Chapter 8

Unconstrained Optimization

Unconstrained optimization deals with the problem of minimizing or maximizing a function in the absence of any restrictions. In this chapter we discuss both the minimization of a function of one variable and a function of several variables. Even though most practical optimization problems have side restrictions that must be satisfied, the study of techniques for unconstrained optimization is important for several reasons. Many algorithms solve a constrained problem by converting it into a sequence of unconstrained problems via Lagrangian multipliers, as illustrated in Chapter 6, or via penalty and barrier functions, as will be discussed in more detail in Chapter 9. Furthermore, another class of methods proceeds by finding a direction and then minimizing along this direction. This line search is equivalent to minimizing a function of one variable without constraints or with simple constraints, such as lower and upper bounds on the variable. Finally, several unconstrained optimization techniques can be extended in a natural way to provide and motivate solution procedures for constrained problems.

The following is an outline of the chapter.

SECTION 8.1: **Line Search Without Using Derivatives** We discuss several procedures for minimizing strictly quasiconvex functions of $\boxed{\text{one}}$ variable without using derivatives. Uniform search, dichotomous search, the golden section method and the Fibonacci method are covered.

SECTION 8.2: **Line Search Using Derivatives** Differentiability is assumed, and the bisecting search method and Newton's method are discussed.

SECTION 8.3: **Closedness of the Line Search Algorithmic Map** We show that the line search algorithmic map is closed, a property that is essential in convergence analysis. Readers who are not interested in convergence may skip this section.

SECTION 8.4: **Multidimensional Search Without Using derivatives** The cyclic coordinate method, the method of Hooke and Jeeves, and Rosenbrock's method are discussed. Convergence of these methods is also established.

SECTION 8.5: **Multidimensional Search Using Derivatives** We develop the steepest descent method and the method of Newton and show their convergence.

SECTION 8.6: **Methods Using Conjugate Directions** The important concept of conjugacy is introduced. If the objective function is quadratic, then methods using conjugate directions are shown to converge in a finite number of steps. The Davidon-Fletcher-Powell method, the conjugate gradient method of Fletcher and Reeves, and the method of Zangwill are covered and their convergence established.

8.1 Line Search Without Using Derivatives

One-dimensional search is the backbone of many algorithms for solving a nonlinear programming problem. Many nonlinear programming algorithms proceed as follows. Given a point $\mathbf{x}_k$, find a direction vector $\mathbf{d}_k$ and then a suitable step size λ_k, yielding a new point $\mathbf{x}_{k+1}=\mathbf{x}_k+\lambda_k\mathbf{d}_k$; the process is then repeated. Finding the step size λ_k involves solving the subproblem to minimize $f(\mathbf{x}_k+\lambda\mathbf{d}_k)$, which is a one-dimensional search problem in the variable λ. The minimization may be over all real λ, nonnegative λ, or λ such that $\mathbf{x}_k+\lambda\mathbf{d}_k$ is feasible.

Consider a function θ of one variable λ to be minimized. One approach to minimizing θ is to let the derivative θ' be equal to 0 and then solve for λ. Note, however, that θ is usually defined implicitly in terms of a function f of several variables. In particular, given the vectors $\mathbf{x}$ and $\mathbf{d}$, $\theta(\lambda)=f(\mathbf{x}+\lambda\mathbf{d})$. If f is not differentiable, then θ will not be differentiable. If f is differentiable, then $\theta'(\lambda)=\mathbf{d}^t\nabla f(\mathbf{x}+\lambda\mathbf{d})$. Therefore to find a point λ with $\theta'(\lambda)=0$, we have to solve the equation $\mathbf{d}^t\nabla f(\mathbf{x}+\lambda\mathbf{d})=0$, which is usually nonlinear in λ. Furthermore, λ satisfying $\theta'(\lambda)=0$ is not necessarily a minimum; it may be a local minimum, a local maximum, or even a saddle point. For these reasons, and except for some special cases, we avoid minimizing θ by letting its derivative be equal to zero. Instead, we resort to some numerical techniques for minimizing the function θ.

In this section we discuss several methods that do not use derivatives for minimizing a function θ of one variable over a closed bounded interval. These methods fall under the categories of simultaneous line search and sequential line search problems. In the former case, the readings are determined a priori, whereas in the sequential search, the values of the function at the previous iterations are used to determine the succeeding readings.

The Interval of Uncertainty

Consider the line search problem to minimize $\theta(\lambda)$ subject to $a \le \lambda \le b$. Since the exact location of the minimum of θ over $[a, b]$ is not known, this interval is called the *interval of uncertainty*. During the search procedure if we could exclude portions of this interval that do not contain the minimum, then the interval of uncertainty is reduced. In general, $[a, b]$ is called the interval of uncertainty, if a minimum point $\bar{\lambda}$ lies in $[a, b]$, though its exact value is not known.

Theorem 8.1.1 below shows that if the function θ is strictly quasiconvex, then the interval of uncertainty could be reduced by evaluating θ at two points within the interval.

8.1.1 Theorem

Let $\theta: E_1 \to E_1$ be strictly quasiconvex over the interval $[a, b]$. Let $\lambda, \mu \in [a, b]$ be such that $\lambda < \mu$. If $\theta(\lambda) > \theta(\mu)$, then $\theta(z) \ge \theta(\mu)$ for all $z \in [a, \lambda)$. If $\theta(\lambda) \le \theta(\mu)$, then $\theta(z) \ge \theta(\lambda)$ for all $z \in (\mu, b]$.

Proof

Suppose that $\theta(\lambda) > \theta(\mu)$, and let $z \in [a, \lambda)$. By contradiction suppose that $\theta(z) < \theta(\mu)$. Since λ could be written as a convex combination of z and μ, and by strict quasiconvexity of θ, we have

$$\theta(\lambda) < \text{maximum}\,\{\theta(z), \theta(\mu)\} = \theta(\mu)$$

contradicting $\theta(\lambda) > \theta(\mu)$. Hence $\theta(z) \ge \theta(\mu)$. The second part of the theorem can be proved similarly.

From the above theorem, under strict quasiconvexity if $\theta(\lambda) > \theta(\mu)$, the new interval of uncertainty is $[\lambda, b]$. On the other hand, if $\theta(\lambda) \le \theta(\mu)$, the new interval of uncertainty is $[a, \mu]$. These two cases are illustrated in Figure 8.1.

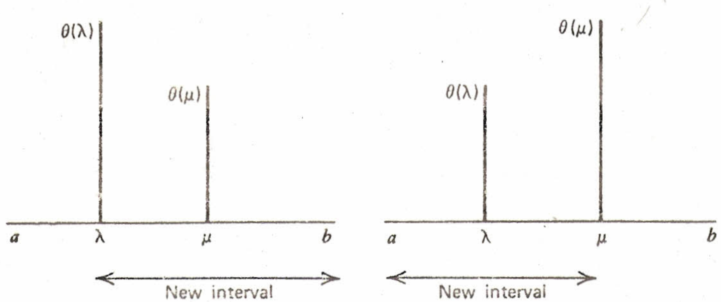


Figure 8.1 Reducing the interval of uncertainty.

Literature on nonlinear programming frequently uses the concept of *unimodality* of θ to reduce the interval of uncertainty. In this book we are using the equivalent concept of strict quasiconvexity. In Exercise 8.4 the definition of unimodality is given and the relationship with strict quasiconvexity is stated.

We now present several procedures for minimizing a strictly quasiconvex function over a closed bounded interval by iteratively reducing the interval of uncertainty.

An Example of a Simultaneous Search: Uniform Search

Uniform search is an example of simultaneous search, where we decide beforehand the points at which the functional evaluations are to be made. The interval of uncertainty $[a_1, b_1]$ is divided into smaller sub-intervals via the *grid points* $a_1 + k\delta$ for $k = 1, \ldots, n$ where $b_1 = a_1 + (n+1)\delta$, as illustrated in Figure 8.2. The function θ is evaluated at each of the n grid points. Let $\hat{\lambda}$ be a grid point with the smallest value of θ. If θ is strictly quasiconvex, it follows that a minimum of θ lies in the interval $[\hat{\lambda} - \delta, \hat{\lambda} + \delta]$.

The Choice of the Grid Length δ

We see that the interval of uncertainty $[a_1, b_1]$ is reduced, after n functional evaluations to an interval of length 2δ. Noting that $n = [(b_1 - a_1)/\delta] - 1$, if we desire a small final interval of uncertainty, then a large number of function evaluations n must be made. One technique that is often used to reduce the computational effort is to utilize a large grid size first and then switch to a finer grid size.

tamanho

Sequential Search

As may be expected, more efficient procedures which utilize the information generated at the previous iterations in placing the subsequent reading, could be

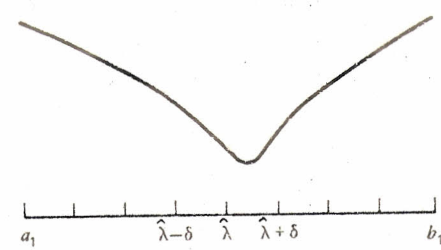


Figure 8.2 Uniform search.

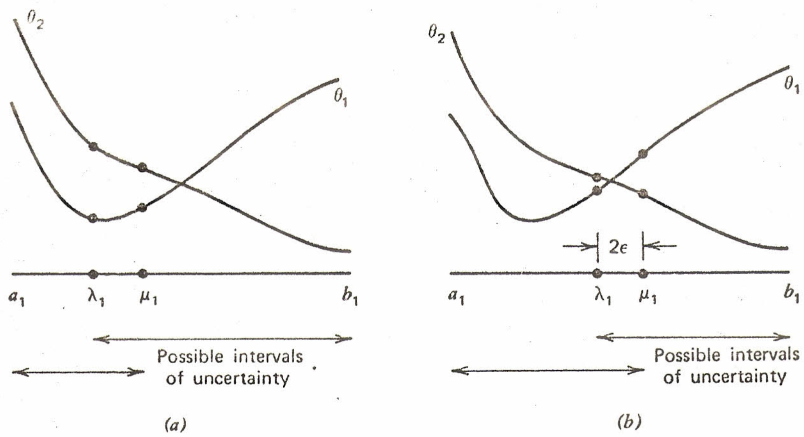


Figure 8.3 Possible intervals of uncertainty

devised. Here we discuss the following sequential search procedures: dichotomous search, the golden section method, and the Fibonacci method.

Dichotomous Search

Consider $\theta: E_1 \to E_1$ to be minimized over the interval $[a_1, b_1]$. Suppose that θ is strictly quasiconvex. Obviously, the smallest number of functional evaluations that is needed to reduce the interval of uncertainty is two. In Figure 8.3 we consider the location of the two readings λ_1 and μ_1. In Figure 8.3*a*, for $\theta = \theta_1$, note that $\theta(\lambda_1) < \theta(\mu_1)$, and hence by Theorem 8.1.1, the new interval of uncertainty is $[a_1, \mu_1]$. However, for $\theta = \theta_2$, note that $\theta(\lambda_1) > \theta(\mu_1)$, and hence by Theorem 8.1.1 the new interval of uncertainty is $[\lambda_1, b_1]$. Thus depending on the function θ, the length of the new interval of uncertainty is equal to $\mu_1 - a_1$ or $b_1 - \lambda_1$.

Note, however, that we do not know, a priori, whether $\theta(\lambda_1) < \theta(\mu_1)$ or $\theta(\lambda_1) > \theta(\mu_1)$.* Thus the *optimal strategy* is to place λ_1 and μ_1 in such a way to guard against the worst possible outcome, that is, to minimize the maximum of $\mu_1 - a_1$ and $b_1 - \lambda_1$. This could be accomplished by placing λ_1 and μ_1 at the midpoint of the interval $[a_1, b_1]$. If we do this, however, we would only have one reading and would not be able to reduce the interval of uncertainty. Therefore, as shown in Figure 8.3*b*, λ_1 and μ_1 are placed symmetrically, each at a distance $\varepsilon > 0$ from the midpoint. Here $\varepsilon > 0$ is a scalar that is sufficiently small so that the new length of uncertainty $\varepsilon + (b_1 - a_1)/2$ is close enough to the

* If the equality $\theta(\lambda_1) = \theta(\mu_1)$ is true, then the interval of uncertainty can be further reduced to $[\lambda_1, \mu_1]$. It may be noted, however, that exact equality is quite unlikely to occur in practice.

theoretical optimal value of $(b_1 - a_1)/2$ and, in the meantime, would make the functional evaluations $\theta(\lambda_1)$ and $\theta(\mu_1)$ distinguishable.

In dichotomous search, we place each of the first two observations, λ_1 and μ_1, symmetrically at a distance ε from the midpoint $(a_1 + b_1)/2$. Depending on the values of θ at λ_1 and μ_1, a new interval of uncertainty is obtained. The process is then repeated by placing two new observations.

Summary of the Dichotomous Search Method

The following is a summary of the dichotomous method of minimizing a strictly quasiconvex function θ over the interval $[a_1, b_1]$.

Initialization Step Choose the distinguishability constant, $2\varepsilon > 0$, and the allowable final length of uncertainty, $l > 0$. Let $[a_1, b_1]$ be the initial interval of uncertainty, let $k = 1$, and go to the main step.

Main Step

1. If $b_k - a_k < l$, stop; the minimum points lies in the interval $[a_k, b_k]$. Otherwise, consider λ_k and μ_k defined below, and go to step 2.

$$\lambda_k = \frac{a_k + b_k}{2} - \varepsilon \qquad \mu_k = \frac{a_k + b_k}{2} + \varepsilon$$

2. If $\theta(\lambda_k) < \theta(\mu_k)$, let $a_{k+1} = a_k$ and $b_{k+1} = \mu_k$. Otherwise, let $a_{k+1} = \lambda_k$ and $b_{k+1} = b_k$. Replace k by $k+1$, and go to step 1.

Note that the length of uncertainty at the beginning of iteration $k+1$ is given by

$$(b_{k+1} - a_{k+1}) = \frac{1}{2^k}(b_1 - a_1) + 2\varepsilon\left(1 - \frac{1}{2^k}\right)$$

This formula can be used to determine the number of iterations needed to achieve the desired accuracy. Since each iteration requires two readings, the formula can also be used to determine the number of readings.

The Golden Section Method

In order to compare the various line search procedures, the following reduction ratio will be of use.

$$\frac{\text{length of interval of uncertainty after } \nu \text{ observations are taken}}{\text{length of interval of uncertainty before taking the observations}}$$

Obviously, more efficient schemes correspond to small ratios. In dichotomous search, the above reduction ratio is approximately $(0.5)^{\nu/2}$. We now describe the more efficient golden section method, whose reduction ratio is given by $(0.618)^{\nu-1}$. The method could be used for minimizing a strictly quasiconvex function.

At a general iteration k of the golden section method, let the interval of uncertainty be $[a_k, b_k]$. By Theorem 8.1.1 the new interval of uncertainty $[a_{k+1}, b_{k+1}]$ is given by $[\lambda_k, b_k]$ if $\theta(\lambda_k) > \theta(\mu_k)$ and by $[a_k, \mu_k]$ if $\theta(\lambda_k) \leq \theta(\mu_k)$. The points λ_k and μ_k are selected such that

1. The length of the new interval of uncertainty $b_{k+1} - a_{k+1}$ does not depend upon the outcome of the kth iteration, that is, on whether $\theta(\lambda_k) > \theta(\mu_k)$ or $\theta(\lambda_k) \leq \theta(\mu_k)$. Therefore, we must have $b_k - \lambda_k = \mu_k - a_k$. Thus, if λ_k is of the form

$$\lambda_k = a_k + (1-\alpha)(b_k - a_k) \tag{8.1}$$

where $\alpha \in (0, 1)$, then μ_k must be of the form

$$\mu_k = a_k + \alpha(b_k - a_k) \tag{8.2}$$

so that

$$b_{k+1} - a_{k+1} = \alpha(b_k - a_k).$$

2. As λ_{k+1} and μ_{k+1} are selected for the purpose of new iteration, either λ_{k+1} coincides with μ_k or μ_{k+1} coincides with λ_k. If this could be realized, then during iteration $k+1$, only one extra observation is needed. To illustrate, consider Figure 8.4 and the following two cases.

Case 1: $\theta(\lambda_k) > \theta(\mu_k)$

In this case, $a_{k+1} = \lambda_k$ and $b_{k+1} = b_k$. In order to satisfy $\lambda_{k+1} = \mu_k$, and applying (8.1) with k replaced by $k+1$, we get

$$\mu_k = \lambda_{k+1} = a_{k+1} + (1-\alpha)(b_{k+1} - a_{k+1}) = \lambda_k + (1-\alpha)(b_k - \lambda_k)$$

Substituting the expressions of λ_k and μ_k from (8.1) and (8.2) into the above equation, we get $\alpha^2 + \alpha - 1 = 0$.

Case 2: $\theta(\lambda_k) \leq \theta(\mu_k)$

In this case, $a_{k+1} = a_k$ and $b_{k+1} = \mu_k$. In order to satisfy $\mu_{k+1} = \lambda_k$, and applying (8.2) with k replaced by $k+1$, we get

$$\lambda_k = \mu_{k+1} = a_{k+1} + \alpha(b_{k+1} - a_{k+1}) = a_k + \alpha(\mu_k - a_k)$$

Noting (8.1) and (8.2), the above equation gives $\alpha^2 + \alpha - 1 = 0$. The roots of the equation $\alpha^2 + \alpha - 1 = 0$ are $\alpha \cong 0.618$ and $\alpha \cong -1.618$. Since α must be in the interval (0, 1) then $\alpha \cong 0.618$. To summarize, if at iteration k, μ_k

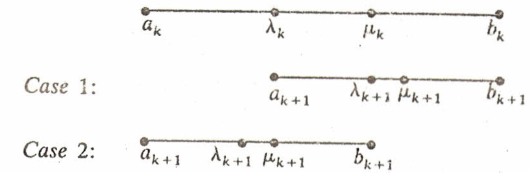


Figure 8.4 Illustration of the golden section rule.

and λ_k are chosen according to (8.1) and (8.2), where $\alpha = 0.618$, then the interval of uncertainty is reduced by a factor of 0.618. At the first iteration, two readings are needed at λ_1 and μ_1, but at each subsequent iteration, only one evaluation is needed, since either $\lambda_{k+1} = \mu_k$ or $\mu_{k+1} = \lambda_k$.

Summary of the Golden Section Method

The following is a summary of the golden section method for minimizing a strictly quasiconvex function over the interval $[a_1, b_1]$.

Initialization Step Choose an allowable final length of uncertainty $l > 0$. Let $[a_1, b_1]$ be the initial interval of uncertainty, and let $\lambda_1 = a_1 + (1-\alpha)(b_1 - a_1)$ and $\mu_1 = a_1 + \alpha(b_1 - a_1)$, where $\alpha = 0.618$. Evaluate $\theta(\lambda_1)$ and $\theta(\mu_1)$, let $k = 1$, and go to the main step.

Main Step

1. If $b_k - a_k < l$, stop; the optimal solution lies in the interval $[a_k, b_k]$. Otherwise, if $\theta(\lambda_k) > \theta(\mu_k)$, go to step 2, and if $\theta(\lambda_k) \leq \theta(\mu_k)$, go to step 3.
2. Let $a_{k+1} = \lambda_k$ and $b_{k+1} = b_k$. Furthermore, let $\lambda_{k+1} = \mu_k$, and let $\mu_{k+1} = a_{k+1} + \alpha(b_{k+1} - a_{k+1})$. Evaluate $\theta(\mu_{k+1})$, and go to step 4.
3. Let $a_{k+1} = a_k$, and $b_{k+1} = \mu_k$. Furthermore, let $\mu_{k+1} = \lambda_k$, and let $\lambda_{k+1} = a_{k+1} + (1-\alpha)(b_{k+1} - a_{k+1})$. Evaluate $\theta(\lambda_{k+1})$, and go to step 4.
4. Replace k by $k+1$, and go to step 1.

8.1.2 Example

Consider the following problem:

$l = 0.2$

Minimize $\quad \lambda^2 + 2\lambda$

subject to $\quad -3 \leq \lambda \leq 5$

Clearly the function θ to be minimized is strictly quasiconvex, and the initial interval of uncertainty is of length 8. We reduce this interval of uncertainty to one whose length is at most 0.2. The first two observations are located at

$$\lambda_1 = -3 + 0.382(8) = 0.056 \qquad \mu_1 = -3 + 0.618(8) = 1.944$$

TABLE 8.1 **Summary of Computations for the Golden Section Method**

Iteration k	a_k	b_k	λ_k	μ_k	$\theta(\lambda_k)$	$\theta(\mu_k)$
1	−3.000	5.000	0.056	1.944	0.115*	7.667*
2	−3.000	1.944	−1.112	0.056	−0.987*	0.115
3	−3.000	0.056	−1.832	−1.112	−0.308*	−0.987
4	−1.832	0.056	−1.112	−0.664	−0.987	−0.887*
5	−1.832	−0.664	−1.384	−1.112	−0.853*	−0.987
6	−1.384	−0.664	−1.112	−0.936	−0.987	−0.996*
7	−1.112	−0.664	−0.936	−0.840	−0.996	−0.974*
8	−1.112	−0.840	−1.016	−0.936	−1.000*	−0.996
9	−1.112	−0.936				

Note that $\theta(\lambda_1) < \theta(\mu_1)$. Hence, the new interval of uncertainty is $[-3, 1.944]$. The process is repeated, and the computations are summarized in Table 8.1. The values of θ that are computed at each iteration are indicated by an asterisk. After eight iterations involving nine observations, the interval of uncertainty is $[-1.112, -0.936]$, so that the minimum could be estimated to be the midpoint -1.024. Note that the true minimum is, in fact, -1.0.

The Fibonacci Search

The Fibonacci method is a line search procedure for minimizing a strictly quasiconvex function θ over a closed bounded interval. Similar to the golden section method, the Fibonacci search procedure makes two functional evaluations at the first iteration and then only one evaluation at each of the subsequent iterations. However, the procedure differs from the golden section method in that the reduction of the interval of uncertainty varies from one iteration to another.

The procedure is based on the Fibonacci sequence $\{F_\nu\}$ defined as follows:

$$F_{\nu+1} = F_\nu + F_{\nu-1} \qquad \nu = 1, 2, \ldots \tag{8.3}$$

$$F_0 = F_1 = 1$$

The sequence is therefore 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, ... At iteration k, suppose that the interval of uncertainty is $[a_k, b_k]$. Consider the two points λ_k and μ_k given below, where n is the total number of functional evaluations planned.

$$\lambda_k = a_k + \frac{F_{n-k-1}}{F_{n-k+1}}(b_k - a_k) \qquad k = 1, \ldots, n-1 \tag{8.4}$$

$$\mu_k = a_k + \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k) \qquad k = 1, \ldots, n-1 \tag{8.5}$$

By Theorem 8.1.1, the new interval of uncertainty $[a_{k+1}, b_{k+1}]$ is given by $[\lambda_k, b_k]$ if $\theta(\lambda_k) > \theta(\mu_k)$ and is given by $[a_k, \mu_k]$ if $\theta(\lambda_k) \le \theta(\mu_k)$. In the former case, noting (8.4) and letting $\nu = n-k$ in (8.3), we get

$$b_{k+1} - a_{k+1} = b_k - \lambda_k = b_k - a_k - \frac{F_{n-k-1}}{F_{n-k+1}}(b_k - a_k) = \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k) \tag{8.6}$$

In the latter case, noting (8.5) we get

$$b_{k+1} - a_{k+1} = \mu_k - a_k = \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k) \tag{8.7}$$

Thus, in either case, the interval of uncertainty is reduced by the factor F_{n-k}/F_{n-k+1}.

We now show that at iteration $k+1$, either $\lambda_{k+1} = \mu_k$ or $\mu_{k+1} = \lambda_k$, so that only one functional evaluation is needed. Suppose that $\theta(\lambda_k) > \theta(\mu_k)$. Then, by Theorem 8.1.1, $a_{k+1} = \lambda_k$ and $b_{k+1} = b_k$. Thus, applying (8.4) with k replaced by $k+1$, we get

$$\lambda_{k+1} = a_{k+1} + \frac{F_{n-k-2}}{F_{n-k}}(b_{k+1} - a_{k+1})$$

$$= \lambda_k + \frac{F_{n-k-2}}{F_{n-k}}(b_k - \lambda_k)$$

Substituting for λ_k from (8.4), we get

$$\lambda_{k+1} = a_k + \frac{F_{n-k-1}}{F_{n-k+1}}(b_k - a_k) + \frac{F_{n-k-2}}{F_{n-k}}\left(1 - \frac{F_{n-k-1}}{F_{n-k+1}}\right)(b_k - a_k)$$

Letting $\nu = n-k$ in (8.3), it follows that $[1 - (F_{n-k-1}/F_{n-k+1})] = F_{n-k}/F_{n-k+1}$. Substituting in the above equation, we get

$$\lambda_{k+1} = a_k + \left(\frac{F_{n-k-1} + F_{n-k-2}}{F_{n-k+1}}\right)(b_k - a_k)$$

Now let $\nu = n-k-1$ in (8.3), and noting (8.5), it follows that

$$\lambda_{k+1} = a_k + \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k) = \mu_k$$

Similarly, if $\theta(\lambda_k) \le \theta(\mu_k)$, the reader can easily verify that $\mu_{k+1} = \lambda_k$. Thus, in either case, only one observation is needed at iteration $k+1$.

To summarize, at the first iteration, two observations are made, and at each subsequent iteration, only one observation is necessary. Thus, at the end of iteration $n-2$, we have completed $n-1$ functional evaluations. Further, for $k = n-1$, it follows from (8.4) and (8.5) that $\lambda_{n-1} = \mu_{n-1} = \frac{1}{2}(a_{n-1} + b_{n-1})$. Since either $\lambda_{n-1} = \mu_{n-2}$ or $\mu_{n-1} = \lambda_{n-2}$, theoretically no new observations are to be made at this stage. However, in order to further reduce the interval of uncertainty, the last observation is placed slightly to the right or to the left of the midpoint $\lambda_{n-1}=\mu_{n-1}$, so that $\frac{1}{2}(b_{n-1}-a_{n-1})$ is the length of the final interval of uncertainty $[a_n, b_n]$.

Choosing the Number of Observations

Unlike the dichotomous search method and the golden section procedure, the Fibonacci method requires that the total number of observations n be chosen beforehand. This is because the placement of the observations is given by (8.4) and (8.5) and hence is dependent on n. From (8.6) and (8.7), the length of the interval of uncertainty is reduced at iteration k by the factor F_{n-k}/F_{n-k+1}. Hence, at the end of $n-1$ iterations, where n total observations have been made, the length of the interval of uncertainty is reduced from b_1-a_1 to $b_n-a_n=(b_1-a_1)/F_n$. Therefore, n must be chosen such that $(b_1-a_1)/F_n$ reflects the accuracy required.

Summary of the Fibonacci Search Method

The following is a summary of the Fibonacci search method for minimizing a strictly quasiconvex function over the interval $[a_1, b_1]$.

Initialization Step Choose an allowable final length of uncertainty $l>0$ and a distinguishability constant $\varepsilon>0$. Let $[a_1, b_1]$ be the initial interval of uncertainty, and choose the number of observations n to be taken such that $F_n>(b_1-a_1)/l$. Let $\lambda_1=a_1+(F_{n-2}/F_n)(b_1-a_1)$, and $\mu_1=a_1+(F_{n-1}/F_n)(b_1-a_1)$. Evaluate $\theta(\lambda_1)$ and $\theta(\mu_1)$, let $k=1$, and go to the main step.

Main Step

1. If $\theta(\lambda_k)>\theta(\mu_k)$, go to step 2, and if $\theta(\lambda_k)\le\theta(\mu_k)$, go to step 3.
2. Let $a_{k+1}=\lambda_k$ and $b_{k+1}=b_k$. Furthermore, let $\lambda_{k+1}=\mu_k$, and let $\mu_{k+1}=a_{k+1}+(F_{n-k-1}/F_{n-k})(b_{k+1}-a_{k+1})$. If $k=n-2$, go to step 5; otherwise, evaluate $\theta(\mu_{k+1})$, and go to step 4.
3. Let $a_{k+1}=a_k$ and $b_{k+1}=\mu_k$. Furthermore, let $\mu_{k+1}=\lambda_k$, and let $\lambda_{k+1}=a_{k+1}+(F_{n-k-2}/F_{n-k})(b_{k+1}-a_{k+1})$. If $k=n-2$, go to step 5; otherwise, evaluate $\theta(\lambda_{k+1})$, and go to step 4.
4. Replace k by $k+1$, and go to step 1.
5. Let $\lambda_n=\lambda_{n-1}$, and $\mu_n=\lambda_{n-1}+\varepsilon$. If $\theta(\lambda_n)>\theta(\mu_n)$, let $a_n=\lambda_n$ and $b_n=b_{n-1}$. Otherwise, if $\theta(\lambda_n)\le\theta(\mu_n)$, let $a_n=a_{n-1}$ and $b_n=\lambda_n$. Stop; the optimal solution lies in the interval $[a_n, b_n]$.

8.1.3 Example

Consider the following problem:

Minimize $\lambda^2+2\lambda$
subject to $-3\le\lambda\le5$

Note that the function is strictly quasiconvex on the interval and that the true minimum occurs at $\lambda=-1$. We reduce the interval of uncertainty to one whose length is, at most, 0.2. Hence, we must have $F_n>8/0.2=40$, so that $n=9$. We adopt the distinguishability constant $\varepsilon=0.01$.

The first two observations are located at

$$\lambda_1=-3+\frac{F_7}{F_9}(8)=0.054545 \qquad \mu_1=-3+\frac{F_8}{F_9}(8)=1.945454$$

Note that $\theta(\lambda_1)<\theta(\mu_1)$. Hence the new interval of uncertainty is $[-3.000000, 1.945454]$. The process is repeated, and the computations are summarized in Table 8.2. The values of θ that are computed at each iteration are indicated by an asterisk. Note that at $k=8$, $\lambda_k=\mu_k=\lambda_{k-1}$, so that no functional evaluations are needed at this stage. For $k=9$, $\lambda_k=\lambda_{k-1}=-0.963636$ and $\mu_k=\lambda_k+\varepsilon=-0.953636$. Since $\theta(\mu_k)>\theta(\lambda_k)$, the final interval of uncertainty $[a_9, b_9]$ is $[-1.109091, -0.963636]$, whose length $l=0.145455$. We approximate the minimum to be the midpoint -1.036364. Note from Example 8.1.2 that with the same number of observations $n=9$, the golden section method gave a final interval of uncertainty whose length is 0.176.

Comparison of Derivative-Free Line Search Methods

Given a function θ that is strictly quasiconvex on the interval $[a_1, b_1]$, obviously each of the methods discussed in this section will yield a point λ in a finite

TABLE 8.2 **Summary of Computations for the Fibonacci Search Method**

Iteration k	a_k	b_k	λ_k	μ_k	$\theta(\lambda_k)$	$\theta(\mu_k)$
1	−3.000000	5.000000	0.054545	1.945454	0.112065*	7.675699*
2	−3.000000	1.945454	−1.109091	0.054545	−0.988099*	0.112065
3	−3.000000	0.054545	−1.836363	−1.109091	−0.300497*	−0.988099
4	−1.836363	0.054545	−1.109091	−0.672727	−0.988099	−0.892892*
5	−1.836363	−0.672727	−1.399999	−1.109091	−0.840001*	−0.988099
6	−1.399999	−0.672727	−1.109091	−0.963636	−0.988099	−0.998677*
7	−1.109091	−0.672727	−0.963636	−0.818182	−0.998677	−0.966942*
8	−1.109091	−0.818182	−0.963636	−0.963636	−0.998677	−0.998677
9	−1.109091	−0.963636	−0.963636	−0.953636	−0.998677	−0.997850*

number of steps such that $|\lambda - \bar{\lambda}| \leq l$, where l is the length of the final interval of uncertainty and $\bar{\lambda}$ is the minimum point over the interval. In particular, given the length l of the final interval of uncertainty, which reflects the desired degree of accuracy, the required number of observations n could be computed as the smallest positive integer satisfying the following relationships.

Uniform search method: $$n \geq \frac{b_1 - a_1}{l/2} - 1$$

Dichotomous search method: $$(1/2)^{n/2} \geq \frac{l}{b_1 - a_1}$$

Golden section method: $$(0.618)^{n-1} \geq \frac{l}{b_1 - a_1}$$

Fibonacci search method: $$F_n \geq \frac{b_1 - a_1}{l}$$

Eficiencia

From the above expressions, we see that the number of observations needed is a function of the ratio $(b_1 - a_1)/l$. Hence for a fixed ratio $(b_1 - a_1)/l$, the smaller the number of observations required, the more efficient the algorithm. It should be evident that the most efficient algorithm is the Fibonacci method, followed by the golden section procedure, the dichotomous search method, and finally the uniform search method.

Also note that for n large enough, $1/F_n$ is asymptotic to $(0.618)^{n-1}$, so that the Fibonacci search method and the golden section method are almost identical.

It is worth mentioning that among the derivative-free methods that minimize strict quasiconvex functions over a closed bounded interval, the Fibonacci search method is the most efficient in that it requires the smallest number of observations for a given reduction in the length of the interval of uncertainty.

General Functions

The procedures discussed above all rely on the strict quasiconvexity assumption. In many problems, this assumption does not hold true, and in any case, it cannot be easily verified. One way to handle this difficulty, especially if the initial interval of uncertainty is large, is to divide it into smaller intervals, find the minimum over each subinterval, and then choose the smallest of the minima over the subintervals.

8.2 Line Search Using Derivatives

In the previous section we discussed several line search procedures that use functional evaluations. In this section we discuss the bisecting search method and Newton's method, both of which need derivative information.

The Bisecting Search Method

Suppose that we wish to minimize a function θ over a closed and bounded interval. Furthermore, suppose that θ is pseudoconvex, and hence differentiable. At iteration k, let the interval of uncertainty be $[a_k, b_k]$. Suppose that the derivative $\theta'(\lambda_k)$ is known, and consider the following three possible cases:

1. If $\theta'(\lambda_k) = 0$, then by pseudoconvexity of θ, λ_k is a minimum point.
2. If $\theta'(\lambda_k) > 0$, then for $\lambda > \lambda_k$, we have $\theta'(\lambda_k)(\lambda - \lambda_k) > 0$, and by pseudoconvexity of θ, it follows that $\theta(\lambda) \geq \theta(\lambda_k)$. In other words, the minimum occurs to the left of λ_k, so that the new interval of uncertainty $[a_{k+1}, b_{k+1}]$ is given by $[a_k, \lambda_k]$.
3. If $\theta'(\lambda_k) < 0$, then for $\lambda < \lambda_k$, $\theta'(\lambda_k)(\lambda - \lambda_k) > 0$, so that $\theta(\lambda) \geq \theta(\lambda_k)$. Thus the minimum occurs to the right of λ_k, so that the new interval of uncertainty $[a_{k+1}, b_{k+1}]$ is given by $[\lambda_k, b_k]$.

The position of λ_k in the interval $[a_k, b_k]$ must be chosen so that the maximum possible length of the new interval of uncertainty is minimized. That is, λ_k must be chosen so as to minimize the maximum of $\lambda_k - a_k$ and $b_k - \lambda_k$. Obviously, the optimal location of λ_k is the midpoint $\frac{1}{2}(a_k + b_k)$.

To summarize, at any iteration k, θ' is evaluated at the midpoint of the interval of uncertainty. Based on the value of θ', we either stop or construct a new interval of uncertainty whose length is half that of the previous iteration. Note that this procedure is very similar to the dichotomous search method except that at each iteration, only one derivative evaluation is required, as opposed to two functional evaluations for the dichotomous search method.

Convergence of the Bisecting Search Method

Note that the length of the interval of uncertainty after n observations is equal to $(1/2)^n(b_1 - a_1)$, so that the method converges to a minimum point within any desired degree of accuracy. In particular, if the length of the final interval of uncertainty is fixed at l, then n must be chosen to be the smallest integer such that $(1/2)^n \leq l/(b_1 - a_1)$.

Summary of the Bisecting Search Method

We now summarize the bisecting search procedure for minimizing a pseudoconvex function θ over a closed and bounded interval.

Initialization Step Let $[a_1, b_1]$ be the initial interval of uncertainty, and let l be the allowable final interval of uncertainty. Let n be the smallest positive integer such that $(\frac{1}{2})^n \leq l/(b_1 - a_1)$. Let $k = 1$, and go to the main step.

Main Step

1. Let $\lambda_k = \frac{1}{2}(a_k + b_k)$, and evaluate $\theta'(\lambda_k)$. If $\theta'(\lambda_k)=0$, stop; λ_k is an optimal solution. Otherwise, go to step 2 if $\theta'(\lambda_k)>0$, and go to step 3 if $\theta'(\lambda_k)<0$.
2. Let $a_{k+1}=a_k$, and $b_{k+1}=\lambda_k$. Go to step 4.
3. Let $a_{k+1}=\lambda_k$, and $b_{k+1}=b_k$. Go to step 4.
4. If $k=n$, stop; the minimum lies in the interval $[a_{n+1}, b_{n+1}]$. Otherwise, replace k by $k+1$, and repeat step 1.

8.2.1 Example

Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & \lambda^2+2\lambda \\ \text{subject to} & -3\leq\lambda\leq 6 \end{array}$$

Suppose we want to reduce the interval of uncertainty to an interval whose length l is less than or equal to 0.2. Hence, the number of observations n satisfying $(\frac{1}{2})^n \leq l/(b_1-a_1)=0.2/9=0.0222$ is given by $n=6$. A summary of the computations using the bisecting search method is given in Table 8.3. Note that the final interval of uncertainty is $[-1.0313, -0.8907]$, so that the minimum could be taken as the midpoint -0.961.

Newton's Method

Newton's method is based on exploiting the quadratic approximation of the function θ at a given point λ_k. This quadratic approximation q is given by

$$q(\lambda)=\theta(\lambda_k)+\theta'(\lambda_k)(\lambda-\lambda_k)+\tfrac{1}{2}\theta''(\lambda_k)(\lambda-\lambda_k)^2$$

The point λ_{k+1} is taken to be the point where the derivative of q is equal to

TABLE 8.3 **Summary of Computations for the Bisecting Search Method**

Iteration k	a_k	b_k	λ_k	$\theta'(\lambda_k)$
1	−3.0000	6.0000	1.5000	5.0000
2	−3.0000	1.5000	−0.7500	0.5000
3	−3.0000	−0.7500	−1.8750	−1.7500
4	−1.8750	−0.7500	−1.3125	−0.6250
5	−1.3125	−0.7500	−1.0313	−0.0625
6	−1.0313	−0.7500	−0.8907	0.2186
7	−1.0313	−0.8907		

zero. This yields $\theta'(\lambda_k)+\theta''(\lambda_k)(\lambda_{k+1}-\lambda_k)=0$, so that

$$\lambda_{k+1}=\lambda_k-\frac{\theta'(\lambda_k)}{\theta''(\lambda_k)} \tag{8.8}$$

The procedure is terminated when $|\lambda_{k+1}-\lambda_k|<\varepsilon$ or when $|\theta'(\lambda_k)|<\varepsilon$ where ε is a prespecified termination scalar.

Note that the above procedure can only be applied for twice differentiable functions. Furthermore, the procedure is well defined only if $\theta''(\lambda_k)\neq 0$ for each k.

8.2.2 Example

Consider the function θ defined below:

$$\theta(\lambda)=\begin{cases} 4\lambda^3-3\lambda^4 & \text{if } \lambda\geq 0 \\ 4\lambda^3+3\lambda^4 & \text{if } \lambda<0 \end{cases}$$

Note that θ is twice differentiable everywhere. We apply Newton's method, starting from two different points. In the first case, $\lambda_1=0.40$, and as shown in Table 8.4, the procedure produced the point 0.002807 after six iterations. The reader can verify that the procedure indeed converges to the stationary point $\lambda=0$. In the second case, $\lambda_1=0.60$, and the procedure oscillates between the points 0.60 and −0.60, as shown in Table 8.5.

Convergence of Newton's Method

The method of Newton, in general, does not converge to a stationary point starting with an arbitrary initial point. The reason for this is that, in general, Theorem 7.2.3 cannot be applied as a result of the unavailability of a descent function. However, as shown in Theorem 8.2.3 below, if the starting point is

TABLE 8.4 **Summary of Computations for Newton's Method Starting from $\lambda_1=0.4$**

Iteration k	λ_k	$\theta'(\lambda_k)$	$\theta''(\lambda_k)$	λ_{k+1}
1	0.400000	1.152000	3.840000	0.100000
2	0.100000	0.108000	2.040000	0.047059
3	0.047059	0.025324	1.049692	0.022934
4	0.022934	0.006167	0.531481	0.011331
5	0.011331	0.001523	0.267322	0.005634
6	0.005634	0.000379	0.134073	0.002807

TABLE 8.5 **Summary of Computations for Newton's Method Starting from $\lambda_1 = 0.6$**

Iteration k	λ_k	$\theta'(\lambda_k)$	$\theta''(\lambda_k)$	λ_{k+1}
1	0.600	1.728	1.440	−0.600
2	−0.600	1.728	−1.440	0.600
3	0.600	1.728	1.440	−0.600
4	−0.600	1.728	−1.440	0.600

sufficiently close to a stationary point, then a suitable descent function could be devised so that the method converges.

8.2.3 Theorem

Let $\theta: E_1 \to E_1$ be twice continuously differentiable. Consider Newton's algorithm defined by the map $\mathbf{A}(\lambda) = \lambda - \theta'(\lambda)/\theta''(\lambda)$. Let $\bar{\lambda}$ be such that $\theta'(\bar{\lambda}) = 0$ and $\theta''(\bar{\lambda}) \neq 0$. Let the starting point λ_1 be sufficiently close to $\bar{\lambda}$ so that there exists $k_1, k_2 > 0$ with $k_1 k_2 < 1$ such that

1. $\dfrac{1}{|\theta''(\lambda)|} \leq k_1$

2. $\dfrac{|\theta'(\bar{\lambda}) - \theta'(\lambda) - \theta''(\lambda)(\bar{\lambda} - \lambda)|}{|\bar{\lambda} - \lambda|} \leq k_2$

for each λ satisfying $|\lambda - \bar{\lambda}| \leq |\lambda_1 - \bar{\lambda}|$. Then the algorithm converges to $\bar{\lambda}$.

Proof

Let the solution set $\Omega = \{\bar{\lambda}\}$, and let $X = \{\lambda : |\lambda - \bar{\lambda}| \leq |\lambda_1 - \bar{\lambda}|\}$. We prove convergence by using Theorem 7.2.3. Note that X is compact and that the map $\mathbf{A}$ is closed on X. We now show that $\alpha(\lambda) = |\lambda - \bar{\lambda}|$ is indeed a descent function. Now let $\lambda \in X$ and suppose that $\lambda \neq \bar{\lambda}$. Let $\hat{\lambda} \in \mathbf{A}(\lambda)$. Then, by definition of $\mathbf{A}$ and since $\theta'(\bar{\lambda}) = 0$, we get

$$\hat{\lambda} - \bar{\lambda} = (\lambda - \bar{\lambda}) - \frac{1}{\theta''(\lambda)}[\theta'(\lambda) - \theta'(\bar{\lambda})]$$

$$= \frac{1}{\theta''(\lambda)}[\theta'(\bar{\lambda}) - \theta'(\lambda) - \theta''(\lambda)(\bar{\lambda} - \lambda)]$$

Noting (1) and (2), it then follows that

$$|\hat{\lambda} - \bar{\lambda}| = \frac{1}{|\theta''(\lambda)|} \frac{|\theta'(\bar{\lambda}) - \theta'(\lambda) - \theta''(\lambda)(\bar{\lambda} - \lambda)|}{|\bar{\lambda} - \lambda|} |\lambda - \bar{\lambda}| \leq k_1 k_2 \, |\lambda - \bar{\lambda}| < |\lambda - \bar{\lambda}|$$

Therefore α is indeed a descent function, and the result follows immediately by the corollary to Theorem 7.2.3.

8.3 Closedness of the Line Search Algorithmic Map

In the previous two sections we discussed several procedures for minimizing a function of one variable. Since the one-dimensional search is a component of most nonlinear programming algorithms, we show in this section that line search procedures define a closed map.

Consider the line search problem to minimize $\theta(\lambda)$ subject to $\lambda \in L$, where $\theta(\lambda) = f(\mathbf{x} + \lambda\mathbf{d})$, and L is a closed interval in E_1. This line search problem can be defined by the algorithmic map $\mathbf{M}: E_n \times E_n \to E_n$ defined by

$$\mathbf{M}(\mathbf{x}, \mathbf{d}) = \{\mathbf{y} : \mathbf{y} = \mathbf{x} + \bar{\lambda}\mathbf{d} \text{ for some } \bar{\lambda} \in L, \text{ and } f(\mathbf{y}) \leq f(\mathbf{x} + \lambda\mathbf{d}) \text{ for each } \lambda \in L\}$$

Note that $\mathbf{M}$ is generally a point-to-set map because there can be more than one minimizing point $\mathbf{y}$. Theorem 8.3.1 below shows that the map $\mathbf{M}$ is closed. Thus, if the map $\mathbf{D}$ that determines the direction $\mathbf{d}$ is also closed, then by Theorem 7.3.2 the overall algorithmic map $\mathbf{A} = \mathbf{MD}$ is closed.

8.3.1 Theorem

Let $f: E_n \to E_1$, and let L be a closed interval in E_1. Consider the line search map $\mathbf{M}: E_n \times E_n \to E_n$ defined by

$$\mathbf{M}(\mathbf{x}, \mathbf{d}) = \{\mathbf{y} : \mathbf{y} = \mathbf{x} + \bar{\lambda}\mathbf{d} \text{ for some } \bar{\lambda} \in L, \text{ and } f(\mathbf{y}) \leq f(\mathbf{x} + \lambda\mathbf{d}) \text{ for each } \lambda \in L\}$$

If f is continuous at $\mathbf{x}$, and $\mathbf{d} \neq \mathbf{0}$, then $\mathbf{M}$ is closed at $(\mathbf{x}, \mathbf{d})$.

Proof

Suppose that $(\mathbf{x}_k, \mathbf{d}_k) \to (\mathbf{x}, \mathbf{d})$ and that $\mathbf{y}_k \to \mathbf{y}$, where $\mathbf{y}_k \in \mathbf{M}(\mathbf{x}_k, \mathbf{d}_k)$. We want to show that $\mathbf{y} \in \mathbf{M}(\mathbf{x}, \mathbf{d})$. First, note that $\mathbf{y}_k = \mathbf{x}_k + \lambda_k \mathbf{d}_k$, where $\lambda_k \in L$. Since $\mathbf{d} \neq \mathbf{0}$, then $\mathbf{d}_k \neq \mathbf{0}$ for k large enough, and hence, $\lambda_k = \|\mathbf{y}_k - \mathbf{x}_k\|/\|\mathbf{d}_k\|$. Taking the limit as $k \to \infty$, then $\lambda_k \to \bar{\lambda}$, where $\bar{\lambda} = \|\mathbf{y} - \mathbf{x}\|/\|\mathbf{d}\|$, and hence, $\mathbf{y} = \mathbf{x} + \bar{\lambda}\mathbf{d}$. Furthermore, since $\lambda_k \in L$ for each k, and since L is closed, then $\bar{\lambda} \in L$. Now let $\lambda \in L$ and note that $f(\mathbf{y}_k) \leq f(\mathbf{x}_k + \lambda\mathbf{d}_k)$. Taking the limit as $k \to \infty$ and noting the continuity of f, we conclude that $f(\mathbf{y}) \leq f(\mathbf{x} + \lambda\mathbf{d})$. Thus, $\mathbf{y} \in \mathbf{M}(\mathbf{x}, \mathbf{d})$, and the proof is complete.

In nonlinear programming, line search is typically perfomed over one of the following intervals

$L = \{\lambda : \lambda \in E_1\}$

$L = \{\lambda : \lambda \geq 0\}$

$L = \{\lambda : a \leq \lambda \leq b\}$

In each of the above cases, L is closed, and the theorem applies.

In the above theorem, we required that the vector $\mathbf{d}$ be nonzero. Example 8.3.2 below presents a case in which $\mathbf{M}$ is not closed if $\mathbf{d} = \mathbf{0}$. In most cases, the direction vector $\mathbf{d} \neq \mathbf{0}$ over points outside the solution set Ω. Thus, $\mathbf{M}$ is closed at these points, and Theorem 7.2.3 can be applied to prove convergence.

8.3.2 Example

Consider the following problem:

Minimize $(x-2)^4$

Here $f(x) = (x-2)^4$. Now consider the sequence $(x_k, d_k) = (1/k, 1/k)$. Clearly x_k converges to $x = 0$, and d_k converges to $d = 0$. Consider the line search map $\mathbf{M}$ defined in Theorem 8.3.1, where $L = \{\lambda : \lambda \geq 0\}$. The point y_k is obtained by solving the problem to minimize $f(x_k + \lambda d_k)$ subject to $\lambda \geq 0$. The reader can verify that $y_k = 2$. Note, however, that $\mathbf{M}(0, 0) = \{0\}$, so that $y \notin \mathbf{M}(x, d)$. This shows that $\mathbf{M}$ is not closed.

8.4 Multidimensional Search Without Using Derivatives

In this section we consider the problem of minimizing a function f of several variables without using derivatives. The methods described in this section proceed in the following manner. Given a vector $\mathbf{x}$, a suitable direction $\mathbf{d}$ is first determined, and then f is minimized from $\mathbf{x}$ in the direction $\mathbf{d}$ by one of the techniques discussed earlier in this chapter.

Throughout the book we are required to solve a line search problem of the form minimize $f(\mathbf{x} + \lambda \mathbf{d})$ subject to $\lambda \in L$, where L is typically of the form $L = E_1$, $L = \{\lambda : \lambda \geq 0\}$, or $L = \{\lambda : a \leq \lambda \leq b\}$. In the statements of the algorithms, for the purpose of simplicity, we have assumed that a minimum point $\bar{\lambda}$ exists. However, this may not be the case. Here the optimal objective value of the line search problem may be unbounded, or else the optimal objective value may be finite but not achieved at any particular λ. In the first case, the original problem is unbounded and we may stop. In the latter case, λ could be chosen as $\bar{\lambda}$ such that $f(\mathbf{x} + \bar{\lambda}\mathbf{d})$ is sufficiently close to the $\inf\{f(\mathbf{x} + \lambda \mathbf{d}) : \lambda \in L\}$.

The Cyclic Coordinate Method

This method uses the coordinate axes as the search directions. More specifically, the method searches along the directions $\mathbf{d}_1, \ldots, \mathbf{d}_n$, where $\mathbf{d}_j$ is a vector of zeros except for a one at the jth position. Thus, along the search direction $\mathbf{d}_j$, the variable x_j is changed, while all other variables are kept fixed. The method is illustrated schematically in Figure 8.5 for the problem of Example 8.4.1.

Summary of the Cyclic Coordinate Method

We summarize below the cyclic coordinate method for minimizing a function of several variables without using any derivative information. As we show shortly, if the function is differentiable, then the method converges to a stationary point.

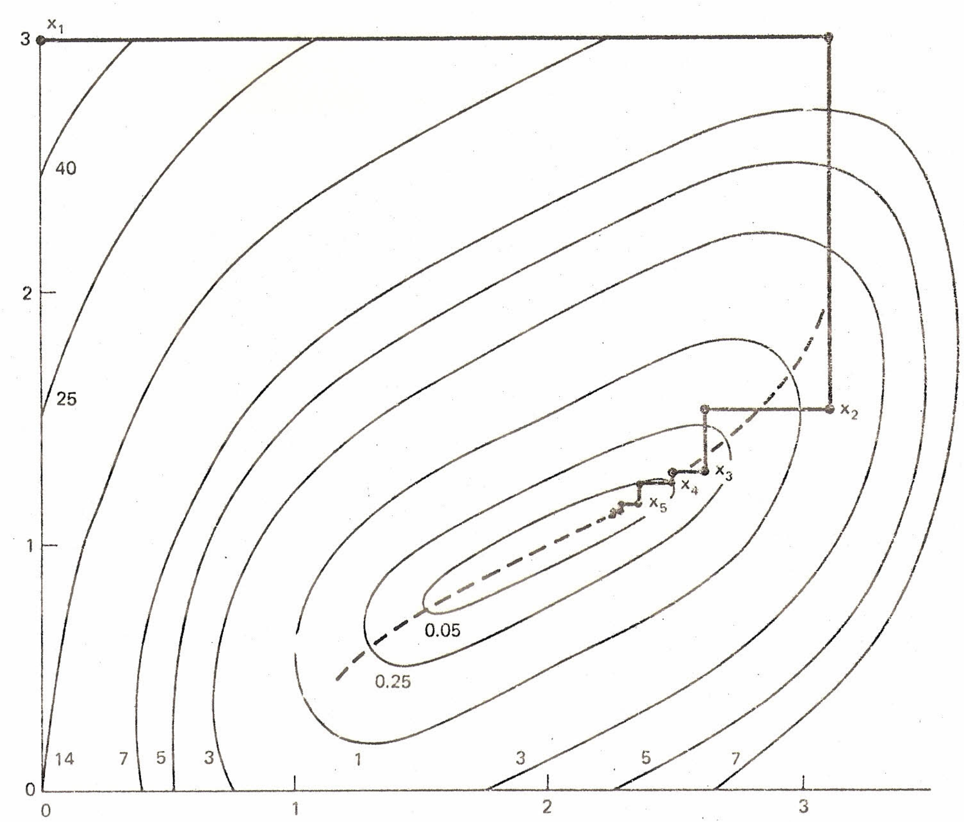


Figure 8.5 Illustration of the cyclic coordinate method.

As discussed in Section 7.2, several criteria could be used for terminating the algorithm. In the statement of the algorithm below, the termination criterion $\|\mathbf{x}_{k+1}-\mathbf{x}_k\|<\varepsilon$ is used. Obviously, any of the other criterion could be used to stop the procedure.

Initialization Step Choose a scalar $\varepsilon>0$ to be used for terminating the algorithm, and let $\mathbf{d}_1, \ldots, \mathbf{d}_n$ be the coordinate directions. Choose an initial point $\mathbf{x}_1$, let $\mathbf{y}_1=\mathbf{x}_1$, let $k=j=1$, and go to the main step.

Main Step

1. Let λ_j be an optimal solution to the problem to minimize $f(\mathbf{y}_j+\lambda \mathbf{d}_j)$ subject to $\lambda \in E_1$, and let $\mathbf{y}_{j+1}=\mathbf{y}_j+\lambda_j\mathbf{d}_j$. If $j<n$, replace j by $j+1$, and repeat step 1. Otherwise, if $j=n$, go to step 2.
2. Let $\mathbf{x}_{k+1}=\mathbf{y}_{n+1}$. If $\|\mathbf{x}_{k+1}-\mathbf{x}_k\|<\varepsilon$, then stop. Otherwise, let $\mathbf{y}_1=\mathbf{x}_{k+1}$, let $j=1$, replace k by $k+1$, and repeat step 1.

8.4.1 Example

Consider the following problem:

$$\text{Minimize} \quad (x_1-2)^4+(x_1-2x_2)^2$$

Note that the optimal solution to the above problem is $(2, 1)$ with objective value equal to zero. Table 8.6 gives a summary of computations for the cyclic coordinate method starting from the initial point $(0, 3)$. Note that at each iteration, the vectors $\mathbf{y}_2$ and $\mathbf{y}_3$ are obtained by performing a line search in the directions $(1, 0)$ and $(0, 1)$, respectively. Also note that significant progress is made during the first few iterations, whereas much slower progress is made during later iterations. After seven iterations, the point $(2.22, 1.11)$, whose objective value is 0.0023, is reached.

In Figure 8.5, the contours of the objective function are given, and the points generated above by the cyclic coordinate method are shown. Note that at later iterations, slow progress is made because of the short orthogonal movements along the valley indicated by the dotted lines.

Convergence of the Cyclic Coordinate Method

Convergence of the cyclic coordinate method to a stationary point follows immediately from Theorem 7.3.5 under the following assumptions.

1. The minimum of f along any line in E_n is unique.
2. The sequence of points generated by the algorithm is contained in a compact subset of E_n.

TABLE 8.6 **Summary of Computations for the Cyclic Coordinate Method**

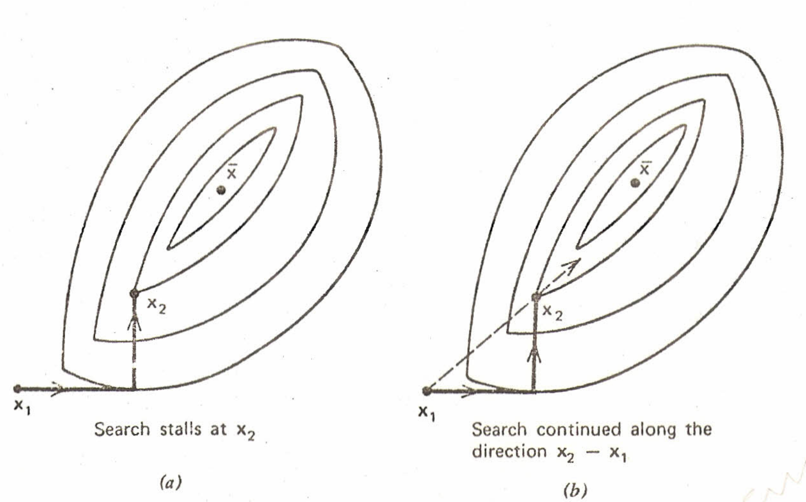
Iteration k	$\mathbf{x}_k$ / $f(\mathbf{x}_k)$	j	$\mathbf{d}_j$	$\mathbf{y}_j$	λ_j	$\mathbf{y}_{j+1}$
1	(0.00, 3.00)	1	(1.0, 0.0)	(0.00, 3.00)	3.13	(3.13, 3.00)
	52.00	2	(0.0, 1.0)	(3.13, 3.00)	−1.44	(3.13, 1.56)
2	(3.13, 1.56)	1	(1.0, 0.0)	(3.13, 1.56)	−0.50	(2.63, 1.56)
	1.63	2	(0.0, 1.0)	(2.63, 1.56)	−0.25	(2.63, 1.31)
3	(2.63, 1.31)	1	(1.0, 0.0)	(2.63, 1.31)	−0.19	(2.44, 1.31)
	0.16	2	(0.0, 1.0)	(2.44, 1.31)	−0.09	(2.44, 1.22)
4	(2.44, 1.22)	1	(1.0, 0.0)	(2.44, 1.22)	−0.09	(2.35, 1.22)
	0.04	2	(0.0, 1.0)	(2.35, 1.22)	−0.05	(2.35, 1.17)
5	(2.35, 1.17)	1	(1.0, 0.0)	(2.35, 1.17)	−0.06	(2.29, 1.17)
	0.015	2	(0.0, 1.0)	(2.29, 1.17)	−0.03	(2.29, 1.14)
6	(2.29, 1.14)	1	(1.0, 0.0)	(2.29, 1.14)	−0.04	(2.25, 1.14)
	0.007	2	(0.0, 1.0)	(2.25, 1.14)	−0.02	(2.25, 1.12)
7	(2.25, 1.12)	1	(1.0, 0.0)	(2.25, 1.12)	−0.03	(2.22, 1.12)
	0.004	2	(0.0, 1.0)	(2.22, 1.12)	−0.01	(2.22, 1.11)

Note that the search directions used at each iteration are the coordinate vectors, so that the matrix of search directions $\mathbf{D}=\mathbf{I}$. Obviously, assumption (1) of Theorem 7.3.5 holds true.

As an alternative approach, Theorem 7.2.3 could have been used to prove convergence after showing that the overall algorithmic map is closed at each $\mathbf{x}$ satisfying $\nabla f(\mathbf{x}) \neq \mathbf{0}$. In this case, the descent function α is taken as f itself, and the solution set $\Omega=\{\mathbf{x}: \nabla f(\mathbf{x})=\mathbf{0}\}$.

Acceleration Step

We learned from the foregoing analysis that the cyclic coordinate method, when applied to a differentiable function, will converge to a point with zero gradient. In the absence of differentiability, however, the method can stall at a nonoptimal point. As shown in Figure 8.6*a*, searching along any of the coordinate axes at the point $\mathbf{x}_2$ leads to no improvement of the objective



vencida

Figure 8.6 Illustration of the effect of a ridge.

function and results in premature termination. The reason for this premature termination is the presence of a valley caused by nondifferentiability of f. As illustrated in Figure 8.6b, this difficulty could be overcome by searching along the direction $\mathbf{x}_2 - \mathbf{x}_1$. mínimo

The search along a direction $\mathbf{x}_{k+1} - \mathbf{x}_k$ is frequently used in applying the cyclic coordinate method, even in the case where f is differentiable. The usual rule of thumb is to apply it at every pth iteration. This modification to the cyclic coordinate method frequently accelerates convergence, particularly when the sequence of points generated zigzag along a valley. Such a step is usually referred to as an *acceleration step.*

The Method of Hooke and Jeeves

The method of Hooke and Jeeves performs two types of search, exploratory search and pattern search. The first two iterations of the procedure are illustrated in Figure 8.7. Given $\mathbf{x}_1$, exploratory search along the coordinate directions produces the point $\mathbf{x}_2$. Now a pattern search along a direction $\mathbf{x}_2 - \mathbf{x}_1$ leads to the point $\mathbf{y}$. Another exploratory search starting from $\mathbf{y}$ gives the point $\mathbf{x}_3$. The next pattern search is along the direction $\mathbf{x}_3 - \mathbf{x}_2$, yielding $\mathbf{y}'$. The process is then repeated.

Summary of the Method of Hooke and Jeeves Using Line Search

As orginally proposed by Hooke and Jeeves, the method does not perform line search but rather takes discrete steps along the search directions, as will be discussed later. Here we present a continuous version of the method using line search along the coordinate directions $\mathbf{d}_1, \ldots, \mathbf{d}_n$ and the pattern direction.

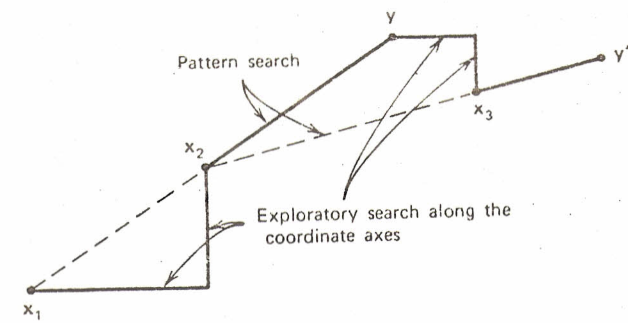


Figure 8.7 Illustration of the method of Hooke and Jeeves.

Initialization Step Choose a scalar $\varepsilon > 0$ to be used in terminating the algorithm. Choose a starting point $\mathbf{x}_1$, let $\mathbf{y}_1 = \mathbf{x}_1$, let $k = j = 1$, and go to the main step.

Main Step

1. Let λ_j be an optimal solution to the problem to minimize $f(\mathbf{y}_j + \lambda \mathbf{d}_j)$ subject to $\lambda \in E_1$, and let $\mathbf{y}_{j+1} = \mathbf{y}_j + \lambda_j \mathbf{d}_j$. If $j < n$, replace j by $j+1$, and repeat step 1. Otherwise, if $j = n$, let $\mathbf{x}_{k+1} = \mathbf{y}_{n+1}$. If $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$, stop; otherwise, go to step 2.
2. Let $\mathbf{d} = \mathbf{x}_{k+1} - \mathbf{x}_k$, and let $\hat{\lambda}$ be an optimal solution to the problem to minimize $f(\mathbf{x}_{k+1} + \lambda \mathbf{d})$ subject to $\lambda \in E_1$. Let $\mathbf{y}_1 = \mathbf{x}_{k+1} + \hat{\lambda}\mathbf{d}$, let $j = 1$, replace k by $k+1$, and repeat step 1.

8.4.2 Example

Consider the following problem:

$$\text{Minimize} \quad (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

Note that the optimal solution is (2.00, 1.00) with objective value equal to zero. Table 8.7 summarizes the computations for the method of Hooke and Jeeves, starting from the initial point (0.00, 3.00). At each iteration, an exploratory search along the coordinate directions gives the points $\mathbf{y}_2$ and $\mathbf{y}_3$, and a pattern search along the direction $\mathbf{d} = \mathbf{x}_{k+1} - \mathbf{x}_k$ gives the point $\mathbf{y}_1$, except at iteration $k = 1$, where $\mathbf{y}_1 = \mathbf{x}_1$. Note that four iterations were required to move from the initial point to the optimal point (2.00, 1.00) whose objective value is zero. At this point $\|\mathbf{x}_5 - \mathbf{x}_4\| = 0.045$, and the procedure is terminated.

Figure 8.8 illustrates the points generated by the method of Hooke and Jeeves using line search. Note that the pattern search has substantially improved convergence by moving along a direction that is almost parallel to the valley shown in dotted lines.