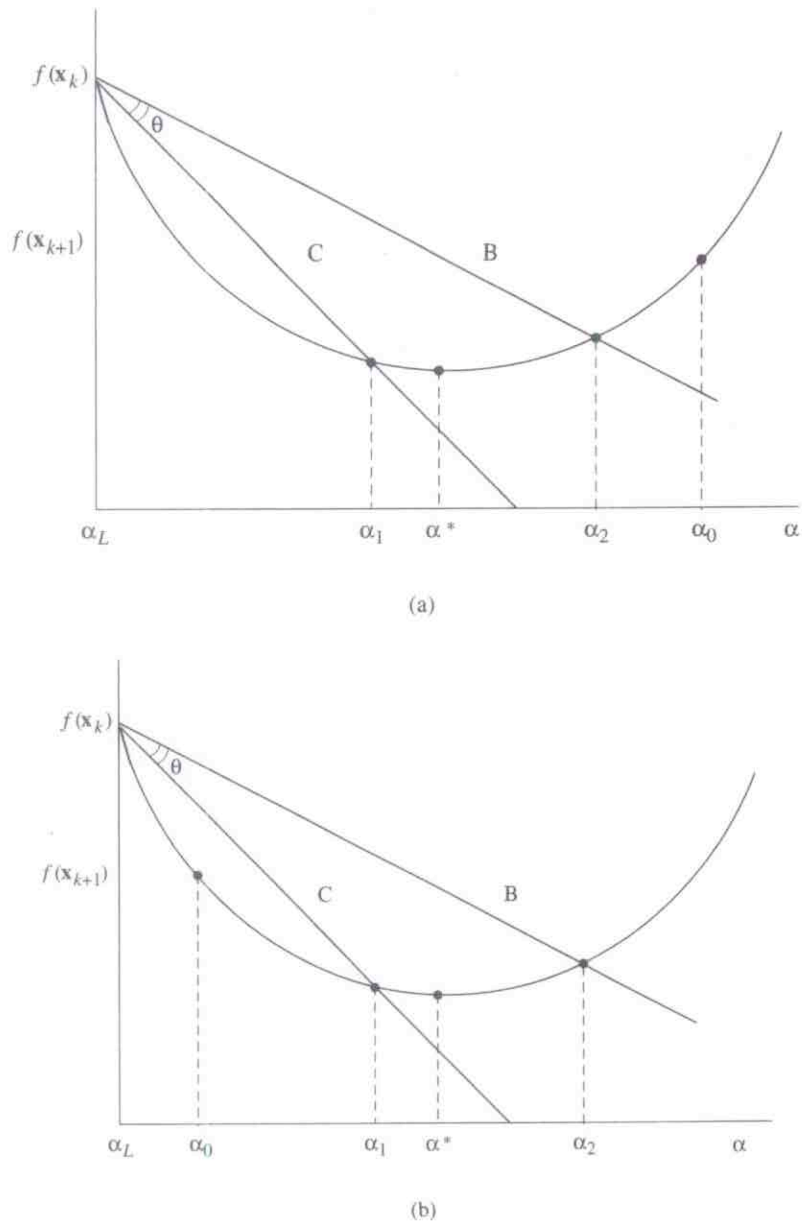

Figure 4.15. Goldstein tests violated: (a) with $\alpha_0 > \alpha_2$, (b) with $\alpha_0 < \alpha_1$.

and for $\alpha_0 < \alpha_1$ the extrapolation formula

$$\breve{\alpha}_0 = \alpha_0 + \frac{(\alpha_0 - \alpha_L) f_0'}{(f_L' - f_0')} \tag{4.58}$$

can be used, where

$$\begin{aligned} f_L &= f(\mathbf{x}_k + \alpha_L \mathbf{d}_k), \quad f_L' = f'(\mathbf{x}_k + \alpha_L \mathbf{d}_k) = \mathbf{g}(\mathbf{x}_k + \alpha_L \mathbf{d}_k)^T \mathbf{d}_k \\ f_0 &= f(\mathbf{x} + \alpha_0 \mathbf{d}_k), \quad f_0' = f'(\mathbf{x}_k + \alpha_0 \mathbf{d}_k) = \mathbf{g}(\mathbf{x}_k + \alpha_0 \mathbf{d}_k)^T \mathbf{d}_k \end{aligned}$$

hand, if
is likely
polation.
own for
deduced

(4.57)

(see Sec. 4.5).

Repeated application of the above procedure will eventually yield a value of $\breve{\alpha}_0$ such that $\alpha_1 < \breve{\alpha}_0 < \alpha_2$ and the inexact line search is terminated.

A useful theorem relating to the application of the Goldstein tests in an inexact line search is as follows:

Theorem 4.1 ***Convergence of inexact line search*** *If*

(a) $f(\mathbf{x}_k)$ *has a lower bound,*
(b) $\mathbf{g}_k$ *is uniformly continuous on set* $\{\mathbf{x} : \ f(\mathbf{x}) < f(\mathbf{x}_0)\}$,
(c) directions $\mathbf{d}_k$ *are not orthogonal to* $-\mathbf{g}_k$ *for all* k,

then a descent algorithm using an inexact line search based on Eqs. (4.55) and (4.56) will converge to a stationary point as $k \to \infty$.

The proof of this theorem is given by Fletcher [9]. The theorem does not guarantee that a descent algorithm will converge to a minimizer since a saddle point is also a stationary point. Nevertheless, the theorem is of importance since it demonstrates that inaccuracies due to the inexactness of the line search are not detrimental to convergence.

Conditions (a) and (b) of Theorem 4.1 are normally satisfied but condition (c) may be violated. Nevertheless, the problem can be avoided in practice by changing direction $\mathbf{d}_k$. For example, if θ_k is the angle between $\mathbf{d}_k$ and $-\mathbf{g}_k$ and

$$\theta_k = \cos^{-1} \frac{-\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\| \, \|\mathbf{d}_k\|} = \frac{\pi}{2}$$

then $\mathbf{d}_k$ can be modified slightly to ensure that

$$\theta_k = \frac{\pi}{2} - \mu$$

where $\mu > 0$.

The Goldstein conditions sometimes lead to the situation illustrated in Fig. 4.16, where α^* is not in the range $[\alpha_1, \ \alpha_2]$. Evidently, in such a case a value α_0 in the interval $[\alpha^*, \ \alpha_1]$ will not terminate the line search even though the reduction in $f(\mathbf{x}_k)$ would be larger than that for any α_0 in the interval $[\alpha_1, \ \alpha_2]$. Although the problem is not serious, since convergence is assured by Theorem 4.1, the amount of computation may be increased. The problem can be eliminated by replacing the second Goldstein condition, namely, Eq. (4.56), by the condition

$$\mathbf{g}_{k+1}^T \mathbf{d}_k \geq \sigma \mathbf{g}_k^T \mathbf{d}_k \tag{4.59}$$

where $0 < \sigma < 1$ and $\sigma \geq \rho$. This modification to the second Goldstein condition was proposed by Fletcher [10]. It is illustrated in Fig. 4.17. The scalar $\mathbf{g}_k^T \mathbf{d}_k$ is the derivative of $f(\mathbf{x}_k + \alpha \mathbf{d}_k)$ at $\alpha = 0$, and since $0 < \sigma < 1$,

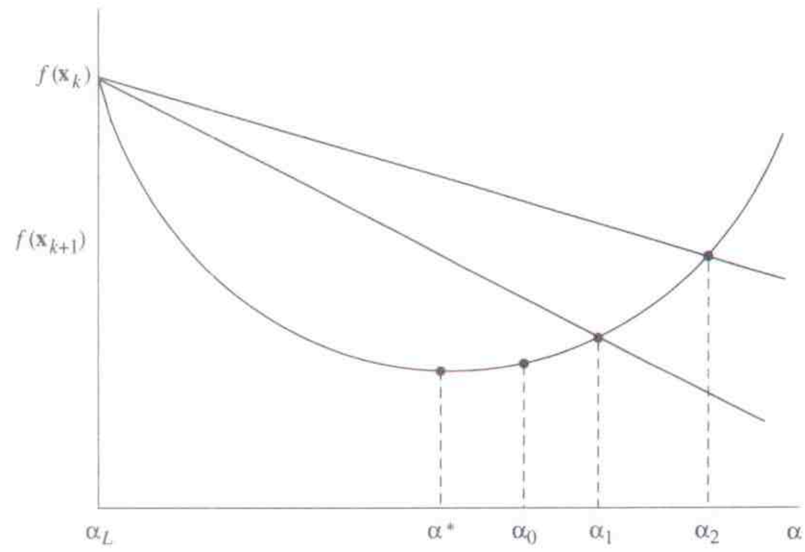

Figure 4.16. Goldstein tests violated with $\alpha^* < \alpha_1$.

$\sigma \mathbf{g}_k^T \mathbf{d}_k$ is the derivative of $f(\mathbf{x}_k + \alpha \mathbf{d}_k)$ at some value of α, say, α_1, such that $\alpha_1 < \alpha^*$. Now if the condition in Eq. (4.59) is satisfied at some point

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_0 \mathbf{d}_k$$

then the slope of $f(\mathbf{x}_k + \alpha \mathbf{d}_k)$ at $\alpha = \alpha_0$ is less negative (more positive) than the slope at $\alpha = \alpha_1$ and, consequently, we conclude that $\alpha_1 \leq \alpha_0$. Now if Eq. (4.55) is also satisfied, then we must have $\alpha_1 < (\alpha^* \text{ or } \alpha_0) < \alpha_2$, as depicted in Fig. 4.17.

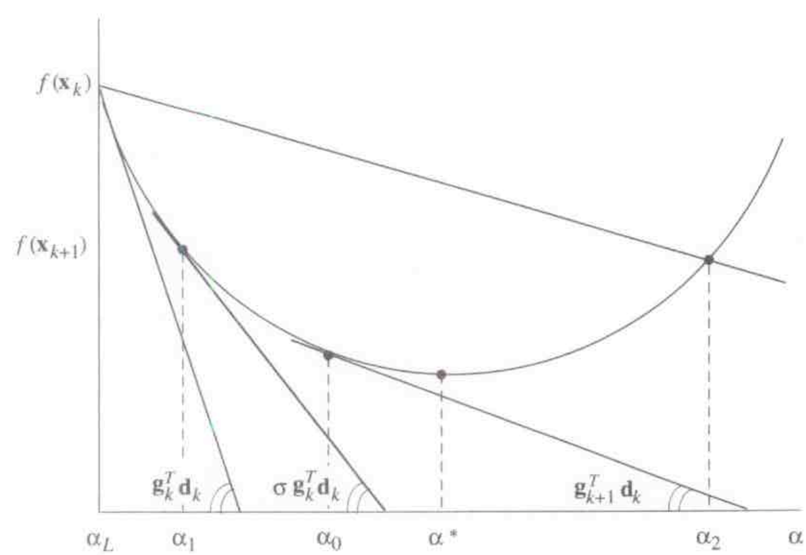

Figure 4.17. Fletcher's modification of the Goldstein tests.

The precision of a line search based on Eqs. (4.55) and (4.59) can be increased by reducing the value of σ. While $\sigma = 0.9$ results in a somewhat imprecise line search, the value $\sigma = 0.1$ results in a fairly precise line search. Note, however, that a more precise line search could slow down the convergence.

A disadvantage of the condition in Eq. (4.59) is that it does not lead to an exact line search as $\sigma \rightarrow 0$. An alternative condition that eliminates this problem is obtained by modifying the condition in Eq. (4.59) as

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| \leq -\sigma \mathbf{g}_k^T \mathbf{d}_k$$

In order to demonstrate that an exact line search can be achieved with the above condition, let us assume that $\mathbf{g}_k^T \mathbf{d}_k < 0$. If $\mathbf{g}_{k+1}^T \mathbf{d}_k < 0$, the line search will not terminate until

$$-|\mathbf{g}_{k+1}^T \mathbf{d}_k| \geq \sigma \mathbf{g}_k^T \mathbf{d}_k$$

and if $\mathbf{g}_{k+1}^T \mathbf{d}_k > 0$, the line search will not terminate until

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| \leq -\sigma \mathbf{g}_k^T \mathbf{d}_k \tag{4.60}$$

Now if $\sigma \mathbf{g}_k^T \mathbf{d}_k$, $\mathbf{g}_{k+1}^T \mathbf{d}_k$, and $-\sigma \mathbf{g}_k^T \mathbf{d}_k$ are the derivatives of $f(\mathbf{x}_k + \alpha \mathbf{d}_k)$ at points $\alpha = \alpha_1$, $\alpha = \alpha_0$, and $\alpha = \alpha_2$, respectively, we have $\alpha_1 \leq \alpha_0 \leq \alpha_2$ as depicted in Fig. 4.18. In effect, Eq. (4.60) overrides both of the Goldstein conditions in Eqs. (4.55) and (4.56). Since interval $[\alpha_1,\ \alpha_2]$ can be reduced as much as desired by reducing σ, it follows that α^* can be determined as accurately as desired, and as $\sigma \rightarrow 0$, the line search becomes exact. In such a case, the amount of computation would be comparable to that required by any other exact line search and the computational advantage of using an inexact line search would be lost.

An inexact line search based on Eqs. (4.55) and (4.59) due to Fletcher [10] is as follows:

Algorithm 4.6 Inexact line search
Step 1
Input $\mathbf{x}_k$, $\mathbf{d}_k$, and compute $\mathbf{g}_k$.
Initialize algorithm parameters ρ, σ, τ, and χ.
Set $\alpha_L = 0$ and $\alpha_U = 10^{99}$.
Step 2
Compute $f_L = f(\mathbf{x}_k + \alpha_L \mathbf{d}_k)$.
Compute $f_L' = \mathbf{g}(\mathbf{x}_k + \alpha_L \mathbf{d}_k)^T \mathbf{d}_k$.
Step 3
Estimate α_0.

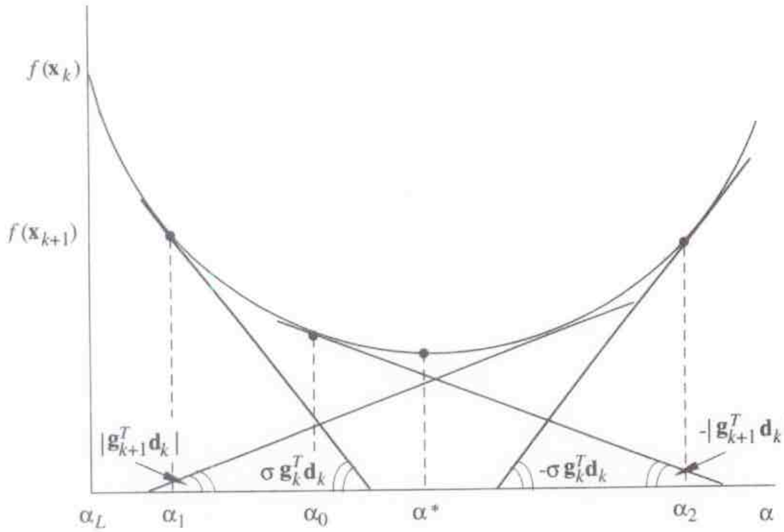

Figure 4.18. Conversion of inexact line search into an exact line search.

Step 4
Compute $f_0 = f(\mathbf{x}_k + \alpha_0\mathbf{d}_k)$.
Step 5 (Interpolation)
If $f_0 > f_L + \rho(\alpha_0 - \alpha_L)f'_L$, then do:
 a. If $\alpha_0 < \alpha_U$, then set $\alpha_U = \alpha_0$.
 b. Compute $\breve{\alpha}_0$ using Eq. (4.57).
 c. If $\breve{\alpha}_0 < \alpha_L + \tau(\alpha_U - \alpha_L)$ then set $\breve{\alpha}_0 = \alpha_L + \tau(\alpha_U - \alpha_L)$.
 d. If $\breve{\alpha}_0 > \alpha_U - \tau(\alpha_U - \alpha_L)$ then set $\breve{\alpha}_0 = \alpha_U - \tau(\alpha_U - \alpha_L)$.
 e. Set $\alpha_0 = \breve{\alpha}_0$ and go to Step 4.
Step 6
Compute $f'_0 = \mathbf{g}(\mathbf{x}_k + \alpha_0\mathbf{d}_k)^T\mathbf{d}_k$.
Step 7 (Extrapolation)
If $f'_0 < \sigma f'_L$, then do:
 a. Compute $\Delta\alpha_0 = (\alpha_0 - \alpha_L)f'_0/(f'_L - f'_0)$ (see Eq. (4.58)).
 b. If $\Delta\alpha_0 < \tau(\alpha_0 - \alpha_L)$, then set $\Delta\alpha_0 = \tau(\alpha_0 - \alpha_L)$.
 c. If $\Delta\alpha_0 > \chi(\alpha_0 - \alpha_L)$, then set $\Delta\alpha_0 = \chi(\alpha_0 - \alpha_L)$.
 d. Compute $\breve{\alpha}_0 = \alpha_0 + \Delta\alpha_0$.
 e. Set $\alpha_L = \alpha_0$, $\alpha_0 = \breve{\alpha}_0$, $f_L = f_0$, $f'_L = f'_0$, and go to Step 4.
Step 8
Output α_0 and $f_0 = f(\mathbf{x}_k + \alpha_0\mathbf{d}_k)$, and stop.

The precision to which the minimizer is determined depends on the values of ρ and σ. Small values like $\rho = \sigma = 0.1$ will yield a relatively precise line search whereas values like $\rho = 0.3$ and $\sigma = 0.9$ will yield a somewhat imprecise line search. The values $\rho = 0.1$ and $\sigma = 0.7$ give good results.

An estimate of α_0 in Step 3 can be determined by assuming that $f(\mathbf{x})$ is a convex quadratic function and using $\alpha_0 = \|\mathbf{g}_0\|^2/(\mathbf{g}_0^T\mathbf{H}_0\mathbf{g}_0)$ which is the minimum point for a convex quadratic function.

In Step 5, $\breve{\alpha}_0$ is checked and if necessary it is adjusted through a series of interpolations to ensure that $\alpha_L < \breve{\alpha}_0 < \alpha_U$. A suitable value for τ is 0.1. This assures that $\breve{\alpha}_0$ is no closer to α_L or α_U than 10 percent of the permissible range. A similar check is applied in the case of extrapolation, as can be seen in Step 7. The value for χ suggested by Fletcher is 9.

The algorithm maintains a running bracket (or range of uncertainty) $[\alpha_L,\ \alpha_U]$ that contains the minimizer which is initially set to $[0,\ 10^{99}]$ in Step 1. This is gradually reduced by reducing α_U in Step $5a$ and increasing α_L in Step $7e$.

In Step $7e$, known data that can be used in the next iteration are saved, i.e., α_0, f_0, and f_0' become α_L, f_L, and f_L', respectively. This keeps the amount of computation to a minimum.

Note that the Goldstein condition in Eq. (4.55) is modified as in Step 5 to take into account the fact that α_L assumes a value greater than zero when extrapolation is applied at least once.

References

1 D. M. Himmelblau, *Applied Nonlinear Programming*, McGraw-Hill, New York, 1972.
2 B. S. Gottfried and J. Weisman, *Introduction to Optimization Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1973.
3 P. R. Adby and M. A. H. Dempster, *Introduction to Optimization Methods*, Chapman and Hall, London, 1974.
4 C. S. Beightler, D. T. Phillips, and D. J. Wilde, *Foundations of Optimization*, Prentice-Hall, Englewood Cliffs, N.J., 1979.
5 M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming, Theory and Algorithms*, Wiley, New York, 1979.
6 P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, 1981.
7 G. P. McCormick, *Nonlinear Programming*, Wiley, New York, 1983.
8 M. J. Box, D. Davies, and W. H. Swann, *Nonlinear Optimization Techniques*, Oliver and Boyd, London, 1969.
9 R. Fletcher, *Practical Methods of Optimization*, 2nd ed., Wiley, New York, 1987.
10 R. Fletcher, *Practical Methods of Optimization*, vol. 1, Wiley, New York, 1980.

Problems

4.1 (a) Assuming that the ratio of two consecutive Fibonacci numbers, F_{k-1}/F_k, converges to a finite limit α, use Eq. (4.4) to show that

$$\lim_{k\to\infty} \frac{F_{k-1}}{F_k} = \alpha = \frac{2}{\sqrt{5}+1} \approx 0.6180$$

(b) Use MATLAB to verify the value of α in part (a).