Explicitly,

$$\tau_1 = \frac{1+\sqrt{5}}{2}, \qquad \tau_2 = \frac{1-\sqrt{5}}{2}.$$

(The number $\tau_1 = 1.618$ is known as the *golden section* ratio and was considered by early Greeks to be the most aesthetic value for the ratio of two adjacent sides of a rectangle.)

For large N the first term on the right side of (4) dominates the second, and hence

$$\lim_{N\to\infty} \frac{F_{N-1}}{F_N} = \frac{1}{\tau_1} \simeq 0.618.$$

It follows from (1) that the interval of uncertainty at any point in the process has width

$$d_k = \left(\frac{1}{\tau_1}\right)^{k-1} d_1, \tag{5}$$

and from this it follows that

$$\frac{d_{k+1}}{d_k} = \frac{1}{\tau_1} = 0.618. \tag{6}$$

Therefore, we conclude that, with respect to the width of the uncertainty interval, the search by golden section converges linearly (see Section 6.6) to the overall minimum of the function f with convergence ratio $1/\tau_1 = 0.618$.

7.2 LINE SEARCH BY CURVE FITTING

The Fibonacci search method has a certain amount of theoretical appeal, since it assumes only that the function being searched is unimodal and with respect to this broad class of functions the method is, in some sense, optimal. In most problems, however, it can be safely assumed that the function being searched, as well as being unimodal, possesses a certain degree of smoothness, and one might, therefore, expect that more efficient search techniques exploiting this smoothness can be devised; and indeed they can. Techniques of this nature are usually based on curve fitting procedures where a smooth curve is passed through the previously measured points in order to determine an estimate of the minimum point. A variety of such techniques can be devised depending on whether or not derivatives of the function as well as the values can be measured, how many previous points are used to determine the fit, and the criterion used to determine the fit. In this section a number of possibilities are outlined and analyzed. All of them have orders of convergence greater than unity.

Newton's Method

Suppose that the function f of a single variable x is to be minimized, and suppose that at a point x_k where a measurement is made it is possible to evaluate the three numbers $f(x_k)$, $f'(x_k)$, $f''(x_k)$. It is then possible to construct a quadratic function q which at x_k agrees with f up to second derivatives, that is

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \tfrac{1}{2}f''(x_k)(x - x_k)^2. \tag{7}$$

We may then calculate an estimate x_{k+1} of the minimum point of f by finding the point where the derivative of q vanishes. Thus setting

$$0 = q'(x_{k+1}) = f'(x_k) + f''(x_k)(x_{k+1} - x_k),$$

we find

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}. \tag{8}$$

This process, which is illustrated in Fig. 7.3, can then be repeated at x_{k+1}.

We note immediately that the new point x_{k+1} resulting from Newton's method does not depend on the value $f(x_k)$. The method can more simply be viewed as a technique for iteratively solving equations of the form

$$g(x) = 0,$$

where, when applied to minimization, we put $g(x) \equiv f'(x)$. In this notation Newton's method takes the form

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}. \tag{9}$$

This form is illustrated in Fig. 7.4.

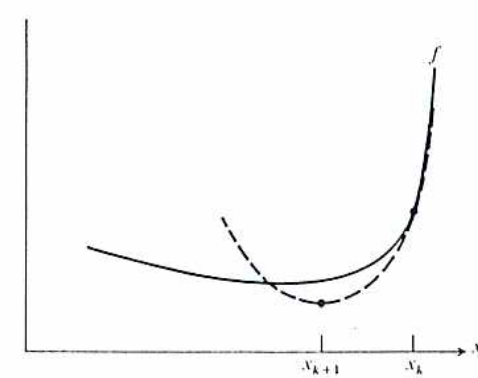

Fig. 7.3 Newton's method for minimization

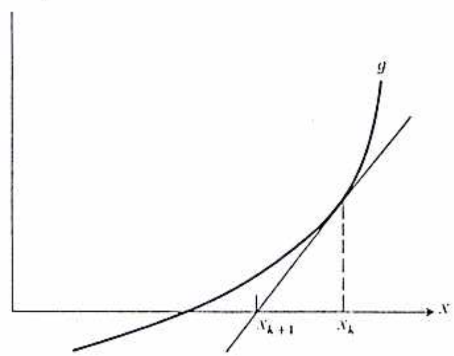

Fig. 7.4 Newton's method for solving equations

We now show that Newton's method has order two convergence:

Proposition. *Let the function g have a continuous second derivative, and let x^* satisfy $g(x^*) = 0$, $g'(x^*) \neq 0$. Then, provided x_0 is sufficiently close to x^*, the sequence $\{x_k\}_{k=0}^{\infty}$ generated by Newton's method (9) converges to x^* with an order of convergence at least two.*

Proof. For points ξ in a region near x^* there is a k_1 such that $|g''(\xi)| < k_1$ and a k_2 such that $|g'(\xi)| > k_2$. Then since $g(x^*) = 0$ we can write

$$x_{k+1} - x^* = x_k - x^* - \frac{g(x_k) - g(x^*)}{g'(x_k)}$$
$$= -[g(x_k) - g(x^*) + g'(x_k)(x^* - x_k)]/g'(x_k).$$

The term in brackets is, by Taylor's theorem, zero to first-order. In fact, using the remainder term in a Taylor series expansion about x_k, we obtain

$$x_{k+1} - x^* = -\frac{1}{2}\frac{g''(\xi)}{g'(x_k)}(x_k - x^*)^2$$

for some ξ between x^* and x_k. Thus in the region near x^*,

$$|x_{k+1} - x^*| \leqslant \frac{k_1}{2k_2}|x_k - x^*|^2.$$

We see that if $|x_k - x^*|k_1/2k_2 < 1$, then $|x_{k+1} - x^*| < |x_k - x^*|$ and thus we conclude that if started close enough to the solution, the method will converge to x^* with an order of convergence at least two. ∎

Method of False Position

Newton's method for minimization is based on fitting a quadratic on the basis of information at a single point; by using more points, less information is required at each of them. Thus, using $f(x_k)$, $f'(x_k)$, $f'(x_{k-1})$ it is possible to fit the quadratic

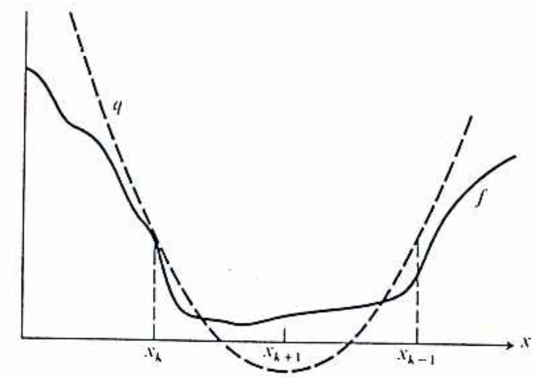

Fig. 7.5 False position for minimization

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{f'(x_{k-1}) - f'(x_k)}{x_{k-1} - x_k} \cdot \frac{(x - x_k)^2}{2},$$

which has the same corresponding values. An estimate x_{k+1} can then be determined by finding the point where the derivative of q vanishes; thus

$$x_{k+1} = x_k - f'(x_k)\left[\frac{x_{k-1} - x_k}{f'(x_{k-1}) - f'(x_k)}\right]. \quad (10)$$

(See Fig. 7.5.) Comparing this formula with Newton's method, we see again that the value $f(x_k)$ does not enter; hence, our fit could have been passed through either $f(x_k)$ or $f(x_{k-1})$. Also the formula can be regarded as an approximation to Newton's method where the second derivative is replaced by the difference of two first derivatives.

Again, since this method does not depend on values of f directly, it can be regarded as a method for solving $f'(x) \equiv g(x) = 0$. Viewed in this way the method, which is illustrated in Fig. 7.6, takes the form

$$x_{k+1} = x_k - g(x_k)\left[\frac{x_k - x_{k-1}}{g(x_k) - g(x_{k-1})}\right]. \quad (11)$$

We next investigate the order of convergence of the method of false position and discover that it is order $\tau_1 \simeq 1.618$, the golden mean.

Proposition. *Let g have a continuous second derivative and suppose x^* is such that $g(x^*) = 0$, $g'(x^*) \neq 0$. Then for x_0 sufficiently close to x^*, the sequence $\{x_k\}_{k=0}^{\infty}$ generated by the method of false position (11) converges to x^* with order $\tau_1 \simeq 1.618$.*

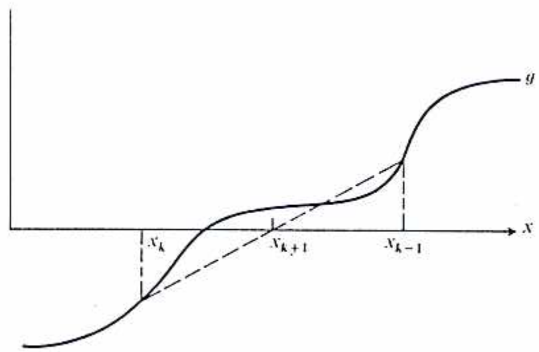

Fig. 7.6 False position for solving equations

Proof. Introducing the notation

$$g[a, b] = \frac{g(b) - g(a)}{b - a}, \tag{12}$$

we have

$$\begin{aligned} x_{k+1} - x^* &= x_k - x^* - g(x_k)\left[\frac{x_k - x_{k-1}}{g(x_k) - g(x_{k-1})}\right] \\ &= (x_k - x^*)\left\{\frac{g[x_{k-1}, x_k] - g[x_k, x^*]}{g[x_{k-1}, x_k]}\right\}. \end{aligned} \tag{13}$$

Further, upon the introduction of the notation

$$g[a, b, c] = \frac{g[a, b] - g[b, c]}{a - c},$$

we may write (13) as

$$x_{k+1} - x^* = (x_k - x^*)(x_{k-1} - x^*)\left\{\frac{g[x_{k-1}, x_k, x^*]}{g[x_{k-1}, x_k]}\right\}.$$

Now, by the mean value theorem with remainder, we have (see Exercise 2)

$$g[x_{k-1}, x_k] = g'(\xi_k) \tag{14}$$

and

$$g[x_{k-1}, x_k, x^*] = \tfrac{1}{2}g''(\eta_k), \tag{15}$$

where ξ_k and η_k are convex combinations of x_k, x_{k-1} and x_k, x_{k-1}, x^*, respectively. Thus

$$x_{k+1} - x^* = \frac{g''(\eta_k)}{2g'(\xi_k)}(x_k - x^*)(x_{k-1} - x^*). \tag{16}$$

It follows immediately that the process converges if it is started sufficien close to x^*.

To determine the order of convergence, we note that for large k E (16) becomes approximately

$$x_{k+1} - x^* = M(x_k - x^*)(x_{k-1} - x^*),$$

where

$$M = \frac{g''(x^*)}{2g'(x^*)}.$$

Thus defining $\varepsilon_k = (x_k - x^*)$ we have, in the limit,

$$\varepsilon_{k+1} = M\varepsilon_k\varepsilon_{k-1}. \quad ($$

Taking the logarithm of this equation we have, with $y_k = \log M\varepsilon_k$,

$$y_{k+1} = y_k + y_{k-1}, \quad ($$

which is the Fibonacci difference equation discussed in Section 7.1. solution to this equation will satisfy

$$y_{k+1} - \tau_1 y_k \to 0.$$

Thus

$$\log M\varepsilon_{k+1} - \tau_1 \log M\varepsilon_k \to 0 \quad \text{or} \quad \log\frac{M\varepsilon_{k+1}}{(M\varepsilon_k)^{\tau_1}} \to 0,$$

and hence

$$\frac{\varepsilon_{k+1}}{\varepsilon_k^{\tau_1}} \to M^{(\tau_1 - 1)}. \blacksquare$$

Having derived the error formula (17) by direct analysis, it is now a propriate to point out a short-cut technique, based on symmetry and oth considerations, that can sometimes be used in even more complicated s uations. The right side of error formula (17) must be a polynomial in ε_k ar ε_{k-1}, since it is derived from approximations based on Taylor's theorer Furthermore, it must be second order, since the method reduces to Newton method when $x_k = x_{k-1}$. Also, it must go to zero if either ε_k or ε_{k-1} go zero, since the method clearly yields $\varepsilon_{k+1} = 0$ in that case. Finally, it mu be symmetric in ε_k and ε_{k-1}, since the order of points is irrelevant. The on formula satisfying these requirements is $\varepsilon_{k+1} = M\varepsilon_k\varepsilon_{k-1}$.

Cubic Fit

Given the points x_{k-1} and x_k together with the values $f(x_{k-1}$ $f'(x_{k-1})$, $f(x_k)$, $f'(x_k)$, it is possible to fit a cubic equation to the poin having corresponding values. The next point x_{k+1} can then be determine as the relative minimum point of this cubic. This leads to

$$x_{k+1} = x_k - (x_k - x_{k-1})\left[\frac{f'(x_k) + u_2 - u_1}{f'(x_k) - f'(x_{k-1}) + 2u_2}\right], \tag{19}$$

where

$$u_1 = f'(x_{k-1}) + f'(x_k) - 3\frac{f(x_{k-1}) - f(x_k)}{x_{k-1} - x_k}$$
$$u_2 = [u_1^2 - f'(x_{k-1})f'(x_k)]^{1/2},$$

which is easily implementable for computations.

It can be shown (see Exercise 3) that the order of convergence of the cubic fit method is 2.0. Thus, although the method is exact for cubic functions indicating that its order might be three, its order is actually only two.

Quadratic Fit

The scheme that is often most useful in line searching is that of fitting a quadratic through three given points. This has the advantage of not requiring any derivative information. Given x_1, x_2, x_3 and corresponding values $f(x_1) = f_1, f(x_2) = f_2, f(x_3) = f_3$ we construct the quadratic passing through these points

$$q(x) = \sum_{i=1}^{3} f_i \frac{\prod_{j\neq i}(x - x_j)}{\prod_{j\neq i}(x_i - x_j)}, \tag{20}$$

and determine a new point x_4 as the point where the derivative of q vanishes. Thus

$$x_4 = \frac{1}{2}\frac{b_{23}f_1 + b_{31}f_2 + b_{12}f_3}{a_{23}f_1 + a_{31}f_2 + a_{12}f_3}, \tag{21}$$

where $a_{ij} = x_i - x_j$, $b_{ij} = x_i^2 - x_j^2$.

Define the errors $\varepsilon_i = x^* - x_i$, $i = 1, 2, 3, 4$. The expression for ε_4 must be a polynomial in $\varepsilon_1, \varepsilon_2, \varepsilon_3$. It must be second order (since it is a quadratic fit). It must go to zero if any two of the errors $\varepsilon_1, \varepsilon_2, \varepsilon_3$ is zero. (The reader should check this.) Finally, it must be symmetric (since the order of points is relevant). It follows that near a minimum point x^* of f, the errors are related approximately by

$$\varepsilon_4 = M(\varepsilon_1\varepsilon_2 + \varepsilon_2\varepsilon_3 + \varepsilon_1\varepsilon_3), \tag{22}$$

where M depends on the values of the second and third derivatives of f at x^*.

If we assume that $\varepsilon_k \to 0$ with an order greater than unity, then for large k the error is governed approximately by

$$\varepsilon_{k+2} = M\varepsilon_k\varepsilon_{k-1}.$$

Letting $y_k = \log M\varepsilon_k$ this becomes

$$y_{k+2} = y_k + y_{k-1}$$

with characteristic equation

$$\lambda^3 - \lambda - 1 = 0.$$

The largest root of this equation is $\lambda \simeq 1.3$ which thus determines the rate of growth of y_k and is the order of convergence of the quadratic fit method.

7.3 GLOBAL CONVERGENCE OF CURVE FITTING

Above, we analyzed the convergence of various curve fitting procedures in the neighborhood of the solution point. If, however, any of these procedures were applied in pure form to search a line for a minimum, there is the danger—alas, the most likely possibility—that the process would diverge or wander about meaninglessly. In other words, the process may never get close enough to the solution for our detailed local convergence analysis to be applicable. It is therefore important to artfully combine our knowledge of the local behavior with conditions guaranteeing global convergence to yield a workable and effective procedure.

The key to guaranteeing global convergence is the Global Convergence Theorem of Chapter 6. Application of this theorem in turn hinges on the construction of a suitable descent function and minor modifications of a pure curve fitting algorithm. We offer below a particular blend of this kind of construction and analysis, taking as departure point the quadratic fit procedure discussed in Section 7.2 above.

Let us assume that the function f that we wish to minimize is strictly unimodal and has continuous second partial derivatives. We initiate our search procedure by searching along the line until we find three points x_1, x_2, x_3 with $x_1 < x_2 < x_3$ such that $f(x_1) \geqslant f(x_2) \leqslant f(x_3)$. In other words, the value at the middle of these three points is less than that at either end. Such a sequence of points can be determined in a number of ways—see Exercise 7.

The main reason for using points having this pattern is that a quadratic fit to these points will have a minimum (rather than a maximum) and the minimum point will lie in the interval $[x_1, x_3]$. See Fig. 7.7. We modify the pure quadratic fit algorithm so that it always works with points in this basic *three-point pattern*.

The point x_4 is calculated from the quadratic fit in the standard way and $f(x_4)$ is measured. Assuming (as in the figure) that $x_2 < x_4 < x_3$, and accounting for the unimodal nature of f, there are but two possibilities:

1. $f(x_4) \leqslant f(x_2)$
2. $f(x_2) < f(x_4) \leqslant f(x_3)$.

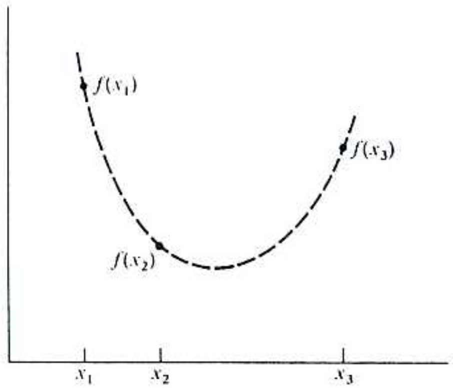

Fig. 7.7 Three-point pattern

In either case a new three-point pattern, $\bar{x}_1, \bar{x}_2, \bar{x}_3$, involving x_4 and two of the old points, can be determined: In case (1) it is

$$(\bar{x}_1, \bar{x}_2, \bar{x}_3) = (x_2, x_4, x_3),$$

while in case (2) it is

$$(\bar{x}_1, \bar{x}_2, \bar{x}_3) = (x_1, x_2, x_4).$$

We then use this three-point pattern to fit another quadratic and continue. The pure quadratic fit procedure determines the next point from the current point and the previous two points. In the modification above, the next point is determined from the current point and the two out of three last points that form a three-point pattern with it. This simple modification leads to global convergence.

To prove convergence, we note that each three-point pattern can be thought of as defining a vector $\mathbf{x}$ in E^3. Corresponding to an $\mathbf{x} = (x_1, x_2, x_3)$ such that (x_1, x_2, x_3) form a three-point pattern with respect to f, we define $\mathbf{A}(\mathbf{x}) = (\bar{x}_1, \bar{x}_2, \bar{x}_3)$ as discussed above. For completeness we must consider the case where two or more of the x_i, $i = 1, 2, 3$ are equal, since this may occur. The appropriate definitions are simply limiting cases of the earlier ones. For example, if $x_1 = x_2$, then (x_1, x_2, x_3) form a three-point pattern if $f(x_2) \leqslant f(x_3)$ and $f'(x_2) < 0$ (which is the limiting case of $f(x_2) < f(x_1)$). A quadratic is fit in this case by using the values at the two distinct points and the derivative at the duplicated point. In case $x_1 = x_2 = x_3$, (x_1, x_2, x_3) forms a three-point pattern if $f'(x_2) = 0$ and $f''(x_2) \geqslant 0$. With these definitions, the map $\mathbf{A}$ is well defined. It is also continuous, since curve fitting depends continuously on the data.

We next define the solution set $\Gamma \subset E^3$ as the points $\mathbf{x}^* = (x^*, x^*, x^*)$ where $f'(x^*) = 0$.

Finally, we let $Z(\mathbf{x}) = f(x_1) + f(x_2) + f(x_3)$. It is easy to see that Z is

a descent function for $\mathbf{A}$. After application of $\mathbf{A}$ one of the valu
$f(x_1), f(x_2), f(x_3)$ will be replaced by $f(x_4)$, and by construction, and t
assumption that f is unimodal, it will replace a strictly larger value.
course, at $\mathbf{x}^* = (x^*, x^*, x^*)$ we have $\mathbf{A}(\mathbf{x}^*) = \mathbf{x}^*$ and hence $Z(\mathbf{A}(\mathbf{x}^*)) = Z(\mathbf{x}$

Since all points are contained in the initial interval, we have all t
requirements for the Global Convergence Theorem. Thus the process co
verges to the solution. The order of convergence may not be destroyed
this modification, if near the solution the three-point pattern is always form
from the previous three points. In this case we would still have convergen
of order 1.3. This cannot be guaranteed, however.

It has often been implicitly suggested, and accepted, that when usi
the quadratic fit technique one should require

$$f(x_{k+1}) < f(x_k)$$

so as to guarantee convergence. If the inequality is not satisfied at sor
cycle, then a special local search is used to find a better x_{k+1} that do
satisfy it. This philosophy amounts to taking $Z(\mathbf{x}) = f(x_3)$ in our gene
framework and, unfortunately, this is not a descent function even for u
modal functions, and hence the special local search is likely to be necessa
several times. It is true, of course, that a similar special local search ma
occasionally, be required for the technique we suggest in regions of multip
minima, but it is never required in a unimodal region.

The above construction, based on the pure quadratic fit technique, c
be emulated to produce effective procedures based on other curve fitti
techniques. For application to smooth functions these techniques seem
be the best available in terms of flexibility to accommodate as much deri
ative information as is available, fast convergence, and a guarantee of glob
convergence.

7.4 CLOSEDNESS OF LINE SEARCH ALGORITHMS

Since searching along a line for a minimum point is a component part
most nonlinear programming algorithms, it is desirable to establish at on
that this procedure is closed; that is, that the end product of the iterati
procedures outlined above, when viewed as a single algorithmic step findi
a minimum along a line, define closed algorithms. That is the objective
this section.

To initiate a line search with respect to a function f, two vectors mu
be specified: the initial point $\mathbf{x}$ and the direction $\mathbf{d}$ in which the search is
be made. The result of the search is a new point. Thus we define the searc
algorithm $\mathbf{S}$ as a mapping from E^{2n} to E^n.

We assume that the search is to be made over the semi-infinite lir
emanating from $\mathbf{x}$ in the direction $\mathbf{d}$. We also assume, for simplicity, tha
the search is not made in vain; that is, we assume that there is a minimur