3.2 Numerical integration

Let us turn to numerical integration. In general, we want to obtain the numerical value of an integral, defined in the region $[a, b]$,

$$S = \int_a^b f(x)\,dx. \tag{3.21}$$

We can divide the region $[a, b]$ into n slices, evenly spaced with an interval h. If we label the data points as x_i with $i = 0, 1, \ldots, n$, we can write the entire integral as a summation of integrals, with each over an individual slice,

$$\int_a^b f(x)\,dx = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)\,dx. \tag{3.22}$$

If we can develop a numerical scheme that evaluates the summation over several slices accurately, we will have solved the problem. Let us first consider each slice separately. The simplest quadrature is obtained if we approximate $f(x)$ in the region $[x_i, x_{i+1}]$ linearly, that is, $f(x) \simeq f_i + (x - x_i)(f_{i+1} - f_i)/h$. After integrating over every slice with this linear function, we have

$$S = \frac{h}{2} \sum_{i=0}^{n-1} (f_i + f_{i+1}) + O(h^2), \tag{3.23}$$

where $O(h^2)$ comes from the error in the linear interpolation of the function. The above quadrature is commonly referred to as the *trapezoid rule*, which has an overall accuracy up to $O(h^2)$.

We can obtain a quadrature with a higher accuracy by working on two slices together. If we apply the Lagrange interpolation to the function $f(x)$ in the region $[x_{i-1}, x_{i+1}]$, we have

$$\begin{aligned} f(x) = {} & \frac{(x - x_i)(x - x_{i+1})}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} f_{i-1} + \frac{(x - x_{i-1})(x - x_{i+1})}{(x_i - x_{i-1})(x_i - x_{i+1})} f_i \\ & + \frac{(x - x_{i-1})(x - x_i)}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)} f_{i+1} + O(h^3). \end{aligned} \tag{3.24}$$

If we carry out the integration for every pair of slices together with the integrand given from the above equation, we have

$$S = \frac{h}{3} \sum_{j=0}^{n/2-1} (f_{2j} + 4f_{2j+1} + f_{2j+2}) + O(h^4), \tag{3.25}$$

which is known as the *Simpson rule*. The third-order term vanishes because of cancelation. In order to pair up all the slices, we have to have an even number of slices. What happens if we have an odd number of slices, or an even number of points in $[a, b]$? One solution is to isolate the last slice and we then have

$$\int_{b-h}^{b} f(x)\,dx = \frac{h}{12}(-f_{n-2} + 8f_{n-1} + 5f_n). \tag{3.26}$$

The expression for $f(x)$ in Eq. (3.24), constructed from the last three points of the function, has been used in order to obtain the above result. The following program is an implementation of the Simpson rule for calculating an integral.

We have used $f(x) = \sin x$ as the integrand in the above example program and $[0, \pi/2]$ as the integration region. The output of the above program is 1.000 008, which has six digits of accuracy compared with the exact result 1. Note that we have used only nine mesh points to reach such a high accuracy.

In some cases, we may not have the integrand given at uniform data points. The Simpson rule can easily be generalized to accommodate cases with nonuniform data points. We can rewrite the interpolation in Eq. (3.24) as

$$f(x) = ax^2 + bx + c, \tag{3.27}$$

where

$$a = \frac{h_{i-1} f_{i+1} - (h_{i-1} + h_i) f_i + h_i f_{i-1}}{h_{i-1} h_i (h_{i-1} + h_i)}, \tag{3.28}$$

$$b = \frac{h_{i-1}^2 f_{i+1} + \left(h_i^2 - h_{i-1}^2\right) f_i - h_i^2 f_{i-1}}{h_{i-1} h_i (h_{i-1} + h_i)}, \tag{3.29}$$

$$c = f_i, \tag{3.30}$$

with $h_i = x_{i+1} - x_i$. We have taken $x_i = 0$ because the integral

$$S_i = \int_{x_{i-1}}^{x_{i+1}} f(x)\, dx \tag{3.31}$$

is independent of the choice of the origin of the coordinates. Then we have

$$S_i = \int_{-h_{i-1}}^{h_i} f(x)\, dx = \alpha f_{i+1} + \beta f_i + \gamma f_{i-1}, \tag{3.32}$$

$$\alpha = \frac{2h_i^2 + h_i h_{i-1} - h_{i-1}^2}{6h_i}, \tag{3.33}$$

$$\beta = \frac{(h_i + h_{i-1})^3}{6h_i h_{i-1}}, \tag{3.34}$$

$$\gamma = \frac{-h_i^2 + h_i h_{i-1} + 2h_{i-1}^2}{6h_i}. \tag{3.35}$$

The last slice needs to be treated separately if $n + 1$ is even, as with the case of uniform data points. Then we have

$$S_n = \int_0^{h_{n-1}} f(x)\,dx = \alpha f_n + \beta f_{n-1} + \gamma f_{n-2}, \tag{3.36}$$

where

$$\alpha = \frac{h_{n-1}}{6}\left(3 - \frac{h_{n-1}}{h_{n-1} + h_{n-2}}\right), \tag{3.37}$$

$$\beta = \frac{h_{n-1}}{6}\left(3 + \frac{h_{n-1}}{h_{n-2}}\right), \tag{3.38}$$

$$\gamma = -\frac{h_{n-1}}{6}\frac{h_{n-1}^2}{h_{n-2}(h_{n-1} + h_{n-2})}. \tag{3.39}$$

The equations appear quite tedious but implementing them in a program is quite straightforward following the program for the case of uniform data points.

Even though we can make an order-of-magnitude estimate of the error occurring in either the trapezoid rule or the Simpson rule, it is not possible to control it because of the uncertainty involved in the associated coefficient. We can, however, develop an adaptive scheme based on either the trapezoid rule or the Simpson rule to make the error in the evaluation of an integral controllable. Here we demonstrate such a scheme with the Simpson rule and leave the derivation of a corresponding scheme with the trapezoid rule to Exercise 3.9.

If we expand the integrand $f(x)$ in a Taylor series around $x = a$, we have

$$\begin{aligned} S &= \int_a^b f(x)\,dx \\ &= \int_a^b \left[\sum_{k=0}^{\infty} \frac{(x-a)^k}{k!} f^{(k)}(a)\right] dx \\ &= \sum_{k=0}^{\infty} \frac{h^{k+1}}{(k+1)!} f^{(k)}(a), \end{aligned} \tag{3.40}$$

where $h = b - a$. If we apply the Simpson rule with $x_{i-1} = a$, $x_i = c = (a+b)/2$, and $x_{i+1} = b$, we have the zeroth-level approximation

$$\begin{aligned} S_0 &= \frac{h}{6}[f(a) + 4f(c) + f(b)] \\ &= \frac{h}{6}\left[f(a) + \sum_{k=0}^{\infty} \frac{h^k}{k!}\left(\frac{1}{2^{k-2}} - 1\right) f^{(k)}(a)\right]. \end{aligned} \tag{3.41}$$

We have expanded both $f(c)$ and $f(b)$ in a Taylor series around $x = a$ in the above equation. Now if we take the difference between S and S_0 and keep only the leading term, we have

$$\Delta S_0 = S - S_0 \approx -\frac{h^5}{2880} f^{(4)}(a). \tag{3.42}$$

We can continue to apply the Simpson rule in the regions $[a, c]$ and $[c, b]$. Then we obtain the first-level approximation

$$S_1 = \frac{h}{12} [f(a) + 4f(d) + 2f(c) + 4f(e) + f(b)], \tag{3.43}$$

where $d = (a + c)/2$ and $e = (c + b)/2$, and

$$\begin{aligned} \Delta S_1 &= S - S_1 \\ &\approx -\frac{(h/2)^5}{2880} f^{(4)}(a) - \frac{(h/2)^5}{2880} f^{(4)}(c) \\ &\approx -\frac{1}{2^4} \frac{h^5}{2880} f^{(4)}(a). \end{aligned} \tag{3.44}$$

We have used that $f^{(4)}(a) \approx f^{(4)}(c)$. The difference between the first-level and zeroth-level approximations is

$$S_1 - S_0 \approx -\frac{15}{16} \frac{h^5}{2880} f^{(4)}(a) \approx \frac{15}{16} \Delta S_0 \approx 15 \Delta S_1. \tag{3.45}$$

The above result can be used to set up the criterion for the error control in an adaptive algorithm. Consider, for example, that we want the error $|\Delta S| \leq \delta$. First we can carry out S_0 and S_1. Then we check whether $|S_1 - S_0| \leq 15\delta$, that is, whether $|\Delta S_1| \leq \delta$. If it is true, we return S_1 as the approximation for the integral. Otherwise, we continue the adaptive process to divide the region into two halves, four quarters, and so on, until we reach the desired accuracy with $|S - S_n| \leq \delta$, where S_n is the nth-level approximation of the integral. Let us here take the evaluation of the integral

$$S = \int_0^{\pi} \frac{[1 + a_0(1 - \cos x)^2]\, dx}{(1 + a_0 \sin^2 x)\sqrt{1 + 2a_0(1 - \cos x)}}, \tag{3.46}$$

for any $a_0 \geq 0$, as an example. The following program is an implementation of the adaptive Simpson rule for the integral above.