Studies in Optimization

D. M. BURLEY

Intertext Books

Chapter II

Hill Climbing without Constraints

2.1 Introduction

The present chapter is concerned with problems where the function to be optimized is a complicated one; examples 2, 3, 16 of chapter 1 illustrate this. A further example shows a problem where the function evaluation is a formidable task and the derivative calculation almost impossible.

Example 17 A pendulum consists of a rod of length $2a$ and mass M and it can swing about one end. A regulator consists of a mass m attached at a distance x from the pivot. Find the period of (large) oscillations of the pendulum if it is given an angular velocity ω while hanging in equilibrium. Has this period a local maximum or minimum as x varies?

If the rod makes an angle θ with the vertical then the equation of motion is

$$I\ddot{\theta} + g(Ma + mx)\sin\theta = 0$$

where $I = \frac{4}{3}Ma^2 + mx^2$. After two integrations the period $t(x)$ is given by

$$t(x) = 4\int_0^{\alpha} \frac{d\theta}{(\omega^2 - 4k^2\sin^2\frac{1}{2}\theta)^{\frac{1}{2}}},$$

where $k^2 = g(Ma + mx)/(\frac{4}{3}Ma^2 + mx^2)$ and $\alpha = 2\sin^{-1}(\omega/2k)$.

While it is theoretically possible to evaluate dt/dx the effort involved is large even if full use of elliptic integrals is made. Some indication of the type of behaviour expected can be obtained by looking at the small oscillation limit when $t(x) = 2\pi/k = A[(\frac{4}{3}C + y^2)/(C + y)]^{\frac{1}{2}}$, where A is a constant, $C = M/m$ and $y = x/a$. This function has a negative gradient at $y = 0$ and tends to $+\infty$ as $y \to +\infty$ and hence has a local minimum somewhere

between. It may be noted that such a minimum may be out of the range of interest since the regulator is fixed to the rod so that $0 \leq x \leq 2a$.

In other examples, example 2 for instance, the derivative is easy to obtain but both function and derivative evaluations are major tasks. Similarly in more than one dimension, example 16 for instance, the same sort of problems remain. The object, therefore, is to develop a series of numerical algorithms which give a reasonable assurance that the extrema can be obtained. It is clear that if an analytic solution can be obtained then this is preferable since it gives a much clearer mathematical or physical picture of the problem under study. If, however, the problem has to be studied numerically then it is essential that the work be done efficiently.

The present chapter is not exhaustive by any means (the important method of conjugate gradients is omitted completely) but is meant to give an indication of the main methods used. Indeed there are almost as many methods as there are workers in the field. Comprehensive reviews and bibliographies can be found in Beveridge and Schechter (1970), Murray (1972). One method, for instance, which will not be looked at in detail, involves a random search. Suppose it is required to maximize a given function. A point is chosen at random, the function evaluated there and this process is repeated as many times as possible. If the evaluation at the current point gives the largest function value yet calculated then it is retained as the best point, otherwise the next random point is chosen. The method has the considerable advantages of simplicity and of requiring little storage; it is, however, comparatively inefficient since it discards all the previously obtained information. Even though the amount of work increases exponentially with the number of independent variables, when this number is very large it is often the only one possible.

Değişken

2.2 Single Variable Problems

2.2.1 Bracketing

Aralığı Almak

It is most important as a first step in a calculation to get a rough idea of where to look for an extremum; a useful idea is to find two values that bracket the extremum. Suppose the maximum of a function $f(x)$ is sought and it is known that the maximum is in the region $x \geq a$. Choose an increment h and evaluate f at points $x_1 = a$, $x_2 = x_1 + h$, $x_3 = x_2 + 2h$, $x_4 = x_3 + 4h$, $x_5 = x_4 + 8h, \ldots,$ that is doubling the increment at each stage. The evaluation is stopped if either the maximum is bracketed or if $x_i > X$, where X is a suitable large constant chosen at the start of the calculation. A maximum is bracketed if at some stage $f(x_i) > f(x_{i-1})$ *and* $f(x_i) > f(x_{i+1})$; the bracket is (x_{i-1}, x_{i+1}). If the *gradient* is known also then if f' at x_j and f' at x_{j+1} are positive and negative respectively then the bracket for the maximum is (x_j, x_{j+1}). If the value of f increases until X is reached it is usually assumed that the function increases indefinitely; if on the other hand the value of X is reached and the function is still decreasing then $f(a)$ is usually taken as the maximum value. Figure 2.1 illustrates this procedure.

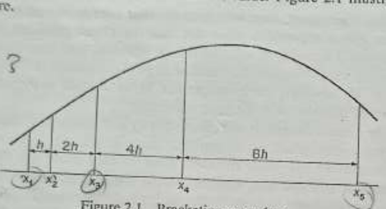

Figure 2.1 Bracketing procedure.

Key decisions must be made at the start of the calculation, the values of a, h, X and the direction of search. Often the search direction cannot be decided on other evidence and both directions must be tried. It is usually better to err on the small side for h and the large side for X, the doubling procedure soon lengthens the interval and soon reaches X. This method, of course, is not foolproof but is found to work well (see problems 1 and 2).

Example 18 Bracket the maximum of $f = \tanh x/(1 + x^2)$.

Now

$$f' = \frac{\operatorname{sech}^2 x}{1 + x^2} - \frac{2x \tanh x}{(1 + x^2)^2}.$$

Since $f' > 0$ when $x = 0$ and since $f \to 0$ as $x \to \infty$ it is safe to assume that the maximum occurs for $x > 0$. Choose $h = 0.1$, $a = 0.1$, $X = 100$.

Table 2.1 Bracketing procedure.

x	0.1	0.2	0.4	0.8	1.6
f	0.098 68	0.189 78	0.327 54	0.404 90	0.258 90
f'	0.960 72	0.851 09	0.511 73	−0.054 14	−0.190 44

Using function values only the bracket for the maximum is (0.4, 1.6), while if the derivative is also used the bracket is (0.4, 0.8).

2.2.2 POLYNOMIAL APPROXIMATION

Once a bracket has been obtained for the extremum it is then required to obtain the extremum to any required accuracy. One simple way of doing this is to use the information obtained by the bracketing procedure directly and approximate this information by a polynomial (see problem 3). Suppose $f_1 = f(z_1)$, $f_2 = f(z_2)$, $f_3 = f(z_3)$ are known and $f_2 > f_1, f_3$ so that (z_1, z_3) brackets the maximum of the function f. A quadratic approximation for f can be written

$$f = (z - z_2)(z - z_3)F_1 + (z - z_3)(z - z_1)F_2 + (z - z_1)(z - z_2)F_3$$

where $F_1 = f_1/(z_1 - z_2)(z_1 - z_3)$, etc. This quadratic has a maximum at $f'(z) = 0$ or at

$$z^* = \tfrac{1}{2}(z_1 + z_2 + z_3) - \frac{\tfrac{1}{2}(z_1F_1 + z_2F_2 + z_3F_3)}{F_1 + F_2 + F_3}. \qquad (2.1)$$

Thus an estimate, $f^* = f(z^*)$, for the extremum is now available. A new calculation can be commenced for the maximum by identifying the 'best' bracket from the values z_1, z_2, z^*, z_3, relabelling appropriately and using (2.1) again.

There are many variants of this algorithm, these depend mainly on the expense of evaluating the function. If the function values are easy to obtain one alternative is to choose the maximum from f_2 and f^*. If $f_2 \geq f^*$ keep z_2 as the best value but if $f_2 < f^*$ put $z_2 = z^*$. A new calculation is then performed by putting $h = \tfrac{1}{2}(z_3 - z_1)$ and then using new values $z_1 = z_2 - h$, $z_3 = z_2 + h$. The formula (2.1) takes the particularly simple form

$$z^* = z_2 + \frac{\tfrac{1}{2}h(f_1 - f_3)}{f_1 - 2f_2 + f_3} \qquad (2.2)$$

and provides a very suitable hand computation method.

Example 18(*a*) Use the quadratic approximation algorithm on the function

$$f = \frac{\tanh z}{1 + z^2}.$$

From the previous bracketing procedure in section 2.2.1 (0.4, 1.6) brackets the maximum.

z	0.4	1.0	1.6
f	0.327 54	0.387 54	0.258 90

$$z(\text{new}) = 1 + \tfrac{1}{2} \cdot 0.6 \cdot \frac{0.068\,64}{-0.175\,16} = 0.88$$

$$f(0.88) = 0.398\,12.$$

$f_x = 7$

Now choose $z_2 = 0.88$, $h = 0.3$.

z	0.58	0.88	1.18	$z(\text{new}) = 0.77$
f	0.391 10	0.398 12	0.345 87	$f(0.77) = 0.406\,15$.

The best value to date is $z = 0.77$, $f = 0.406\,15$. Continuing the calculation on the computer gives $z = 0.741\,60$, $f = 0.406\,53$ to five figure accuracy with 34 function evaluations.

Working directly with the values $z = 0.4, 0.8, 1.6$, obtained from the bracketing in section 2.2.1, an implementation of (2.1) on the computer gives the five figure accuracy with 17 function evaluations.

If the *derivative* is also available, immediate advantage is obtained as can be observed from example 18 in section 2.2.1 where a much more accurate bracket was obtained. Suppose a maximum of f is bracketed by (x_1, x_2) and hence f_1, f'_1, f_2, f'_2 are known. Since there are now four pieces of information known it is possible to approximate the function in the interval by a cubic.

$$f(x) = ax^3 + bx^2 + cx + d \tag{2.3}$$

$$f'(x) = 3ax^2 + 2bx + c. \tag{2.4}$$

Putting the known information into (2.3) and (2.4) the result can be summarized in the matrix equation (2.5)

$$\begin{bmatrix} f_1 \\ f_2 \\ f'_1 \\ f'_2 \end{bmatrix} = \begin{bmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ 3x_1^2 & 2x_1 & 1 & 0 \\ 3x_2^2 & 2x_2 & 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \tag{2.5}$$

which can be solved for a, b, c, d. The maximum of the approximating cubic is given by $f'(x) = 0$ from equation (2.4) or

$$x^* = \frac{-b \pm (b^2 - 3ac)^{\frac{1}{2}}}{3a} \tag{2.6}$$

and the sign is chosen to make $x_1 < x^* < x_2$. The gradients f'_1, f'_*, f'_2 are compared and the two new points are chosen to give gradients of opposite sign. The same procedure can then be repeated until sufficient accuracy is obtained.

This method works extremely well for most functions and is very well used. It is, however, not particularly suitable for a hand computation since a matrix equation (2.5) must be solved and a square root taken in (2.6) (see problem 5). On a computer these problems are straightforward and such an algorithm is usually available in most programme libraries.

Example 18(*b*) Use the cubic approximation algorithm to find the maximum of $f = \tanh x/(1 + x^2)$.

From section 2.2.1, example 18, the maximum is bracketed by (0.4, 0.8).

x	f	f'
0.4	0.327 54	0.511 73
0.8	0.404 90	−0.054 14

Solve (2.5); $a = 0.442\,4$, $b = -1.504$, $c = 1.502$
From (2.6); $x^* = 0.744$.

This new value gives $f_* = 0.406\,53$, $f'_* = -0.002\,44$, which is already better than the quadratic approximation, example 18(a). Since $f'(0.4)$ and $f'(0.744)$ have opposite signs these are chosen for the next approximation.

x	f	f'
0.4	0.327 54	0.511 73
0.744	0.406 53	−0.002 44

If the solution of (2.5) and (2.6) is repeated the new value gives $x^* = 0.741\,7$,
$f_* = 0.406\,53$, $f'_* = -0.000\,11$.

The calculation proceeds on the computer to give $f' = 0$ to 5 decimal places at the next iteration with $x = 0.741\,60$, $f = 0.406\,53$.

2.2.3 FIBONACCI TYPE SEARCH

This class of methods is often called the class of direct search methods and is concerned with optimizing when the derivative is *not* known. Once a bracket has been obtained the aim is to progressively reduce the length of the bracket until it is less than a prescribed limit.

Suppose (a_1, a_2) brackets a required *maximum* of the function $f(x)$. The points a_3, a_4 are symmetrically placed in this interval, so that

$$\left.\begin{aligned} a_3 &= (1-\alpha)a_1 + \alpha a_2, \\ a_4 &= \alpha a_1 + (1-\alpha)a_2, \end{aligned} \quad 0 < \alpha < \tfrac{1}{2}\right\} \tag{2.7}$$

and this division is illustrated in figure 2.2. Calculate $f_i = f(a_i)$, $i = 1, 2, 3, 4$; then *either* $f(a_4) > f(a_3)$ and it is now assumed that (a_3, a_2) brackets the

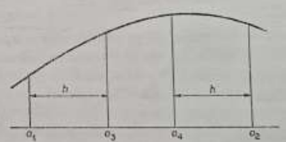

Figure 2.2 Interval division for Fibonacci search.