4

Numerical Evaluation of Derivatives and Integrals

• • •

The mathematics of the Greeks was insufficient to handle the concept of time. Perhaps the clearest demonstration of this is Zeno's Paradox regarding the flight of arrows. Zeno reasoned that since an arrow must cover half the distance between the bow and the target before traveling all the distance and half of that distance (i.e. a quarter of the whole) before that, etc., that the total number of steps the arrow must cover was infinite. Clearly the arrow could not accomplish that in a finite amount of time so that its flight to the target was impossible. This notion of a limiting process of an infinitesimal distance being crossed in an infinitesimal time producing a constant velocity seems obvious to us now, but it was a fundamental barrier to the development of Greek science. The calculus developed in the 17th century by Newton and Leibnitz has permitted, not only a proper handling of time and the limiting process, but the mathematical representation of the world of phenomena which science seeks to describe. While the analytic representation of the calculus is essential in this description, ultimately we must numerically evaluate the analytic expressions that we may develop in order to compare them with the real world.

Again we confront a series of subjects about which books have been written and entire courses of study developed. We cannot hope to provide an exhaustive survey of these areas of numerical analysis, but only develop the basis for the approach to each. The differential and integral operators reviewed in chapter 1 appear in nearly all aspects of the scientific literature. They represent mathematical processes or operations to be carried out on continuous functions and therefore can only be approximated by a series of discrete numerical operations. So, as with any numerical method, we must establish criteria for which the discrete operations will accurately represent the continuous operations of differentiation and integration. As in the case of interpolation, we shall find the criteria in the realm of polynomial approximation.

4.1 Numerical Differentiation

Compared with other subjects to be covered in the study of numerical methods, little is usually taught about numerical differentiation. Perhaps that is because the processes should be avoided whenever possible. The reason for this can be seen in the nature of polynomials. As was pointed out in the last chapter on interpolation, high degree polynomials tend to oscillate between the points of constraint. Since the derivative of a polynomial is itself a polynomial, it too will oscillate between the points of constraint, but perhaps not quite so wildly. To minimize this oscillation, one must use low degree polynomials which then tend to reduce the accuracy of the approximation. Another way to see the dangers of numerical differentiation is to consider the nature of the operator itself. Remember that

$$\frac{df(x)}{dx} = \operatorname*{Lim}_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \quad . \tag{4.1.1}$$

Since there are always computational errors associated with the calculation of f(x), they will tend to be present as $\Delta x \to 0$, while similar errors will not be present in the calculation of Δx itself. Thus the ratio will end up being largely determined by the computational error in f(x). Therefore numerical differentiation should only be done if no other method for the solution of the problem can be found, and then only with considerable circumspection.

a. Classical Difference Formulae

With these caveats clearly in mind, let us develop the formalisms for numerically differentiating a function f(x). We have to approximate the continuous operator with a finite operator and the finite difference operators described in chapter 1 are the obvious choice. Specifically, let us take the finite difference operator to be defined as it was in equation (1.5.1). Then we may approximate the derivative of a function f(x) by

$$\frac{df(x)}{dx} = \frac{\Delta f(x)}{\Delta x} \quad . \tag{4.1.2}$$

The finite difference operators are linear so that repeated operations with the operator lead to

$$\Delta^n f(x) = \Delta[\Delta^{n-1} f(x)] \quad . \tag{4.1.3}$$

This leads to the *Fundamental Theorem of the Finite Difference Calculus* which is

> *The* nth *difference of a polynomial of degree* n *is a constant* (a_n n! h^n)*, and the* (n+1) *st difference is zero.*

Clearly the extent to which equation (4.1.3) is satisfied will depend partly on the value of h. Also the ability to repeat the finite difference operation will depend on the amount of information available. To find a non-trivial nth order finite difference will require that the function be approximated by an nth degree polynomial which has n+1 linearly independent coefficients. Thus one will have to have knowledge of the function for at least n+1 points. For example, if one were to calculate finite differences for the function x^2 at a finite set of points x_i, then one could construct a finite difference table of the form:

Table 4.1

A Typical Finite Difference Table for f(x) = x^2

x_i	$f(x_i)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
2	f(2)=4			
		$\Delta f(2)=5$		
3	f(3)=9		$\Delta^2 f(2)=2$	
		$\Delta f(3)=7$		$\Delta^3 f(2)=0$
4	f(4)=16		$\Delta^2 f(3)=2$	
		$\Delta f(4)=9$		$\Delta^3 f(3)=0$
5	f(5)=25		$\Delta^2 f(4)=2$	
		$\Delta f(5)=11$		
6	f(6)=36			

This table nicely demonstrates the fundamental theorem of the finite difference calculus while pointing out an additional problem with repeated differences. While we have chosen f(x) to be a polynomial so that the differences are exact and the fundamental theorem of the finite difference calculus is satisfied exactly, one can imagine the situation that would prevail should f(x) only approximately be a polynomial. The truncation error that arises from the approximation would be quite significant for $\Delta f(x_i)$ and compounded for $\Delta^2 f(x_i)$. The propagation of the truncation error gets progressively worse as one proceeds to higher and higher differences. The table illustrates an additional problem with finite differences. Consider the values of $\Delta f(x_i)$. They are not equal to the values of the derivative at x_i implied by the definition of the forward difference operator at which they are meant to apply. For example Δf(3)=7 and with h=1 for this table would suggest that f '(3)=7, but simple differentiation of the polynomial will show that f '(3)=6. One might think that this could be corrected by averaging Δf (2) and Δf (3), or by re-defining the difference operator so that it didn't always refer backward. Such an operator is known as the central difference operator which is defined as

$$\delta f(x) \equiv f(x+\tfrac{1}{2}h) - f(x-\tfrac{1}{2}h) \ . \qquad (4.1.4)$$

However, this does not remove the problem that the value of the nth difference, being derived from information spanning a large range in the independent variable, may not refer to the nth derivative at the point specified by the difference operator.

In Chapter 1 we mentioned other finite difference operators, specifically the shift operator E and the identity operator I (see equation 1.5.3). We may use these operators and the relation between them given by equation (1.5.4), and the binomial theorem to see that

$$\Delta^k[f(x)] = [E-I]^k[f(x)] = \sum_{i=0}^{k}(-1)^k\left(\frac{k}{i}\right)E^i[f(x)] = \sum_{i=0}^{k}(-1)^{k-1}\left(\frac{k}{i}\right)f(x+i) \quad , \tag{4.1.5}$$

where (^k_i) is the binomial coefficient which can be written as

$$\left(\frac{k}{i}\right) = \frac{k!}{(k-i)!i!} \quad . \tag{4.1.6}$$

One can use equation (4.1.5) to find the kth difference for equally spaced data without constructing the entire difference table for the function. If a specific value of $f(x_j)$ is missing from the table, and one assumes that the function can be represented by a polynomial of degree k-1, then, since $\Delta^k f\,(x_i) = 0$, equation (4.1.5) can be solved for the missing value of $f(x_j)$.

While equation (4.1.5) can be used to find the differences of any equally spaced function $f(x_i)$ and hence is an estimate of the kth derivative, the procedure is equivalent to finding the value of a polynomial of degree n-k at a specific value of x_i. Therefore, we may use any interpolation formula to obtain an expression for the derivative at some specific point by differentiation of the appropriate formula. If we do this for Lagrangian interpolation, we obtain

$$\Phi'(x) = \sum_{i=1}^{n} f(x_i)L_i^{'}(x) \quad , \tag{4.1.7}$$

where

$$L_i{}'(x) = \sum_{k=1}^{n}\prod_{\substack{j\neq i \\ j\neq k}}^{n} {}^{(x-x_j)}\!\Big/\!{}_{(x_i-x_j)} \quad . \tag{4.1.8}$$

Higher order formulae can be derived by successive differentiation, but one must always use numerical differentiation with great care.

b. Richardson Extrapolation for Derivatives

We will now consider a "clever trick" that enables the improvement of nearly all formulae that we have discussed so far in this book and a number yet to come. It is known as Richardson extrapolation, but differs from what is usually meant by extrapolation. In chapter 3 we described extrapolation in terms of extending some approximation formula beyond the range of the data which constrained that formula. Here we use it to describe a process that attempts to approximate the results of any difference or difference based formula to limit where the spacing h approaches zero. Since h is usually a small number, the extension, or extrapolation, to zero doesn't seem so unreasonable. Indeed, it may not seem very important, but remember the limit of the accuracy on nearly all approximation formulae is set by the influence of round-off error in the case where an approximating interval becomes small. This will be particularly true for problems of the numerical solution of differential equations discussed in the next chapter. However, we can develop and use it here to obtain expressions for derivatives that have greater accuracy and are obtained with greater efficiency than the classical difference formulae. Let us consider the special case where a function f(x) can be represented by a Taylor series so that if

$$x = x_0 + kh \ , \qquad (4.1.9)$$

then

$$f(x_0 + kh) = f(x_0) + khf'(x_0) + \frac{(kh)^2 f''(x_0)}{2!} + \frac{(kh)^3 f^{(3)}(x_0)}{3!} + \cdots + \frac{(kh)^n f^{(n)}(x_0)}{n!} \ . \qquad (4.1.10)$$

Now let us make use of the fact that h appears to an odd power in even terms of equation (4.1.10). Thus if we subtract the a Taylor series for -k from one for +k, the even terms will vanish leaving

$$f(x_0 + kh) - f(x_0 - kh) = 2khf'(x_0) + \frac{2(kh)^3 f^{(3)}(x_0)}{3!} + \cdots + \frac{(kh)^{2n+1} f^{(2n+1)}(x_0)}{(2n+1)!} \ . \qquad (4.1.11)$$

The functional relationship on the left hand side of equation (4.1.11) is considered to be some mathematical function whose value is precisely known, while the right hand side is the approximate relationship for that function. That relationship now only involves odd powers of h so that it converges much faster than the original Taylor series. Now evaluate equation (4.1.11) for k = 1 and 2 explicitly keeping just the first two terms on the right hand side so that

$$\left.\begin{array}{l} f(x_0 + h) - f(x_0 - h) = 2hf'(x_0) + 2h^3 f^{(3)}(x_0)/6 \quad + \cdots + R(h^5) \\ f(x_0 + 2h) - f(x_0 - 2h) = 4hf'(x_0) + 16h^3 f^{(3)}(x_0)/6 \quad + \cdots + \tilde{R}(h^5) \end{array}\right\} \ . \qquad (4.1.12)$$

We now have two equations from which the term involving the third derivative may be eliminated yielding

$$f(x_0\text{-}2h)\text{-}8f(x_0\text{-}h)\text{+}8f(x_0\text{+}h)\text{-}f(x_0\text{+}2h) = -12hf'(x_0)+R(h^5)\text{-}\tilde{R}(h^5) \ , \qquad (4.1.13)$$

and solving for $f'(x_0)$ we get.

$$f'(x_0) = [f(x_0\text{-}2h) - 8f(x_0\text{-}h) + 8f(x_0\text{+}h) - f(x_0\text{+}2h)]/(12h) + O(h^4). \qquad (4.1.14)$$

It is not hard to show that the error term in equation (4.1.13) divided by h is $O(h^4)$. Thus we have an expression for the *derivative* of the function f(x) evaluated at some value of $x = x_0$ which requires four values of the function and is exact for cubic polynomials. This is not too surprising as we have four free parameters with which to fit a Taylor series or alternately a cubic polynomial and such polynomials will be unique. What is surprising is the rapid rate of convergence with decreasing interval h. But what is even more amazing is that this method can be generalized to any approximation formulae that can be written as

$$\left.\begin{array}{l} f(x) = \Phi(x,\alpha h) + Ch^n + O(h^m) \\ m > n,\ \alpha > 0,\ \alpha \neq 1 \end{array}\right\} \ . \qquad (4.1.15)$$

so that

$$f(x) = \frac{\alpha^n \Phi(x,h) - \Phi(x,\alpha h)}{\alpha^n - 1} + O(h^m) \ . \qquad (4.1.16)$$

Indeed, it could be used to obtain an even higher order approximation for the derivative utilizing more tabular points. We shall revisit this method when we consider the solution to differential equations in Chapter 5.

4.2 Numerical Evaluation of Integrals: Quadrature

While the term *quadrature* is an old one, it is the correct term to use for describing the numerical evaluation of integrals. The term *numerical integration* should be reserved for describing the numerical solution of differential equations (see chapter 5). There is a genuine necessity for the distinction because the very nature of the two problems is quite different. Numerically evaluating an integral is a rather common and usually stable task. One is basically assembling a single number from a series of independent evaluations of a function. Unlike numerical differentiation, numerical quadrature tends to average out random computational errors.

Because of the inherent stability of numerical quadrature, students are generally taught only the simplest of techniques and thereby fail to learn the more sophisticated, highly efficient techniques that can be so important when the integrand of the integral is extremely complicated or occasionally the result of a separate lengthy study. Virtually all numerical quadrature schemes are based on the notion of polynomial approximation. Specifically, the quadrature scheme will give the exact value of the integral if the integrand is a polynomial of some degree n. The scheme is then said to have a *degree of precision* equal to n. In general, since a nth degree polynomial has n+1 linearly independent coefficients, a quadrature scheme will have to have n+1 adjustable parameters in order to accurately represent the polynomial and its integral. Occasionally, one comes across a quadrature scheme that has a degree of precision that is greater than the number of adjustable parameters. Such a scheme is said to be hyper-efficient and there are a number of such schemes known for multiple integrals. For single, or one dimensional, integrals, there is only one which we will discuss later.

a. The Trapezoid Rule

The notion of evaluating an integral is basically the notion of evaluating a sum. After all the integral sign ∫ is a stylized S that stands for a continuous "sum". The symbol Σ as introduced in equation (1.5.2) stands for a discrete or finite sum, which, if the interval is taken small enough, will approximate the value for the integral. Such is the motivation for the Trapezoid rule which can be stated as

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n-1} \frac{f(x_{i+1})+f(x_i)}{2}\Delta x_i \quad . \tag{4.2.1}$$

The formula takes the form of the sum of a discrete set of average values of the function each of which is multiplied by some sort of weight W_i. Here the weights play the role of the adjustable parameters of the quadrature formula and in the case of the trapezoid rule the weights are simply the intervals between functional evaluations. A graphical representation of this can be seen below in Figure 4.1

The meaning of the rule expressed by equation (4.2.1) is that the integral is approximated by a series of trapezoids whose upper boundaries in the interval Δx_i are straight lines. In each interval this formula would have a degree of precision equal to 1 (i.e. equal to the number of free parameters in the interval minus one). The other "adjustable" parameter is the 2 used in obtaining the average of $f(x_i)$ in the interval. If we divide the interval a → b equally then the Δx_i's have the particularly simple form

$$\Delta x_i = (b-a)/(n-1) \, . \tag{4.2.2}$$

In Chapter 3, we showed that the polynomic form of the integrand of an integral was unaffected by a linear transformation [see equations (3.3.19) and (3.3.20)]. Therefore, we can rewrite equation (4.2.1) as

$$\int_a^b f(x)\,dx = \frac{(b-a)}{2}\int_{-1}^{+1} f(y)\,dy = \frac{(b-a)}{2}\sum_{i=1}^{n}\frac{f[x(y_{i+1})]+f[x(y_i)]}{2}W'_i \quad , \tag{4.2.3}$$

where the weights for an equally spaced interval are

$$W'_i = 2/(n\text{-}1) \quad . \tag{4.2.4}$$

If we absorb the factor of (b-a)/2 into the weights we see that for both representations of the integral [i.e. equation (4.2.1) and equation (4.2.3)] we get

$$\sum_{i=1}^{n} W_i = b - a \quad . \tag{4.2.5}$$

Notice that the function f(x) plays absolutely no role in determining the weights so that once they are determined; they can be used for the quadrature of any function. Since any quadrature formula that is exact for polynomials of some degree greater than zero must be exact for $f(x) = x^0$, the sum of the weights of any quadrature scheme must be equal to the total interval for which the formula holds.

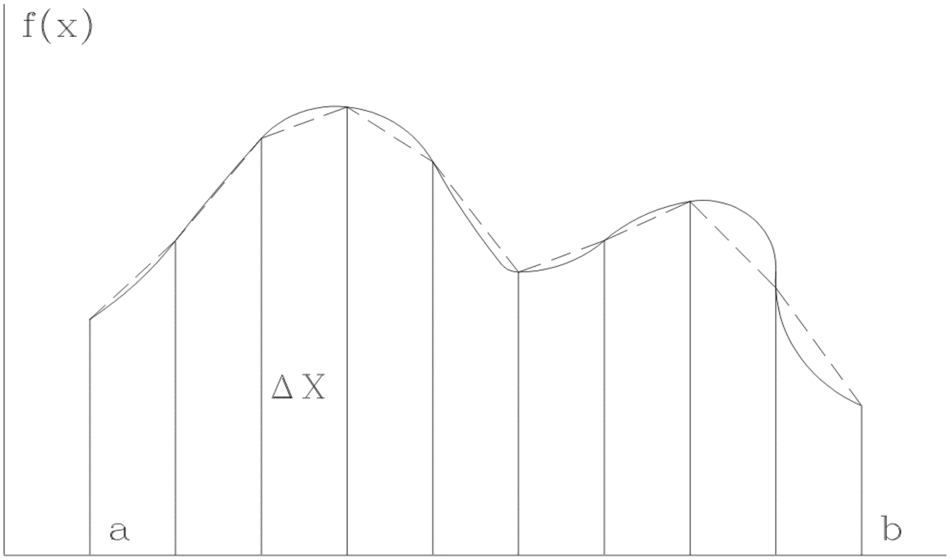

Figure 4.1 shows a function whose integral from a *to* b *is being evaluated by the trapezoid rule. In each interval a straight line approximates the function* Δx_i.

b. Simpson's Rule

The trapezoid rule has a degree of precision of 1 as it fits straight lines to the function in the interval. It would seem that we should be able to do better than this by fitting a higher order polynomial to the function. So instead of using the functional values at the endpoints of the interval to represent the function by a straight line, let us try three equally spaced points. That should allow us to fit a polynomial with three adjustable parameters (i.e. a parabola) and obtain a quadrature formula with a degree of precision of 2. However, we shall see that this quadrature formula actually has a degree of precision of 3 making it a hyper-efficient quadrature formula and the only one known for integrals in one dimension.

In general, we can construct a quadrature formula from an interpolation formula by direct integration. In chapter 3 we developed interpolation formulae that were exact for polynomials of an arbitrary degree n. One of the more general forms of these interpolation formulae was the Lagrange interpolation formula given by equation (3.2.8). In that equation Φ(x) was a polynomial of degree n and was made up of a linear combination of the Lagrange polynomials $L_i(x)$. Since we are interested in using three equally spaced points, n will be 2. Also, we have seen that any finite interval is equivalent to any other for the purposes of fitting polynomials, so let us take the interval to be 2h so that our formula will take the form

$$\int_0^{2h} f(x)\,dx = \sum_{i=0}^{2} f(x_i)W_i = \sum_{i=0}^{2} f(x_i)\int_0^{2h} L_i(x)\,dx \quad . \tag{4.2.6}$$

Here we see that the quadrature weights W_i are given by

$$W_i = \int_0^{2h} L_i(x)\,dx = \int_0^{2h} \prod_{\substack{j\neq i\\ j=0}}^{2} \frac{(x-x_i)}{(x_i-x_j)}\,dx \quad . \tag{4.2.7}$$

Now the three equally spaced points in the interval 2h will have x = 0, h, and 2h. For equal intervals we can use equation (3.2.11) to evaluate the Lagrange polynomials to get

$$\left.\begin{aligned} L_0(x) &= \frac{(x-h)(x-2h)}{2h^2} = \frac{(x^2-3xh+2h^2)}{2h^2} \\ L_1(x) &= \frac{(x-0)(x-2h)}{h^2} = \frac{(x^2-2xh)}{h^2} \\ L_2(x) &= \frac{(x-0)(x-h)}{2h^2} = \frac{(x^2-xh)}{2h^2} \end{aligned}\right\} \quad . \tag{4.2.8}$$

Therefore the weights for Simpson's rule become

$$\left.\begin{aligned} W_0 &= \int_0^{2h} L_0(x)\,dx = \frac{(8h^3/3-12h^3/2+4h^3)}{2h^2} = \frac{h}{3} \\ W_1 &= \int_0^{2h} L_1(x)\,dx = \frac{(8h^3/3-8h^3/2)}{h^2} = \frac{4h}{3} \\ W_2 &= \int_0^{2h} L_2(x)\,dx = \frac{(8h^3/3-4h^3/2)}{h^2} = \frac{h}{3} \end{aligned}\right\} \quad . \tag{4.2.9}$$

Actually we need only to have calculated two of the weights since we know that the sum of the weights had to be 2h. Now since h is only half the interval we can write

$$h = \Delta x/2 \quad , \tag{4.2.10}$$

so that the approximation formula for Simpson's quadrature becomes

$$\int_0^{\Delta x} f(x)\,dx = \sum_{i=0}^{2} f(x_i)W_i = \frac{\Delta x}{6}\left[f(x_0)+4f(x_1)+f(x_2)\right] \;. \tag{4.2.11}$$

Now let us confirm the assertion that Simpson's rule is hyper-efficient. We know that the quadrature formula will yield exact answers for quadratic polynomials, so consider the evaluation of a quartic. We pick the extra power of x in anticipation of the result. Thus we can write

$$\begin{aligned}\int_0^{\Delta x}(\alpha x^3+\beta x^4)\,dx &= \frac{\alpha\Delta x^4}{4}+\frac{\beta\Delta x^5}{5} = \frac{\Delta x}{6}\left(4\alpha\left[\frac{\Delta x}{2}\right]^3+\alpha(\Delta x)^3+4\beta\left[\frac{\Delta x}{2}\right]^4+\beta(\Delta x)^4\right)+R(\Delta x)\\ &= \frac{\alpha(\Delta x)^4}{4}+\frac{5\beta(\Delta x)^5}{24}+R(\Delta x) \qquad .\end{aligned} \tag{4.2.12}$$

Here $R(\Delta x)$ is the error term for the quadrature formula. Completing the algebra in equation (4.2.12) we get

$$R(\Delta x) = \beta(\Delta x)^5/120 \;\; . \tag{4.2.13}$$

Clearly the error in the integral goes as the interval to the fifth power and not the fourth power. So the quadrature formula will have no error for cubic terms in the integrand and the formula is indeed hyper-efficient. Therefore Simpson's rule is a surprisingly good quadrature scheme having a degree of precision of 3 over the interval Δx. Should one wish to span a larger interval (or reduce the spacing for a given interval), one could write

$$\int_0^{h\Delta x} f(x)\,dx = \sum_{i=1}^{n}\int_{(i-1)\Delta x}^{i\Delta x} f(x_i)\,dx = \frac{\Delta x}{6}\left[f(x_1)+4f(x_2)+2f(x_3)+4f(x_4)+\cdots+4f(x_{n-1})+f(x_n)\right] \;. \tag{4.2.14}$$

By breaking the integral up into sub-intervals, the function need only be well approximated locally by a cubic. Indeed, the function need not even be continuous across the separate boundaries of the sub-intervals. This form of Simpson's rule is sometimes called a *running Simpson's rule* and is quite easy to implement on a computer. The hyper-efficiency of this quadrature scheme makes this a good "all purpose" equal interval quadrature algorithm.

c. Quadrature Schemes for Arbitrarily Spaced Functions

As we saw above, it is possible to obtain a quadrature formula from an interpolation formula and maintain the same degree of precision as the interpolation formula. This provides the basis for obtaining quadrature formula for functions that are specified at arbitrarily spaced values of the independent variable x_i. For example, simply evaluating equation (4.2.6) for an arbitrary interval yields

$$\int_a^b f(x)\,dx = \sum_{i=0}^{n} f(x_i)\int_a^b L_i(x)\,dx \quad , \tag{4.2.15}$$

which means that the weights associated with the arbitrarily spaced points x_i are

$$W_i = \int_a^b L_i(x)\,dx \;\; . \tag{4.2.16}$$

However, the analytic integration of $L_i(x)$ can become tedious when n becomes large so we give an alternative strategy for obtaining the weights for such a quadrature scheme. Remember that the scheme is to have a degree of precision of n so that it must give the exact answers for *any* polynomial of degree n. But there can only be one set of weights, so we specify the conditions that must be met for a set of polynomials for which we know the answer - namely x^i. Therefore we can write

$$\int_a^b x^i\,dx = \frac{b^{i+1} - a^{i+1}}{i+1} = \sum_{j=0}^{n} x_j^i W_j \quad , \quad i = 0 \cdots n \ . \tag{4.2.17}$$

The integral on the left is easily evaluated to yield the center term which must be equal to the sum on the right if the formula is to have the required degree of precision n. Equations (4.2.17) represent n+1 linear equations in the n+1 weights W_i. Since we have already discussed the solution of linear equations in some detail in chapter 2, we can consider the problem of finding the weights to be solved.

While the spacing of the points given in equations (4.2.17) is completely arbitrary, we can use these equations to determine the weights for Simpson's rule as an example. Assume that we are to evaluate an integral in the interval $0 \rightarrow 2h$. Then the equations (4.2.17) for the weights would be

$$\int_0^{2h} x^i\,dx = \frac{(2h)^{i+1}}{i+1} = \sum_{j=0}^{n} x_j^i W_j \quad , \quad i = 0 \cdots n \ . \tag{4.2.18}$$

For $x_j = [0,h,2h]$, the equations specifically take the form

$$\left.\begin{aligned} & 2h = W_1 + W_2 + W_3 \\ & \frac{(2h)^2}{2} = 2h^2 = h^2W_2 + h^2W_3 \\ & \frac{(2h)^3}{3} = \frac{8h^3}{3} = h^2W_2 + 4h^2W_3 \end{aligned}\right\} \ . \tag{4.2.19}$$

which upon removal of the common powers of h are

$$\left.\begin{aligned} & 2h = W_1 + W_2 + W_3 \\ & 2h = W_2 + W_3 \\ & \frac{8h}{3} = W_2 + 4W_3 \end{aligned}\right\} \ . \tag{4.2.20}$$

These have the solution

$$W_i = [1/3,\ 4/3,\ 1/3]h \ . \tag{4.2.21}$$

The weights given in equation (4.2.21) are identical to those found for Simpson's rule in equation (4.2.9) which lead to the approximation formula given by equation (4.2.11). The details of finding the weights by this method are sufficiently simple that it is generally preferred over the method discussed in the previous section (section 4.2b).

There are still other alternatives for determining the weights. For example, the integral in equation (4.2.16) is itself the integral of a polynomial of degree n and as such can be evaluated exactly by any quadrature scheme with that degree of precision. It need not have the spacing of the desired scheme at all. Indeed, the integral could be evaluated at a sufficient level of accuracy by using a running Simpson's rule with a sufficient total number of points. Or the weights could be obtained using the highly efficient Gaussian type quadrature schemes described below. In any event, a quadrature scheme can be tailored to fit nearly any problem by writing down the equations of condition that the weights must satisfy in order to have the desired degree of precision. There are, of course, some potential pitfalls with this approach. If very high degrees of precision formulae are sought, the equations (4.2.17) may become nearly singular and be quite difficult to solve with the accuracy required for reliable quadrature schemes. If such high degrees of precision formulae are really required, then one should consider Gaussian quadrature schemes.

d. Gaussian Quadrature Schemes

We turn now to a class of quadrature schemes first suggested by that brilliant 19th century mathematician Karl Friedrich Gauss. Gauss noted that one could obtain a much higher degree of precision for a quadrature scheme designed for a function specified at a given number of points, if the location of those points were regarded as additional free parameters. So, if in addition to the N weights one also had N locations to specify, one could obtain a formula with a degree of precision of 2N-1 for a function specified at only N points. However, they would have to be the proper N points. That is, their location would no longer be arbitrary so that the function would have to be known at a particular set of values of the independent variable x_i. Such a formula would not be considered a hyper-efficient formula since the degree of precision does not exceed the number of adjustable parameters. One has simply enlarged the number of such parameters available in a given problem.

The question then becomes how to locate the proper places for the evaluation of the function given the fact that one wishes to obtain a quadrature formula with this high degree of precision. Once more we may appeal to the notion of obtaining a quadrature formula from an interpolation formula. In section (3.2b) we developed Hermite interpolation which had a degree of precision of 2N-1. (Note: in that discussion the actual numbering if the points began with zero so that N=n+1 where n is the limit of the sums in the discussion.) Since equation (3.2.12) has the required degree of precision, we know that its integral will provide a quadrature formula of the appropriate degree. Specifically

$$\int_a^b \Phi(x)\,dx = \sum_{j=0}^{n} f(x_j)\int_a^b h_j(x)\,dx + \sum_{j=0}^{n} f'(x_j)\int_a^b H_j(x)\,dx \quad . \tag{4.2.22}$$

Now equation (4.2.22) would resemble the desired quadrature formula if the second sum on the right hand side could be made to vanish. While the weight functions $H_j(x)$ themselves will not always be zero, we can ask under what conditions their integral will be zero so that

$$\int_a^b H_j(x)\,dx = 0 \quad . \tag{4.2.23}$$

Here the secret is to remember that those weight functions are polynomials [see equation (3.2.32)] of degree 2n+1 (i.e. 2N-1) and in particular $H_j(x)$ can be written as

$$H_i(x) = \frac{\prod(x)L_i(x)}{\prod_{j\neq i}^{n}(x_i - x_j)} \quad , \tag{4.2.24}$$

where

$$\prod(x) \equiv \prod_{j=0}^{n}(x - x_j) \quad . \tag{4.2.25}$$

Here the additional multiplicative linear polynomial $u_j(x)$ that appears in equation has been included in one of the Lagrange polynomials $L_j(x)$ to produce the n+1 degree polynomial $\Pi(x)$. Therefore the condition for the weights of $f'(x_i)$ to vanish [equation(4.2.23)] becomes

$$\frac{\int_a^b \prod(x)L_i(x)\,dx}{\prod_{j\neq i}^{n}(x_i - x_j)} = 0 \quad . \tag{4.2.26}$$

The product in the denominator is simply a constant which is not zero so it may be eliminated from the equation. The remaining integral looks remarkably like the integral for the definition of orthogonal polynomials [equation (3.3.6)]. Indeed, since $L_i(x)$ is a polynomial of degree n [or (N-1)] and $\Pi(x)$ is a polynomial of degree n+1 (also N), the conditions required for equation (4.2.26) to hold will be met if $\Pi(x)$ is a member of the set of polynomials which are orthogonal in the interval $a \rightarrow b$. But we have not completely specified $\Pi(x)$ for we have not chosen the values x_j where the function f(x) and hence $\Pi(x)$ are to be evaluated. Now it is clear from the definition of $\Pi(x)$ [equation (4.2.25)] that the values of x_j are the roots of a polynomial of degree n+1 (or N) that $\Pi(x)$ represents. Thus, we now know how to choose the x_j's so that the weights of f'(x) will vanish. Simply choose them to be the roots of the (n+1)th degree polynomial which is a member on an orthogonal set on the interval $a \rightarrow b$. This will insure that the second sum in equation (4.2.22) will always vanish and the condition becomes

$$\int_a^b \Phi(x)\,dx = \sum_{j=0}^{n} f(x_j)\int_a^b h_j(x)\,dx \quad . \tag{4.2.27}$$

This expression is exact as long as $\Phi(x)$ is a polynomial of degree 2n+1 (or 2N-1) or less. Thus, Gaussian quadrature schemes have the form

$$\int_a^b f(x)\,dx = \sum_{j=0}^{n} f(x_j)W_j \quad , \tag{4.2.28}$$

where the x_i's are the roots of the Nth degree orthogonal polynomial which is orthogonal in the interval $a \rightarrow b$, and the weights W_i can be written with the aid of equation (3.2.32) as

$$W_i = \int_a^b h_i(x)\,dx = \int_a^b [1 - 2(x - x_i)L'_i(x)L_i^2(x)]\,dx \,. \tag{4.2.29}$$

Now these weights can be evaluated analytically should one have the determination, or they can be evaluated from the equations of condition [equation (4.2.17)] which any quadrature weights must satisfy. Since the extent of the finite interval can always be transformed into the interval –1 → +1 where the appropriate orthonormal polynomials are the Legendre polynomials, and the weights are independent of the function f(x), they will be specified by the value of N alone and may be tabulated once and for all. Probably the most complete tables of the roots and weights for Gaussian quadrature can be found in Abramowitz and Stegun[1] and unless a particularly unusual quadrature scheme is needed these tables will suffice.

Before continuing with our discussion of Gaussian quadrature, it is perhaps worth considering a specific example of such a formula. Since the Gaussian formulae make use of orthogonal polynomials, we should first express the integral in the interval over which the polynomials form an orthogonal set. To that end, let us examine an integral with a finite range so that

$$\int_a^b f(x)\,dx = \left(\frac{b-a}{2}\right)\int_{-1}^{+1} f\{[(b-a)y+(a+b)]/2\}\,dy \quad . \qquad (4.2.30)$$

Here we have transformed the integral into the interval –1 → +1. The appropriate transformation can be obtained by evaluating a linear function at the respective end points of the two integrals. This will specify the slope and intercept of the straight line in terms of the limits and yields

$$\left.\begin{array}{l} y=[2x-(a+b)]/(b-a) \\ dy=[2/(b-a)]dx \end{array}\right\} \quad . \qquad (4.2.31)$$

We have no complicating weight function in the integrand so that the appropriate polynomials are the Legendre polynomials. For simplicity, let us take n=2. We gave the first few Legendre polynomials in Table 3.4 and for n = 2 we have

$$P_2(y) = (3y^2-1)/2 \quad . \qquad (4.2.32)$$

The points at which the integrand is to be evaluated are simply the roots of that polynomial which we can fine from the quadratic formula to be

$$\left.\begin{array}{l} (3y^2-1)/2=0 \\ y_i=\pm\sqrt{3} \end{array}\right\} \quad . \qquad (4.2.33)$$

Quadrature formulae of larger n will require the roots of much larger degree polynomials which have been tabulated by Abramowitz and Stegun[1]. The weights of the quadrature formula are yet to be determined, but having already specified where the function is to be evaluated, we may use equations (4.2.17) to find them. Alternatively, for this simple case we need only remember that the weights sum to the interval so that

$$W_1 + W_2 = 2 \quad . \qquad (4.2.34)$$

Since the weights must be symmetric in the interval, they must both be unity. Substituting the values for y_i and W_i into equation (4.2.28), we get

$$\int_a^b f(x)\,dx \cong \tfrac{(b-a)}{2}\{f[({}^{(b-a)}\!/_{2\sqrt{3}})+\tfrac{1}{2}(a+b)]+f[({}^{(a-b)}\!/_{2\sqrt{3}})+\tfrac{1}{2}(a+b)]\} \quad . \qquad (4.2.35)$$

While equation (4.2.35) contains only two terms, it has a degree of precision of three (2n-1) or the same as the three term hyper-efficient Simpson's rule. This nicely illustrates the efficiency of the Gaussian schemes. They rapidly pass the fixed abscissa formulae in their degree of precision as [(2n-1)/n].

So far we have restricted our discussion of Gaussian quadrature to the finite interval. However, there is nothing in the entire discussion that would affect general integrals of the form

$$I = \int_\alpha^\beta w(x) f(x)\, dx \quad . \tag{4.2.36}$$

Here w(x) is a weight function which may not be polynomic and should not be confused with the quadrature weights W_i. Such integrals can be evaluated exactly as long as f(x) is a polynomial of degree 2N-1. One simply uses a Gaussian scheme where the points are chosen so that the values of x_i are the roots of the Nth degree polynomial that is orthogonal in the interval α → β *relative to the weight function* w(x). We have already studied such polynomials in section 3.3 so that we may use Gaussian schemes to evaluate integrals in the semi-infinite interval [0 → +∞] and full infinite interval [–∞ → +∞] as well as the finite interval [–1 → +1] as long as the appropriate weight function is used. Below is a table of the intervals and weight functions that can be used for some common types of Gaussian quadrature.

Table 4.2

Types of Polynomials for Gaussian Quadrature

INTERVAL	WEIGHT fUNCTION W(X)	TYPE OF POLYNOMIAL
-1 → +1	$(1-x^2)^{-½}$	Chebyschev: 1st kind
-1 → +1	$(1-x^2)^{+½}$	Chebyschev: 2nd kind
0 → +∞	e^{-x}	Laguerre
-∞ → +∞	e^{-x^2}	Hermite

It is worth noting from the entries in Table 4.2 that there are considerable opportunities for creativity available for the evaluation of integrals by a clever choice of the weight function. Remember that it is only f(x) of the product w(x)f(x) making up the integrand that need be well approximated by a polynomial in order for the quadrature formula to yield accurate answers. Indeed the weight function for Gaussian-Chebyschev quadrature of the first kind has singularities at the end points of the interval. Thus if one's integral has similar singularities, it would be a good idea to use Gauss-Chebyschev quadrature instead of Gauss-Legendre quadrature for evaluating the integral. Proper choice of the weight function may simply be used to improve the polynomic behavior of the remaining part of the integrand. This will certainly improve the accuracy of the solution.

In any event, the quadrature formulae can always be written to have the form

$$\int_\alpha^\beta w(x) f(x)\, dx = \sum_{j=0}^{n} f(x_j) W_j \quad , \tag{4.2.37}$$

where the weights, which may include the weight function w(x) can be found from

$$w_i = \int_\alpha^\beta w(x)h_i(x)\,dx \quad . \tag{4.2.38}$$

Here $h_i(x)$ is the appropriate orthogonal polynomial for the weight function and interval.

e. Romberg Quadrature and Richardson Extrapolation

So far we have given explicit formulae for the numerical evaluation of a definite integral. In reality, we wish the result of the application of such formulae to specific problems. Romberg quadrature produces this result without obtaining the actual form for the quadrature formula. The basic approach is to use the general properties of the equal-interval formulae such as the Trapezoid rule and Simpson's rule to generate the results for formulae successively applied with smaller and smaller step size. The results can be further improved by means of Richardson's extrapolation to yield results for formulae of greater accuracy [i.e. higher order $O(h^m)$]. Since the Romberg algorithm generates these results recursively, the application is extremely efficient, readily programmable, and allows an on-going estimate of the error. Let us define a step size that will always yield equal intervals throughout the interval a → b as

$$h_j = (b-a)/2^j \quad . \tag{4.2.39}$$

The general Trapezoid rule for an integral over this range can written as

$$F(b-a) = \int_a^b f(x)\,dx = \frac{h_j}{2}\left[f(a) + f(b) + 2\sum_{i=1}^{j-1} f(a+ih_j) \right] \quad . \tag{4.2.40}$$

The Romberg recursive quadrature algorithm states that the results of applying this formula for successive values of j (i.e. smaller and smaller step sizes h_j) can be obtained from

$$\left.\begin{aligned} F_j^0 &= \tfrac{1}{2}(F_{j-1}^0 + Q_{j-1}) \\ Q_{j-1} &= h_{j-1}\sum_{i=1}^{2^{(j-1)}} f[b+(i-\tfrac{1}{2})h_{j-1}] \\ F_0^0 &= (b-a)[f(a)+f(b)]/2 \end{aligned}\right\} \quad . \tag{4.2.41}$$

Each estimate of the integral will require $2^{(j-1)}$ evaluations of the function and should yield a value for the integral, but can have a degree of precession no greater than $2^{(j-1)}$. Since a sequence of j steps must be execute to reach this level, the efficiency of the method is poor compared to Gaussian quadrature. However the difference $(F_j^0 - F_{j-1}^0)$ does provide an continuous estimate of the error in the integral.

We can significantly improve the efficiency of the scheme by using Romberg extrapolation to improve the nature of the quadrature formulae that the iteration scheme is using. Remember that successive values of h differ by a factor of two. This is exactly the form that we used to develop the Richardson formula for the derivative of a function [equation (4.1.15)]. Thus we can use the generalization of the Richardson algorithm given by equation (4.1.15) and utilizing two successive values of F_j^0 to "extrapolate" to the result for a higher order formula. Each value of integral corresponding to the higher order quadrature formula can, in turn, serve as the basis for an additional extrapolation. This procedure also can be cast as a recurrence formula where

$$F_j^k = \frac{2^{2k} F_{j+1}^{k-1} - F_j^{k-1}}{2^{2k} - 1} \quad . \tag{4.2.42}$$

There is a trade off between the results generated by equation (4.2.42) and equation (4.2.41). Larger values of j produce values for F^k_j which correspond to decreasing values of h (see table 4.3). However, increasing values of k yield values for F^k_j which correspond to quadrature formulae smaller error terms, but with larger values of h. Thus it is not obvious which sequence, equation (4.2.41) or equation (4.2.42) will yield the better value for the integral.

In order to see how this method works, consider applying it to the analytic integral

$$\int_0^{+1} e^{5x}\, dx = \frac{e^5 - 1}{5} = 29.48263182 \quad . \tag{4.2.43}$$

Table 4.3

Sample Results for Romberg Quadrature

I	F_j^0	F_j^1	F_j^2	F_j^3	F_j^4
0	74.7066	33.0238	29.6049	29.4837	29.4827
1	43.4445	29.8186	29.4856	29.4826	
2	33.2251	29.5064	29.4827		
3	30.4361	29.4824			
4	29.722113				

Here it is clear that improving the order of the quadrature formula rapidly leads to a converged solution. The convergence of the non-extrapolated quadrature is not impressive considering the number of evaluations required to reach, say, F^0_4. Table 4.4 gives the results of applying some of the other quadrature methods we have developed to the integral in equation (4.2.43).

We obtain the results for the Trapezoid rule by applying equation (4.2.1) to the integral given by equation (4.2.43). The results for Simpson's rule and the two-point Gaussian quadrature come from equations (4.2.11) and (4.2.35) respectively. In the last two columns of Table 4.4 we have given the percentage error of the method and the number of evaluations of the function required for the determination of the integral. While the Romberg extrapolated integral is five times more accurate that it nearest competitor, it takes twice the number of evaluations. This situation gets rapidly worse so that the Gaussian quadrature becomes the most efficient and accurate scheme when n exceeds about five. The trapezoid rule and Romberg F_0^0 yield identical results as they are the same approximation. Similarly Romberg F^1_0 yields the same results as Simpson's rule. This is to be expected as the Richardson extrapolation of the Romberg quadrature equivalent to the Trapezoid rule should lead to the next higher order quadrature formula which is Simpson's rule.

Table 4.4

Test Results for Various Quadrature Formulae

Type	F(x)	|ΔF(%)|	N[f(x)]
Analytic Result	29.48263182	0.0	1
Trapezoid Rule	74.70658	153.39	2
Simpson's Rule	33.02386	12.01	3
2-point Gauss Quad.	27.23454	7.63	2
Romberg Quadrature F_0^0	74.70658	153.39	2
Romberg Quadrature F_1^1	29.8186	1.14	4

f. Multiple Integrals

Most of the work on the numerical evaluation of multiple integrals has been done in the middle of this century at the University of Wisconsin by Preston C. Hammer and his students. A reasonably complete summary of much of this work can be found in the book by Stroud[2]. Unfortunately the work is not widely known since problems associated with multiple integrals occur frequently in the sciences particularly in the area of the modeling of physical systems. From what we have already developed for quadrature schemes one can see some of the problems. For example, should it take N points to accurately represent an integral in one dimension, then it will take N^m points to calculate an m-dimensional integral. Should the integrand be difficult to calculate, the computation involved in evaluating it at N^m points can be prohibitive. Thus we shall consider only those quadrature formulae that are the most efficient - the Gaussian formulae. The first problem in numerically evaluating multiple integrals is to decide what will constitute an approximation criterion. Like integrals of one dimension, we shall appeal to polynomial approximation. That is, in some sense, we shall look for schemes that are exact for polynomials of the multiple variables that describe the multiple dimensions. However, there are many distinct types of such polynomials so we shall choose a subset. Following Stroud[2] let us look for quadrature schemes that will be exact for polynomials that can be written as simple products of polynomials of a single variable. Thus the approximating polynomial will be a product polynomial in m-dimensions. Now we will not attempt to derive the general theory for multiple Gaussian quadrature, but rather pick a specific space. Let the space be m-dimensional and of the full infinite interval. This allows us, for the moment, to avoid the problem of boundaries. Thus we can represent our integral by

$$V = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} e^{-(x_1^2+x_2^2+\cdots+x_m^2)} f(x_1, x_2, \cdots, x_m)\, dx_1 dx_2 \cdots dx_m \;. \qquad (4.2.44)$$

Now we have seen that we lose no generality by assuming that our nth order polynomial is a monomial of the form x^α so let us continue with this assumption that $f(x_1, x_2, \cdots x_m)$ has the form

$$f(x) = \prod_{i=1}^{n} x_i^{\alpha_i} \quad . \tag{4.2.45}$$

We can then write equation (4.2.44) as

$$V = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} e^{-\sum_{i=1}^{m} x_i^2} \prod_{j=1}^{m} x_j^{\alpha_j} dx_j = \prod_{j=1}^{m}\int_{-\infty}^{+\infty} e^{x_j^2} x_j^{\alpha_j} dx_j \quad . \tag{4.2.46}$$

The right hand side has the relatively simple form due to the linearity of the integral operator. Now make a coordinate transformation to general spherical coordinates by means of

$$\left.\begin{array}{l} x_1 = r\cos\theta_{m-1}\cos\theta_{m-2}\cdots\cos\theta_2\cos\theta_1 \\ x_2 = r\cos\theta_{m-1}\cos\theta_{m-2}\cdots\cos\theta_2\sin\theta_1 \\ \vdots \qquad\quad \vdots \qquad\quad \vdots \\ x_{m-1} = r\cos\theta_{m-1}\sin\theta_{m-2} \\ x_m = r\sin\theta_{m-1} \end{array}\right\} , \tag{4.2.47}$$

which has a Jacobian of the transformation equal to

$$J(x_i \,|\, r,\theta_i) = r^{m-1}\cos^{m-2}(\theta_{m-1})\cos^{m-3}(\theta_{m-2}) \cdots \cos(\theta_2) \ . \tag{4.2.48}$$

This allows the expression of the integral to take the form

$$V = \left(\int_{-\infty}^{+\infty} e^{-r^2} |r|^{m-1} r^{(\sum_{j=1}^{m}\alpha_i)} dr \right) \prod_{i=1}^{m-1} \int_{-\pi/2}^{+\pi/2} (\cos\theta_i)^{i-1} (\cos\theta_i)^{(-\sum_{j=1}^{m}\alpha_i)} (\sin\theta_i)^{\alpha_{i+1}} d\theta_i \ . \tag{4.2.49}$$

Consider how we could represent a quadrature scheme for any single integral in the running product. For example

$$\int_{-\pi/2}^{+\pi/2} (\cos\theta_i)^{i-1} (\cos\theta_i)^{\alpha_i} (\sin\theta_i)^{\alpha_{i+1}} d\theta_i = \sum_{j=1}^{N} B_{ij} (\cos\theta_i)^{\alpha_i} \quad . \tag{4.2.50}$$

Here we have chosen the quadrature points for θ_i to be at θ_{ij} and we have let

$$\alpha = \Sigma\alpha_i \ . \tag{4.2.51}$$

Now make one last transformation of the form $\qquad y_i = \cos\theta_i$, $\qquad$ (4.2.52)
which leads to

$$\int_{-1}^{+1} (1 - y_i^2)^{(i-2)/2} y_i^\alpha \, dy = \sum_{j=1}^{N} B_{ij} y_j = \int_{-1}^{+1} w(y_i) y_i^\alpha dy_i \quad , \quad i = 1\cdots(m-1) \quad . \tag{4.2.53}$$

The integral on the right hand side can be evaluated exactly if we take the y_i's to be the roots of a polynomial of degree $(\alpha+1)/2$ which is a member of an orthogonal set in the interval $-1 \rightarrow +1$, relative to the weight function $w(y_i)$ which is

$$w(y_i) = (1 - y_i^2)^{(i-2)/4} (1 + y_i^2)^{(i-2)/4} . \tag{4.2.54}$$

By considering Table 3.1 we see that the appropriate polynomials will be members of the Jacobi polynomials for α = β = (i-2)/4. The remaining integral over the radial coordinate has the form

$$\int_{-\infty}^{+\infty} e^{-r^2} r^{\alpha'} dr \ , \qquad (4.2.55)$$

which can be evaluated using Gauss-Hermite quadrature. Thus we see that multiple dimensional quadratures can be carried out with a Gaussian degree of precision for product polynomials by considering each integral separately and using the appropriate Gaussian scheme for that dimension. For example, if one desires to integrate over the solid sphere, one would choose Gauss-Hermite quadrature for the radial quadrature, Gauss-Legendre quadrature for the polar angle θ, and Gauss-Chebyschev quadrature for the azimuthal angle φ. Such a scheme can be used for integrating over the surface of spheres or surfaces that can be distorted from a sphere by a polynomial in the angular variables with good accuracy. The use of Gaussian quadrature schemes can save on the order of $N^{m/2}$ evaluations of the functions which is usually significant.

For multi-dimensional integrals, there are a number of hyper-efficient quadrature formulae that are known. However, they depend on the boundaries of the integration and are generally of rather low order. Nevertheless such schemes should be considered when the boundaries are simple and the function well behaved. When the boundaries are not simple, one may have to resort to a modeling scheme such a Monte Carlo method.

It is clear that the number of points required to evaluate an integral in m-dimensions will increase as N^m. It does not take many dimensions for this to require an enormous number of points and hence, evaluations of the integrand. Thus for multiple integrals, efficiency may dictate another approach.

4.3 Monte Carlo Integration Schemes and Other Tricks

The Monte Carlo approach to quadrature is a philosophy as much as it is an algorithm. It is an application of a much more widely used method due to John von Neumann. The method was developed during the Second World War to facilitate the solution to some problems concerning the design of the atomic bomb. The basic philosophy is to describe the problem as a sequence of causally related physical phenomena. Then by determining the probability that each separate phenomenon can occur, the joint probability that all can occur is a simple product. The procedure can be fashioned sequentially so that even probabilities that depend on prior events can be handled. One can conceptualize the entire process by following a series of randomly chosen initial states each of which initiates a causal sequence of events leading to the desired final state. The probability distribution of the final state contains the answer to the problem. While the method derives it name from the casino at Monte Carlo in order to emphasize the probabilistic nature of the method, it is most easily understood by example. One of the simplest examples of Monte Carlo modeling techniques involves the numerical evaluation of integrals.

a. Monte Carlo Evaluation of Integrals

Let us consider a one dimensional integral defined over a finite interval. The graph of the integrand might look like that in Figure 4.2. Now the area under the curve is related to the integral of the function.

Therefore we can replace the problem of finding the integral of the function to that of finding the area under the curve. However, we must place some units on the integral and we do that by finding the *relative area* under the curve. For example, consider the integral

$$\int_a^b f_{max} dx = (b-a) f_{max} \quad . \tag{4.3.1}$$

The graphical representation of this integral is just the area of the rectangle bounded by y = 0, x = a, x = b, and y = f_{max}. Now if we were to *randomly* select values of x_i and y_i, one could ask if

$$y_i \leq. f(x_i) \quad . \tag{4.3.2}$$

If we let ratio of the number of successful trials to the total number of trials be R, then

$$\int_a^b f(x) dx = R(b-a) f_{max} \quad . \tag{4 .3.3}$$

Clearly the accuracy of the integral will depend on the accuracy of R and this will improve with the number N of trials. In general, the value of R will approach its actual value as.N. This emphasizes the major difference between Monte Carlo quadrature and the other types of quadrature. In the case of the quadrature formulae that depend on a direct calculation of the integral, the error of the result is determined by the extent to which the integrand can be approximated by a polynomial (neglecting round-off error). If one is sufficiently determined he/she can determine the magnitude of the error term and thereby place an absolute limit on the magnitude of the error. However, Monte Carlo schemes are not based on polynomial approximation so such an absolute error estimate cannot be made even in principle. The best we can hope for is that there is a certain probability that the value of the integral lies within ε of the correct answer. Very often this is sufficient, but it should always remembered that the certainty of the calculation rests on a statistical basis and that the approximation criterion is different from that used in most areas of numerical analysis.

If the calculation of f(x) is involved, the time required to evaluate the integral may be very great indeed. This is one of the major drawbacks to the use of Monte Carlo methods in general. Another lesser problem concerns the choice of the random variables x_i and y_i. This can become a problem when very large numbers of random numbers are required. Most random number generators are subject to periodicities and other non-random behavior after a certain number of selections have been made. Any non-random behavior will destroy the probabilistic nature of the Monte Carlo scheme and thereby limit the accuracy of the answer. Thus, one may be deceived into believing the answer is better than it is. One should use Monte Carlo methods with great care. It should usually be the method of last choice. However, there are problems that can be solved by Monte Carlo methods that defy solution by any other method. This modern method of modeling the integral is reminiscent of a method used before the advent of modern computers. One simply graphed the integrand on a piece of graph paper and then cut out the area that represented the integral. By comparing the carefully measured weight of the cutout with that of a known area of graph paper, one obtained a crude estimate of the integral.

While we have discussed Monte Carlo schemes for one-dimensional integrals only, the technique can easily be generalized to multiple dimensions. Here the accuracy is basically governed by the number of points required to sample the "volume" represented by the integrand and limits. This sampling can generally be done more efficiently than the N^m points required by the direct multiple dimension quadrature schemes. Thus, the Monte-Carlo scheme is likely to efficiently compete with those schemes as the number of dimensions increases. Indeed, should m > 2, this is likely to be the case.

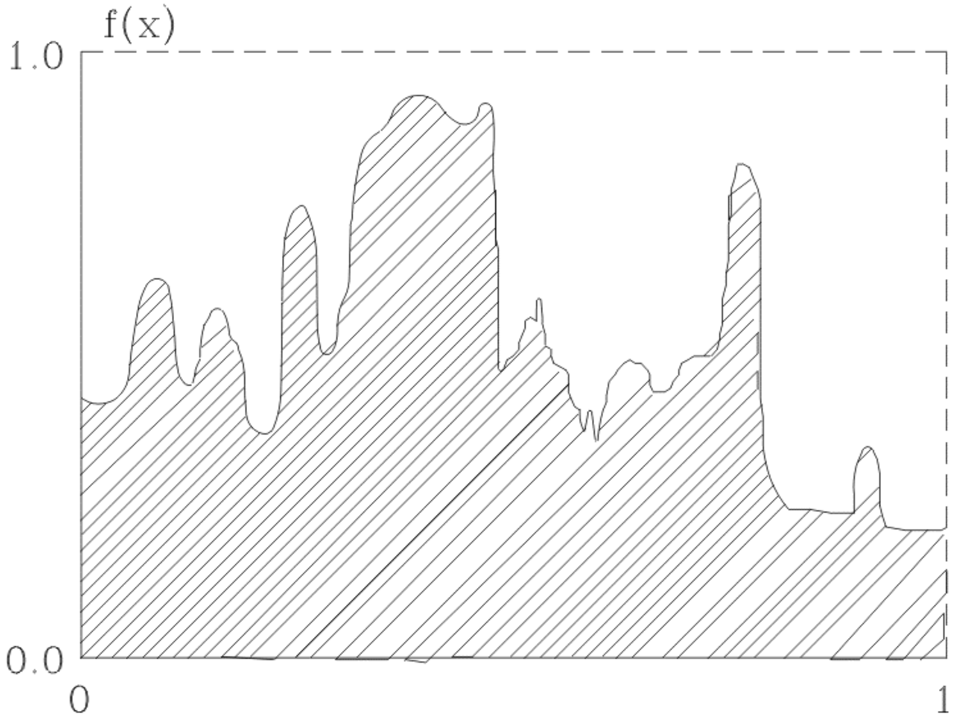

Figure 4.2 shows the variation of a particularly complicated integrand. Clearly it is not a polynomial and so could not be evaluated easily using standard quadrature formulae. However, we may use Monte Carlomethods to determine the ratio area under the curve compared to the area of the rectangle.

One should not be left with the impression that other quadrature formulae are without their problems. We cannot leave this subject without describing some methods that can be employed to improve the accuracy of the numerical evaluation of integrals.

b. The General Application of Quadrature Formulae to Integrals

Additional tricks that can be employed to produce more accurate answers involve the proper choice of the interval. Occasionally the integrand will display pathological behavior at some point in the interval. It is generally a good idea to break the interval at that point and represent the integral by two (or more) separate integrals each of which may separately be well represented by a polynomial. This is particularly useful in dealing with integrals on the semi-infinite interval, which have pathological integrands in the vicinity of zero. One can separate such an integral into two parts so that

$$\int_0^{+\infty} f(x)\,dx = \int_0^a f(x)\,dx + \int_a^{+\infty} f(x)\,dx \ . \tag{4.3.4}$$

The first of these can be transformed into the interval -1→ +1 and evaluated by means of any combination of the finite interval quadrature schemes shown in table 4.2. The second of these integrals can be transformed back into the semi-infinite interval by means of the linear transformation

$$y = x - a \ , \tag{4.3.5}$$

so that

$$\int_{a}^{+\infty} f(x)\,dx = \int_{0}^{+\infty} e^{-y}[e^{+y} f(y+a)\,dy \ . \tag{4.3.6}$$

Gauss-Laguerre quadrature can be used to determine the value of the second integral. By judiciously choosing places to break an integral that correspond to locations where the integrand is not well approximated by a polynomial, one can significantly increase the accuracy and ease with which integrals may be evaluated.

Having decided on the range over which to evaluate the integral, one has to pick the order of the quadrature formula to be used. Unlike the case for numerical differentiation, the higher the degree of precision of the quadrature formula, the better. However, there does come a point where the round-off error involved in the computation of the integrand exceeds the incremental improvement from the increased degree of precision. This point is usually difficult to determine. However, if one evaluates an integral with formulae of increasing degree of precision, the value of the integral will steadily change, reach a plateau, and then change slowly reflecting the influence of round-off error. As a rule of thumb 8 to 10 point Gauss-Legendre quadrature is sufficient to evaluate any integral over a finite range. If this is not the case, then the integral is somewhat pathological and other approaches should be considered. In some instances, one may use very high order quadrature (roots and weights for Legendre polynomials can be found up to N = 212), but these instances are rare. There are many other quadrature formulae that have utility in specific circumstances. However, should the quadrature present special problems, or require highly efficient evaluation, these formulae should be considered.

Chapter 4 Exercises

1. Numerically differentiate the function

   $$f(x) = e^{-x} ,$$

   at the points x = 0, .5, 1, 5, 10. Describe the numerical method you used and why you chose it. Discuss the accuracy by comparing your results with the analytic closed form derivatives.

2. Numerically evaluate

   $$f = \int_0^1 e^{-x}\, dx .$$

   Carry out this evaluation using
   - a. 5-point Gaussian quadrature
   - b. a 5-point equal interval formula that you choose
   - c. 5 point trapezoid rule
   - d. analytically.

   Compare and discuss your results.

3. Repeat the analysis of problem 2 for the integral

   $$\int_{-1}^{+1} |\, x \,|\, dx .$$

   Comment on your results

4. What method would you use to evaluate

   $$\int_1^{+\infty} (x^{-4} + 3x^{-2})\ \mathrm{Tanh}(x)\ dx\ ?$$

   Explain your choice.

5. Use the techniques described in section (4.2e) to find the volume of a sphere. Discuss all the choices you make regarding the type of quadrature use and the accuracy of the result.

Chapter 4 References and Supplemental Reading

1. Abramowitz, M. and Stegun, I.A., "Handbook of Mathematical Functions" National Bureau of Standards Applied Mathematics Series 55 (1964) U.S. Government Printing Office, Washington D.C.

2. Stroud, A.H., "Approximate Calculation of Multiple Integrals", (1971), Prentice-Hall Inc. Englewood Cliffs.

Because to the numerical instabilities encountered with most approaches to numerical differentiation, there is not a great deal of accessible literature beyond the introductory level that is available. For example

3. Abramowitz, M. and Stegun, I.A., "Handbook of Mathematical Functions" National Bureau of Standards Applied Mathematics Series 55 (1964) U.S. Government Printing Office, Washington D.C., p. 877, devote less than a page to the subject quoting a variety of difference formulae.

The situation with regard to quadrature is not much better. Most of the results are in technical papers in various journals related to computation. However, there are three books in English on the subject:

4. Davis, P.J., and Rabinowitz,P., "Numerical Integration", Blaisdell,

5. Krylov, V.I., "Approximate Calculation of Integrals" (1962) (trans. A.H.Stroud), The Macmillian Company

6. Stroud, A.H., and Secrest, D. "Gaussian Quadrature Formulas", (1966), Prentice-Hall Inc., Englewood Cliffs.

Unfortunately they are all out of print and are to be found only in the better libraries. A very good summary of various quadrature schemes can be found in

7. Abramowitz, M. and Stegun, I.A., "Handbook of Mathematical Functions" National Bureau of Standards Applied Mathematics Series 55 (1964) U.S. Government Printing Office, Washington D.C., pp. 885-899.

This is also probably the reference for the most complete set of Gaussian quadrature tables for the roots and weights with the possible exception of the reference by Stroud and Secrest (i.e. ref 4). They also give some hyper-efficient formulae for multiple integrals with regular boundaries. The book by Art Stroud on the evaluation of multiple integrals

6. Stroud, A.H., "Approximate Calculation of Multiple Integrals", (1971), Prentice-Hall Inc., Englewood Cliffs.

represents largely the present state of work on multiple integrals , but it is also difficult to find.