Chapter 3

Numerical integration

3.1 The closed Newton–Cotes method 57
3.2 Composite methods with equidistant nodes 58
3.3 Richardson extrapolation and Romberg's method 65
3.4 Methods with non-equidistant nodes 69
3.5 Introduction to probabilistic integration methods 73
3.6 Exercises 75
3.7 Discussion and bibliography 81

Introduction

Integrals are ubiquitous in science and mathematics. In this chapter, we are concerned with the problem of calculating numerically integrals of the form

$$I = \int_\Omega u(\boldsymbol{x}) \, \mathrm{d} \, \boldsymbol{x}, \tag{3.1}$$

Perhaps somewhat surprisingly, the numerical calculation of such integrals when $n \gg 1$ is still a very active area of research today. In this chapter, however, we will focus for simplicity on the one-dimensional setting where $\Omega = [a, b] \subset \mathbf{R}$. We assume throughout this chapter that the function u is Riemann-integrable. Then, by definition,

$$I = \lim_{h \to 0} \sum_{i=0}^{n-1} u(t_i)(z_{i+1} - z_i),$$

where $a = z_0 < \cdots < z_n = b$ is a partition of the interval $[a, b]$ such that the maximum spacing between successive x values is equal to h, and with $t_i \in [x_i, x_{i+1}]$ for all $i \in \{0, \ldots, n-1\}$.

All the numerical integration formulas that we present in this chapter are based on a deterministic approximation of the form

$$\hat{I} = \sum_{i=0}^{n} w_i u(x_i), \tag{3.2}$$

where $x_0 < \ldots < x_n$ are the *integration points* and $w_0, \ldots, w_n$ are the *integration weights*. In many cases, integration formulas contain a small parameter that can be refined to improve the accuracy of the approximation. In methods based on equidistant interpolation nodes, for example, this parameter encodes the distance between nodes and is typically denoted by h. We shall often use the notation $\hat{I}_h$ to emphasize the dependence of the approximation on h. The difference $E_h = I - \hat{I}_h$ is called the *integration error* or *discretization error*. The *degree of precision*, defined hereafter, is an important measure of the quality of quadrature rule.

Definition 3.1. The *degree of precision* of an integration method is the smallest integer number d such that the integration error is zero for all $u \in \mathbf{P}(d)$, i.e. for all the polynomials of degree less than or equal to d.

We observe that, without loss of generality, we can assume that the integration interval is equal to $[-1, 1]$. Indeed, using the change of variable

$$\begin{aligned} \zeta\colon [-1,1] &\to [a,b]; \\ y &\mapsto \frac{b+a}{2} + \frac{(b-a)}{2} y, \end{aligned} \tag{3.3}$$

we have

$$\int_a^b u(x)\,\mathrm{d}\,x = \int_{-1}^{1} u\big(\zeta(y)\big)\,\zeta'(y)\,\mathrm{d}\,y = \frac{b-a}{2}\int_{-1}^{1} u\circ\zeta(y)\,\mathrm{d}\,y, \tag{3.4}$$

and the right-hand side is the integral of $u \circ \zeta$ over the interval $[-1, 1]$.

3.1 The closed Newton–Cotes method

Given a set of equidistant points $-1 = x_0 < \cdots < x_n = 1$, a natural method for approximating the integral (3.1) of a function $u\colon [-1, 1] \to \mathbf{R}$ is to first construct the interpolating polynomial $\hat{u}$ at the nodes, and then calculate the exact integral of this polynomial. By construction, this method is exact for polynomials of degree up to n, and so the degree of precision is equal to *at least* n. Let $\varphi_0, \ldots, \varphi_n$ denote the Lagrange polynomials associated with the integration nodes. Then we have

$$I \approx \int_{-1}^{1} \hat{u}(x)\,\mathrm{d}\,x = \int_{-1}^{1} \sum_{i=0}^{n} u(x_i)\varphi_i(x)\,\mathrm{d}\,x = \sum_{i=0}^{n} u(x_i) \underbrace{\int_{-1}^{1} \varphi_i(x)\,\mathrm{d}\,x}_{w_i}.$$

The weights are independent of the function u, and so they can be calculated a priori. The class of integration methods obtained using this approach are known as *Newton–Cotes methods*. We present a few particular cases:

- $n = 1$, $d = 1$ (trapezoidal rule):

$$\int_{-1}^{1} u(x)\,dx \approx u(-1) + u(1). \tag{3.5}$$

- $n = 2$, $d = 3$ (Simpson's rule):

$$\int_{-1}^{1} u(x)\,dx \approx \frac{1}{3}u(-1) + \frac{4}{3}u(0) + \frac{1}{3}u(1). \tag{3.6}$$

- $n = 3$, $d = 3$ (Simpson's $\frac{3}{8}$ rule):

$$\int_{-1}^{1} u(x)\,dx \approx \frac{1}{4}u(-1) + \frac{3}{4}u(-1/3) + \frac{3}{4}u(1/3) + \frac{1}{4}u(1).$$

- $n = 4$, $d = 5$ (Boole's rule):

$$\int_{-1}^{1} u(x)\,dx \approx \frac{7}{45}u(-1) + \frac{32}{45}u\left(-\frac{1}{2}\right) + \frac{12}{45}u(0) + \frac{32}{45}u\left(\frac{1}{2}\right) + \frac{7}{45}u(1).$$

Remark 3.1. Note that, although it is based on a quadratic polynomial interpolation, Simpson's rule (3.6) has a degree of precision equal to 3. This is because any integration rule with nodes and weights symmetric around $x = 0$ is exact for odd functions, in particular x^3. Likewise, the degree of precision of Boole's rule is equal to 5.

In principle, this approach could be employed in order to construct integration rules of arbitrary high degree of precision. In practice, however, the weights become more and more imbalanced as the number of interpolation points increases, with some of them becoming negative. As a result, roundoff errors become increasingly detrimental to accuracy. In addition, in cases where the interpolating polynomial does not converge to u, for example if u is Runge's function, the approximate integral may not converge to the correct value in the limit as $n \to \infty$, even in exact arithmetic!

The integration rules presented in this section, which are based on equidistant nodes that include the endpoints of the integration interval, are called *closed Newton–Cotes* methods. A similar approach can be employed in order to construct to integration rules based on equidistant nodes that do not include the endpoints; these are called *open Newton–Cotes* methods.

3.2 Composite methods with equidistant nodes

A natural alternative to the approach presented in Section 3.1 is to construct an integration rule using piecewise polynomial interpolation, which we studied in Section 2.1.7. After partitioning the integration interval in a number of subintervals, the integral can be approximated by using one of the rules presented in Section 3.1 within each subinterval.

Composite trapezoidal rule. Let us illustrate the composite approach with an example. To this end, we introduce a partition $a = x_0 < \cdots < x_n = b$ of the interval $[a, b]$ and assume that

the nodes are equidistant with $x_{i+1} - x_i = h$. Using (3.4) with $a = x_i$ and $b = x_{i+1}$, we first generalize (3.5) to an interval $[x_i, x_{i+1}]$ as follows:

$$\int_{x_i}^{x_{i+1}} u(x)\,\mathrm{d}\,x = \frac{h}{2}\int_{-1}^{1} u\circ\zeta(y)\,\mathrm{d}\,y \approx \frac{h}{2}\Big(u\circ\zeta(-1) + u\circ\zeta(1)\Big) = \frac{h}{2}\Big(u(x_i) + u(x_{i+1})\Big),$$

where $\approx$ in this equation indicates approximation using the trapezoidal rule. Applying this approximation to each subinterval of the partition, we obtain the composite trapezoidal rule:

$$\begin{aligned}\int_a^b u(x)\,\mathrm{d}\,x &= \sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}} u(x)\,\mathrm{d}\,x \approx \frac{h}{2}\sum_{i=0}^{n-1}\Big(u(x_i) + u(x_{i+1})\Big)\\ &= \frac{h}{2}\Big(u(x_0) + 2u(x_1) + 2u(x_2) + \cdots + 2u(x_{n-2}) + 2u(x_{n-1}) + u(x_n)\Big). \quad (3.7)\end{aligned}$$

Like the trapezoidal rule (3.5), the composite trapezoidal rule (3.7) has a degree of precision equal to 1. However, the integration error of the method depends on the parameter h, which represents the width of each subinterval: for very small h, equation (3.7) is expected to provide a good approximation of the integral. An error estimate can be obtained directly from the formula in Theorem 2.3 for the interpolation error, provided that we assume that $u \in C^2[a,b]$.

Theorem 3.1 (Integration error for the composite trapezoidal rule)**.** *Let $\hat{I}_h$ denote the approximate integral calculated using (3.7). Then*

$$\big|I - \hat{I}_h\big| \leqslant \frac{b-a}{12} C_2 h^2, \qquad C_2 \coloneqq \sup_{\xi\in[a,b]} |u''(\xi)|. \quad (3.8)$$

Proof. Denoting by $\hat{u}_h$ the piecewise linear interpolation of u, we have

$$\int_{x_i}^{x_{i+1}} u(x) - \hat{u}(x)\,\mathrm{d}\,x = \frac{1}{2}\int_{x_i}^{x_{i+1}} u''\big(\xi(x)\big)(x - x_i)(x - x_{i+1})\,\mathrm{d}\,x.$$

Since $(x - x_i)(x - x_{i+1})$ is nonpositive over the interval $[x_i, x_{i+1}]$, we deduce that

$$\left|\int_{x_i}^{x_{i+1}} u(x) - \hat{u}(x)\,\mathrm{d}\,x\right| \leqslant \frac{1}{2}\left(\sup_{\xi\in[a,b]} |u''(\xi)|\right)\int_{x_i}^{x_{i+1}} (x - x_i)(x_{i+1} - x)\,\mathrm{d}\,x = C_2\frac{h^3}{12}.$$

Summing the contributions of all the intervals, we obtain

$$\big|I - \hat{I}_h\big| \leqslant \sum_{i=0}^{n-1}\left|\int_{x_i}^{x_{i+1}} u(x) - \hat{u}(x)\,\mathrm{d}\,x\right| \leqslant n \times C_2\frac{h^3}{12} = \frac{b-a}{12}C_2 h^2,$$

which concludes the proof. □

The integration error therefore scales as $\mathcal{O}(h^2)$. (Strictly speaking, we have shown only that the integration error admits an upper bound that scales at $\mathcal{O}(h^2)$, but it turns out that the dependence on h of this bound is optimal).

Example 3.1 (Integration error for the composite trapezoidal rule). Let us approximate the integral of x^2 over $[0,1]$ via the trapezoidal rule. In other words, let us calculate the right-hand side of

$$\int_0^1 x^2 \,\mathrm{d}\, x \approx \frac{h}{2}(x_0^2 + 2x_1^2 + \ldots + 2x_{n-1}^2 + x_n^2),$$

which we shall denote $\hat{I}_h$ hereafter. Using $x_i = ih$, and $x_n = 1$, we obtain

$$\hat{I}_h = \frac{h}{2}\left(2h^2 \sum_{i=1}^{n-1} i^2 + 1\right) = \frac{h}{2}\left(\frac{h^2}{3}(n-1)n(2n-1) + 1\right) = \frac{h}{2}\left(\frac{h^2}{3}(2n^3 - 3n^2 + n) + 1\right).$$

Now, observing that $n = 1/h$ yields

$$\hat{I}_h = \frac{h}{2}\left(\frac{h^2}{3}(2h^{-3} - 3h^{-2} + h^{-1}) + 1\right) = \frac{h}{2}\left(\frac{2}{3}h^{-1} - 1 + \frac{h}{3} + 1\right) = \frac{1}{3} + \frac{h^2}{6},$$

which is exactly the error shown by Theorem 3.1.

Composite Simpson rule. The composite Simpson rule is derived in Exercise 3.2. Given an odd number $n+1$ of equidistant points $a = x_0 < x_1 < \cdots < x_n = b$, this rule is given by

$$\hat{I}_h = \frac{h}{3}\Big(u(x_0) + 4u(x_1) + 2u(x_2) + 4u(x_3) + 2u(x_4) + \cdots + 2u(x_{n-2}) + 4u(x_{n-1}) + u(x_n)\Big). \quad (3.9)$$

This approximation is obtained by integrating the piecewise quadratic interpolant over a partition of the integration interval into $n/2$ subintervals of equal width. Obtaining an optimal error estimate, in terms of the dependence on h, for this integration formula is slightly more involved.

Theorem 3.2 (Integration error for the composite Simpson rule). *Let $\hat{I}_h$ denote the approximate integral calculated using (3.9). Then*

$$\left|I - \hat{I}_h\right| \leqslant (b-a)\frac{C_4 h^4}{180}, \qquad C_4 \coloneqq \sup_{\xi \in [a,b]} \left|u^{(4)}(\xi)\right|. \quad (3.10)$$

Proof. For a given subinterval $[x_{2i}, x_{2i+2}]$, let us denote by $\hat{u}_2$ the quadratic interpolating polynomial at $x_{2i}, x_{2i+1}, x_{2i+2}$, and by $\hat{u}_3(\bullet;\alpha)$ the cubic interpolating polynomial relative to the nodes $x_{2i}, x_{2i+1}, x_{2i+2}, \alpha$, for some $\alpha \in [x_{2i}, x_{2i+1}]$ that does not coincide with the integration nodes. We have

$$\int_{x_{2i}}^{x_{2i+2}} u(x) - \hat{u}_2(x) \,\mathrm{d}\, x = \int_{x_{2i}}^{x_{2i+2}} u(x) - \hat{u}_3(x;\alpha) \,\mathrm{d}\, x + \int_{x_{2i}}^{x_{2i+2}} \hat{u}_3(x;\alpha) - \hat{u}_2(x) \,\mathrm{d}\, x. \quad (3.11)$$

The second term on the right-hand side is zero, because the integrand is a cubic polynomial with zeros at x_{2i}, x_{2i+1} and x_{2i+2}, and because

$$\int_{x_{2i}}^{x_{2i+2}} (x - x_{2i})(x - x_{2i+1})(x - x_{2i+2}) = 0.$$

By Theorem 2.3, the first term in (3.11) is bounded from above as follows:

$$\begin{aligned}\left| \int_{x_{2i}}^{x_{2i+2}} u(x) - \hat{u}_3(x;\alpha) \,\mathrm{d}\, x \right| &\leqslant \int_{x_{2i}}^{x_{2i+2}} \left| \frac{u^{(4)}\big(\xi(x)\big)}{24} (x - x_{2i})(x - x_{2i+1})(x - x_{2i+2})(x - \alpha) \right| \mathrm{d}\, x \\ &\leqslant \frac{C_4}{24} \int_{x_{2i}}^{x_{2i+2}} \big|(x - x_{2i})(x - x_{2i+1})(x - x_{2i+2})(x - \alpha)\big| \,\mathrm{d}\, x.\end{aligned}$$

This inequality is valid for any $\alpha \in A := [x_{2i}, x_{2i+2}] \backslash \{x_{2i}, x_{2i+1}, x_{2i+2}\}$. Denoting by $h(\alpha)$ the integral on the right-hand side, we observe that

$$\lim_{\alpha \to x_{2i+1}} h(\alpha) = \int_{x_{2i}}^{x_{2i+2}} (x - x_{2i})(x - x_{2i+1})^2 (x_{2i+2} - x) \,\mathrm{d}\, x = \frac{4}{15} h^5.$$

Therefore, we conclude that

$$\left| \int_{x_{2i}}^{x_{2i+2}} u(x) - \hat{u}_2(x) \,\mathrm{d}\, x \right| \leqslant \inf_{\alpha \in A} \left| \int_{x_{2i}}^{x_{2i+2}} u(x) - \hat{u}_3(x;\alpha) \,\mathrm{d}\, x \right| \leqslant \frac{C_4}{90} h^5.$$

Summing the contributions of all the subintervals, we finally obtain

$$|I - \hat{I}_h| \leqslant \frac{n}{2} \times \frac{C_4 h^5}{90} = (b - a) \frac{C_4 h^4}{180}, \tag{3.12}$$

which concludes the proof. □

Remark 3.2. The cancellation of the second term in (3.11) also follows from the fact that the degree of precision of the Simpson rule (3.6) is equal to 3, and so

$$\int_{x_{2i}}^{x_{2i+2}} \hat{u}_3(x) - \hat{u}_2(x) \,\mathrm{d}\, x = \frac{1}{3}(\hat{u}_3 - \hat{u}_2)(x_{2i}) + \frac{4}{3}(\hat{u}_3 - \hat{u}_2)(x_{2i+1}) + \frac{1}{3}(\hat{u}_3 - \hat{u}_2)(x_{2i+2}) = 0,$$

where we used the short-hand notation $\hat{u}_3(x) = \hat{u}_3(x;\alpha)$.

General composite quadrature rules. In view of (3.4), any quadrature rule with $N + 1$ integration points of the form

$$\int_{-1}^{1} u(x) \,\mathrm{d}\, x \approx \sum_{j=0}^{N} w_j u(x_j) \tag{3.13}$$

admits a composite version for an arbitrary integration interval $[a, b]$ obtained by applying the rule locally in M equally-sized subintervals:

$$\int_a^b u(x)\,\mathrm{d}\,x \approx \frac{h}{2}\sum_{i=0}^{M-1}\sum_{j=0}^{N} w_j u\left(a + ih + \frac{h}{2} + \frac{x_j h}{2}\right), \qquad h := \frac{b-a}{M}. \tag{3.14}$$

Clearly, the single-interval rule (3.13) and the composite rule (3.14) have the same degree of precision. To conclude this section, we prove a relation between this degree of precision of a rule and the convergence rate with respect to h of the composite rule (3.14).

Theorem 3.3. *Suppose that the degree of precision of the single-interval rule (3.13) is equal to d. Then for any $u \in C^{(d+1)}[a, b]$, there is $c > 0$ such that*

$$\forall M \geqslant 1, \qquad \left|I[u] - \hat{I}_h[u]\right| \leqslant ch^{d+1}, \qquad c := \frac{C_{d+1}(b-a)}{(d+1)!}\left(1 + \frac{1}{2}\sum_{i=0}^{J}|w_j|\right). \tag{3.15}$$

Here $I[u]$ and $\hat{I}_h[u]$ denote respectively the left-hand side and the right-hand side of (3.14), and C_{d+1} is defined in (3.17).

Proof. Fix $i \in \{0, \dots M-1\}$, and let p_i denote the Taylor expansion of degree d of the function u around $z_i := a + ih$:

$$p_i(x) = u(z_i) + u'(z_i)(x - z_i) + \ldots + \frac{u^{(d)}(z_i)}{d!}(x - z_i)^d. \tag{3.16}$$

By the mean-value form of the remainder, it is simple to prove that

$$\forall x \in [z_i, z_{i+1}], \qquad \left|u(x) - p_i(x)\right| = \frac{C_{d+1}}{(d+1)!}h^{d+1}, \qquad C_{d+1} = \sup_{x\in[a,b]}\left|u^{(d+1)}(x)\right|. \tag{3.17}$$

Let us define the local contributions to the total integral as follows:

$$I^{(i)}[u] := \int_{z_i}^{z_{i+1}} u(x)\,\mathrm{d}\,x, \qquad \hat{I}_h^{(i)}[u] = \frac{h}{2}\sum_{i=0}^{N} w_i u\left(z_i + \frac{h}{2} + \frac{x_i h}{2}\right),$$

so that $I[u] = \sum_{i=1}^{M-1} I^{(i)}[u]$ and $\hat{I}_h[u] = \sum_{i=1}^{M-1} \hat{I}_h^{(i)}[u]$. Since p is a polynomial of degree at most d, it follows that

$$\begin{aligned}\left|I^{(i)}[u] - \hat{I}_h^{(i)}[u]\right| &= \left|I^{(i)}[u] - I^{(i)}[p] + \hat{I}_h^{(i)}[p] - \hat{I}_h^{(i)}[u]\right| \\ &\leqslant \left|I^{(i)}[u] - I^{(i)}[p]\right| + \left|\hat{I}_h^{(i)}[u] - \hat{I}_h^{(i)}[p]\right| = \left|I^{(i)}[u-p]\right| + \left|\hat{I}_h^{(i)}[u-p]\right|,\end{aligned}$$

where we used the triangle inequality and linearity in the second line. Thus, using (3.17) we deduce that

$$\left|I^{(i)}[u] - \hat{I}_h^{(i)}[u]\right| \leqslant \frac{C_{d+1}}{(d+1)!} h^{d+1} \int_{z_i}^{z_{i+1}} 1 \,\mathrm{d}\, x + \frac{h}{2} \frac{C_{d+1}}{(d+1)!} h^{d+1} \sum_{i=0}^{J} |w_j|$$
$$= \frac{C_{d+1}}{(d+1)!} h^{d+2} \left(1 + \frac{1}{2} \sum_{i=0}^{J} |w_j|\right).$$

Summing the contributions of the local errors, we obtain

$$\left|I[u] - \hat{I}_h[u]\right| \leqslant \sum_{i=0}^{M-1} \left|I^{(i)}[u] - \hat{I}_h^{(i)}[u]\right| \leqslant M \frac{C_{d+1}}{(d+1)!} h^{d+2} \left(1 + \frac{1}{2} \sum_{i=0}^{J} |w_j|\right),$$

which leads to the result since $Mh = b - a$. □

Remark 3.3. A few comments are in order.

- The constant c in (3.15) is not sharp in general, but the power of h is optimal.
- This result may be viewed as a generalization of Theorems 3.1 and 3.2, albeit with a worse constant prefactor.
- Instead of the polynomial p_i in (3.16), we could have used any polynomial approximation of u such that (3.17) is satisfied, with possibly a different prefactor but the same power of h on the right-hand side. A natural choice, for example, would have been to define p_i by interpolation through (possibly a subset or superset of) the local integration points.
- Theorem 3.3 further motivates why the degree of precision of an integration rule is an interesting metric.

Estimating the error a posteriori. In practice, it is useful to be able to estimate the integration error so that, if the error is deemed too large, a better approximation of the integral can be calculated by using a smaller value for the step size h. Calculating the exact error $I - \hat{I}_h$ is impossible in general, because this would require to know the exact value of the integral, but it is possible to calculate a rough approximation of the error based on two numerical approximations of the integral, as we illustrate formally hereafter for the composite Simpson rule.

Suppose that $\hat{I}_{2h}$ and $\hat{I}_h$ are two approximations of the integral, calculated using the composite Simpson rule with step size $2h$ and h, respectively. If we assume that the error proportionally to $\mathcal{O}(h^4)$ as (3.12) suggests, then it holds approximately that

$$I - \hat{I}_h \approx \frac{1}{2^4}(I - \hat{I}_{2h}). \tag{3.18}$$

This implies that

$$I - \hat{I}_{2h} = (I - \hat{I}_h) + (I_h - \hat{I}_{2h}) \approx \frac{1}{16}(I - \hat{I}_{2h}) + (\hat{I}_h - \hat{I}_{2h}).$$

Rearranging this equation gives an approximation of the error for $\hat{I}_{2h}$:

$$I - \hat{I}_{2h} \approx \frac{16}{15}(\hat{I}_h - \hat{I}_{2h}).$$

Using (3.18), we can then derive an error estimate for $\hat{I}_h$:

$$|I - \hat{I}_h| \approx \frac{1}{15}|\hat{I}_h - \hat{I}_{2h}|. \tag{3.19}$$

The right-hand side can be calculated numerically, because it does not depend on the exact value of the integral. In practice, the two sides of (3.19) are often very close for small h. In the code example below, we approximate the integral

$$I = \int_0^{\frac{\pi}{2}} \cos(x)\,\mathrm{d}\,x = 1 \tag{3.20}$$

for different step sizes and compare the exact error with the approximate error obtained using (3.19). The results obtained are summarized in Table 3.1, which shows a good match between the two quantities.

Table 3.1: Comparison between the exact integration error and the approximate integration error calculated using (3.19).

h	Exact error $\lvert I - \hat{I}_h\rvert$	Approximate error $\frac{1}{15}\lvert\hat{I}_h - \hat{I}_{2h}\rvert$
2^{-4}	$5.166847063531321 \times 10^{-7}$	$5.185892840930961 \times 10^{-7}$
2^{-5}	$3.226500089326123 \times 10^{-8}$	$3.229464703065806 \times 10^{-8}$
2^{-6}	$2.0161285974040766 \times 10^{-9}$	$2.016591486390477 \times 10^{-9}$
2^{-7}	$1.2600120946615334 \times 10^{-10}$	$1.260084925291949 \times 10^{-10}$

```
# Composite Simpson's rule
function composite_simpson(u, a, b, n)
    # Integration nodes
    x = LinRange(a, b, n + 1)
    # Evaluation of u at the nodes
    ux = u.(x)
    # Step size
    h = x[2] - x[1]
    # Approximation of the integral
    return (h/3) * sum([ux[1]; ux[end]; 4ux[2:2:end-1]; 2ux[3:2:end-2]])
end
# Function to integrate
u(x) = cos(x)
# Integration bounds
a, b = 0, π/2
# Exact integral
I = 1.0
```

```
# Number of subintervals
ns = [8; 16; 32; 64; 128]
# Approximate integrals
Ī = composite_simpson.(u, a, b, ns)
# Calculate exact and approximate errors
for i in 2:length(ns)
    println("Exact error: $(I - Ī[i]), ",
            "Approx error: $((Ī[i] - Ī[i-1])/15)")
end
```

3.3 Richardson extrapolation and Romberg's method

In the previous section, we showed how the integration error could be approximated based on two approximations of the integral with different step sizes. The aim of this section is to show that, by cleverly combining two approximations $\hat{I}_h$ and $\hat{I}_{2h}$ of an integral, an approximation even better than $\hat{I}_h$ can be constructed.

This approach is based on *Richardson's extrapolation*, which is a general method for accelerating the convergence of sequences, with applications beyond numerical integration. The idea is the following: assume that $J(h)$ is an approximation with step size h of some unknown quantity $J_* = \lim_{h\to 0} J(h)$, and that we have access to evaluations of J at $h, h/2, h/4, h/8 \dots$. If J extends to a smooth function over $[0, H]$, then by Taylor expansion it holds that

$$J(\eta) = J(0) + J'(0)\eta + J''(0)\frac{\eta^2}{2} + J^{(3)}(0)\frac{\eta^3}{3!} + \cdots + J^{(k)}(0)\frac{\eta^k}{k!} + \mathcal{O}(\eta^{k+1}).$$

Elimination of the linear error term. Let us assume that $J'(0) \neq 0$, so that the leading order term after the constant $J(0)$ scales as η. Then we have

$$\begin{aligned} J(h) &= J(0) + J'(0)h + \mathcal{O}(h^2) \\ J(h/2) &= J(0) + J'(0)\frac{h}{2} + \mathcal{O}(h^2). \end{aligned}$$

We now ask the following question: can we combine linearly $J(h)$ and $J(h/2)$ in order to construct an approximation $J_1(h/2)$ of $J(0)$ with an error scaling as $\mathcal{O}(h^2)$? Employing the ansatz $J_1(h/2) = \alpha J(h) + \beta J(h/2)$, we calculate

$$J_1(h/2) = (\alpha + \beta)J(0) + J'(0)h\left(\alpha + \frac{1-\alpha}{2}\right) + \mathcal{O}(h^2). \tag{3.21}$$

Since we want this expression to approximate $J(0)$ for small h, we need to impose that $\alpha + \beta = 1$. Then, in order for the term multiplying h to cancel out, we require that

$$\alpha + \frac{1-\alpha}{2} = 0 \qquad \Leftrightarrow \qquad \alpha = -1.$$

This yields the formula

$$J_1(h/2) = 2J(h/2) - J(h). \tag{3.22}$$

Notice that, in the case where J is a linear function, $J_1(h/2)$ is exactly equal to $J(0)$. This reveals a geometric interpretation of (3.22): the approximation $J_1(h/2)$ is simply the y intercept of the straight line passing through the points $\big(h/2, J(h/2)\big)$ and $\big(h, J(h)\big)$.

Elimination of the quadratic error term. If we had tracked the coefficient of h^2 in the previous paragraph, we would have obtained instead of (3.21) the following equation:

$$J_1(h/2) = J(0) - J^{(2)}(0)\frac{h^2}{4} + \mathcal{O}(h^3).$$

Provided that we have access also to $J(h/4)$, we can also calculate

$$J_1(h/4) = 2J(h/4) - J(h/2) = J(0) - J^{(3)}(0)\frac{h^2}{16} + \mathcal{O}(h^3).$$

At this point, it is natural to wonder whether we can combine $J_1(h/2)$ and $J_1(h/4)$ in order to produce an even better approximation of $J(0)$. Applying the same reasoning as in the previous section leads us to introduce

$$J_2(h/4) = \frac{4J_1(h/4) - J_1(h/2)}{4-1} = J(0) + \mathcal{O}(h^3).$$

This is an exact approximation of $J(0)$ if J is a quadratic polynomial, indicating that $J_2(h/4)$ is simply the y intercept of the quadratic polynomial interpolating the function J through the three points $\big(h/4, J(h/4)\big)$, $\big(h/2, J(h/2)\big)$ and $\big(h, J(h)\big)$.

Elimination of higher order terms. The procedure above can be repeated in order to eliminate terms of higher and higher orders. The following schematic illustrates, for example, the calculation of an approximation $J_3(h/8) = J(0) + \mathcal{O}(h^4)$.

$$\begin{array}{ccccccc}
J(h) & & & & & & \\
 & \searrow & & & & & \\
J(h/2) & \longrightarrow & J_1(h/2) & & & & \\
 & \searrow & & \searrow & & & \\
J(h/4) & \longrightarrow & J_1(h/4) & \longrightarrow & J_2(h/4) & & \\
 & \searrow & & \searrow & & \searrow & \\
J(h/8) & \longrightarrow & J_1(h/8) & \longrightarrow & J_2(h/8) & \longrightarrow & J_3(h/8) \\
\mathcal{O}(h) & & \mathcal{O}(h^2) & & \mathcal{O}(h^3) & & \mathcal{O}(h^4).
\end{array}$$

Here, the last row indicates the scaling of the error with respect to the parameter h in the limit as $h \to 0$. The linear combination in order to calculate $J_i(h/2^i)$ is always of the form

$$J_i(h/2^i) = \frac{2^i J_{i-1}(h/2^i) - J_{i-1}(h/2^{i-1})}{2^i - 1}, \qquad J_0 = J.$$

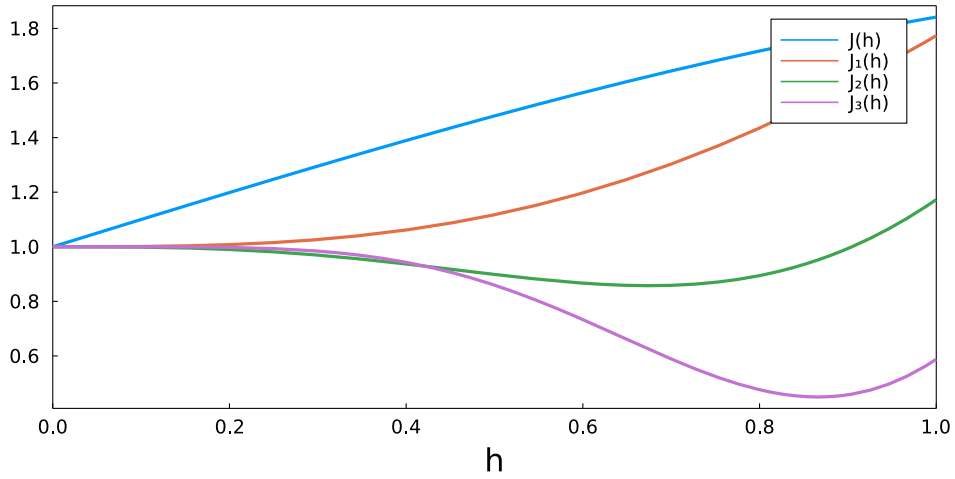

Figure 3.1: Illustration of the functions J_1, J_2 and J_3 constructed by Richardson extrapolation.

In practice we calculate the values taken by $J, J_1, J_2, \ldots$ at specific values of h, but these are in fact functions of h. In Figure 3.1, we plot these functions when $J(h) = 1 + \sin(h)$. It appears clearly from the figure that, for sufficiently small h, $J_3(h)$ provides the most precise approximation of $J(0) = 1$. Constructing the functions in Julia can be achieved in just a few lines of code.

```
J(h) = 1 + sin(h)
J_1(h) = 2J(h) - J(2h)
J_2(h) = (4J_1(h) - J_1(2h))/3
J_3(h) = (8J_2(h) - J_2(2h))/7
```

Generalization. Sometimes, it is known a priori that the Taylor development of the function J around zero contains only even powers of h. In this case, the Richardson extrapolation procedure can be slightly modified in order to produce approximations with errors scaling as $\mathcal{O}(h^4)$, then $\mathcal{O}(h^6)$, then $\mathcal{O}(h^8)$, etc. This procedure is illustrated below:

$$
\begin{array}{ccccccc}
J(h) & & & & & & \\
 & \searrow & & & & & \\
J(h/2) & \longrightarrow & J_1(h/2) & & & & \\
 & \searrow & & \searrow & & & \\
J(h/4) & \longrightarrow & J_1(h/4) & \longrightarrow & J_2(h/4) & & \\
 & \searrow & & \searrow & & \searrow & \\
J(h/8) & \longrightarrow & J_1(h/8) & \longrightarrow & J_2(h/8) & \longrightarrow & J_3(h/8) \\
\mathcal{O}(h^2) & & \mathcal{O}(h^4) & & \mathcal{O}(h^6) & & \mathcal{O}(h^8).
\end{array}
$$

This time, the linear combinations required for populating this table are given by

$$
J_i(h/2^i) = \frac{2^{2i} J_{i-1}(h/2^i) - J_{i-1}(h/2^{i-1})}{2^{2i} - 1}. \tag{3.23}
$$

Application to integration: Romberg's method Romberg's integration method consists of applying Richardson's extrapolation to the function

$$J(h) = \hat{I}_h = u(x_0) + 2u(x_1) + 2u(x_2) + \cdots + 2u(x_{n-1}) + 2u(x_n), \qquad h \in \left\{ \frac{b-a}{n} : n \in \mathbf{N} \right\}.$$

where $a = x_0 < x_1 < \cdots < x_n = b$ are equidistant nodes. The right-hand side of this equation is simply the composite trapezoidal rule with step size h. It is possible to show that $J(h)$ may be expanded as follows:

$$\forall k \in \mathbf{N}, \qquad J(h) = I + \alpha_1 h^2 + \alpha_2 h^4 + \cdots + \alpha_k h^{2k} + \mathcal{O}(h^{2k+2}). \tag{3.24}$$

This is the content of the following result.

Lemma 3.4. *Let $J(h)$ denote the approximation of the integral $I[u]$ by the composite trapezium rule, and assume that $u \in C^\infty[a,b]$. Then $J(h)$ may be expanded as in (3.24).*

Proof. Using the same notation as in Theorem 3.3, we introduce

$$I^{(i)}[u] = \int_{x_i}^{x_{i+1}} u(x)\,\mathrm{d}\,x, \qquad \hat{I}_h^{(i)}[u] = \frac{h}{2}\Big(u(x_i) + u(x_{i+1})\Big), \qquad i = 0, \ldots, n-1.$$

Fix $i \in \{0, \ldots, n-1\}$ and let p_i denote the Taylor expansion of u around $x_{i+\frac{1}{2}} := x_i + \frac{h}{2}$, of degree $2k+1$:

$$p_i(x) = u\Big(x_{i+\frac{1}{2}}\Big) + u'\Big(x_{i+\frac{1}{2}}\Big)\Big(x - x_{i+\frac{1}{2}}\Big) + \cdots + \frac{u^{(2k+1)}\Big(x_{i+\frac{1}{2}}\Big)}{(2k+1)!}\Big(x - x_{i+\frac{1}{2}}\Big)^{2k+1}.$$

From the mean-value form of the remainder, it is clear that

$$I^{(i)}[u] = I^{(i)}[p_i] + \mathcal{O}(h^{2k+3}), \qquad \hat{I}_h^{(i)}[u] = \hat{I}_h^{(i)}[p_i] + \mathcal{O}(h^{2k+3}).$$

Substituting p_i and noting that odd powers cancel when integrating, we have that

$$\begin{aligned} I^{(i)}[p_i] &= u\Big(x_{i+\frac{1}{2}}\Big)h + \omega_2 h^3 + \omega_4 h^5 + \cdots + \omega_{2k} h^{2k+1}, \\ \hat{I}_h^{(i)}[p_i] &= u\Big(x_{i+\frac{1}{2}}\Big)h + \eta_2 h^3 + \eta_4 h^5 + \cdots + \eta_{2k} h^{2k+1}, \end{aligned}$$

for appropriate coefficients. Thus we obtain that

$$\begin{aligned} I[u] - \hat{I}_h[u] &= \sum_{i=0}^{n} I^{(i)}[p_i] - I_h^{(i)}[p_i] \\ &= (\omega_2 - \eta_2)h^2 + (\omega_2 - \eta_2)h^4 + \ldots + (\omega_{2k} - \eta_{2k})h^{2k} + \mathcal{O}(h^{2k+2}), \end{aligned}$$

which concludes the proof. □

Richardson's extrapolation (3.23) can therefore be employed in order to compute approximations of the integral with increasing accuracy. The convergence of Romberg's method for calculating the integral (3.20) is illustrated in Figure 3.2.

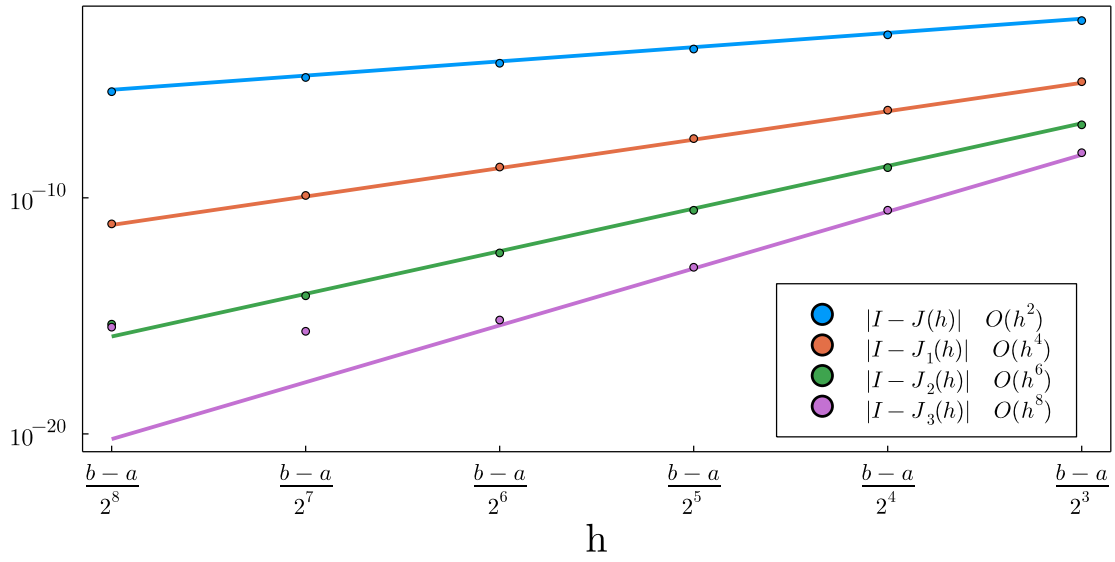

Figure 3.2: Convergence of Romberg's method. The straight lines correspond to the monomial functions $f(h) = C_i h^i$, with $i = 2, 4, 6, 8$ and for appropriate constants C_i. We observe a good agreement between the observed and theoretical convergence rates.

3.4 Methods with non-equidistant nodes

The Newton–Cotes method relies on equidistant integration nodes, and the only degrees of freedom are the integration weights. If the nodes are not fixed, then additional degrees of freedom are available, and these can be leveraged in order to construct a better integration formula. The total number of degrees of freedom for a general integration rule of the form (3.2) is $2n + 2$ which, in principle, should enable to construct an integration rule with a degree of precision equal to $2n + 1$.

A necessary condition for an integration rule of the form (3.2) to have a degree of precision equal to $2n + 1$ is that it integrates exactly all the monomials of degree 0 to $2n + 1$. This condition is also sufficient because, assuming that it is satisfied, we have by linearity of the functionals I and $\hat{I}$ that

$$\begin{aligned}\hat{I}(\alpha_0 + \alpha_1 x + \cdots + \alpha_{2n+1} x^{2n+1}) &= \alpha_0 \hat{I}(1) + \alpha_1 \hat{I}(x) + \cdots + \alpha_{2n+1} \hat{I}(x^{2n+1}) \\ &= \alpha_0 I(1) + \alpha_1 I(x) + \cdots + \alpha_{2n+1} I(x^{2n+1}) \\ &= I(\alpha_0 + \alpha_1 x + \cdots + \alpha_{2n+1} x^{2n+1}),\end{aligned}$$

Here $I(u)$ and $\hat{I}(u)$ denote respectively the exact integral of u and its approximate integral using (3.2). In order to find the nodes and weights of the integration rule, we can therefore solve the following nonlinear system of $2n + 2$ equations with $2n + 2$ unknowns:

$$\sum_{i=0}^{n} w_i x_i^d = \int_{-1}^{1} x^d \,\mathrm{d}\, x, \qquad d = 0, \ldots, 2n+1. \tag{3.25}$$

The quadrature rule thus obtained is called the *Gauss–Legendre quadrature*.

Example 3.2. Let us derive the Gauss–Legendre quadrature with $n+1 = 2$ nodes. The system of equations that we need to solve in this case is the following:

$$w_0 + w_1 = 2, \qquad w_0 x_0 + w_1 x_1 = 0, \qquad w_0 x_0^2 + w_1 x_1^2 = \frac{2}{3}, \qquad w_0 x_0^3 + w_1 x_1^3 = 0.$$

The solution to these equations is given by

$$-x_0 = x_1 = \frac{\sqrt{3}}{3}, \qquad w_0 = w_1 = 1.$$

Connection with orthogonal polynomials. Let $(L_n)_{n\in\mathbf{N}}$ denote the Legendre polynomials, i.e. the orthogonal polynomials with for the inner product

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x) \,\mathrm{d}\, x.$$

The nodes and weights of the Gauss–Legendre quadrature rules can be obtained constructively from Legendre polynomials, as shown in Section 2.2.4. We shall now demonstrate this connection in much more direct manner. Specifically, we prove that the integration nodes are given by the roots of a Legendre polynomial.

Theorem 3.5. *For every $n \in \mathbf{N}$, there exists a unique solution to the system of equations (3.25). The nodes $(x_i)_{i\in\{0,\ldots,n\}}$ are the roots of L_{n+1} and the weights are given by*

$$w_i = \int_{-1}^{1} \ell_i(x) \,\mathrm{d}\, x, \qquad i \in \{1, \ldots, n\}, \tag{3.26}$$

where ℓ_i is the Lagrange polynomial

$$\frac{(x-x_0)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}.$$

In addition, the weights are all positive.

Proof. We show first that the solution to (3.25) exists, then that it is unique, and finally that the weights are positive.

Existence. We begin by showing that, if $(x_i)_{i\in\{0,\ldots n\}}$ are the roots of L_{n+1} and the weights are defined from (3.26), then the equations (3.25) are satisfied. To this end, it is sufficient to show that for all $p \in \mathbf{P}(2n+1)$,

$$\int_{-1}^{1} p(x)\,\mathrm{d}\,x = \sum_{i=0}^{n} w_i p(x_i). \tag{3.27}$$

Take $p \in \mathbf{P}(2n+1)$ and let $q \in \mathbf{P}(n)$ and $r \in \mathbf{P}(n)$ be the polynomials such that

$$p(x) = q(x)L_{n+1}(x) + r(x).$$

The quotient q and the remainder r can be obtained by Euclidean division of p by L_{n+1}. Since L_{n+1} is orthogonal to any polynomial in $\mathbf{P}(n)$, in particular q, and the nodes $(x_i)_{i\in\{0,\dots n\}}$ are the roots of L_{n+1}, it holds that

$$\int_{-1}^{1} p(x)\,\mathrm{d}\,x - \sum_{i=0}^{n} w_i p(x_i) = \int_{-1}^{1} r(x)\,\mathrm{d}\,x - \sum_{i=0}^{n} w_i r(x_i). \tag{3.28}$$

Given that $r \in \mathbf{P}(n)$, the remainder r must coincides with its polynomial interpolation at the points $x_0, \dots x_m$. Therefore,

$$r = r(x_0)\ell_0 + \cdots + r(x_n)\ell_n,$$

and so

$$\int_{-1}^{1} r(x)\,\mathrm{d}\,x = \sum_{i=0}^{n} r(x_i) \int_{-1}^{1} \ell_i(x)\,\mathrm{d}\,x = \sum_{i=0}^{n} r(x_i) w_i,$$

where we used (3.26) in the last equality. Consequently, the right-hand side of (3.28) is zero, and since p was arbitrary this implies that (3.27) is satisfied.

Uniqueness. Next, we show that the nodes are necessarily the roots of L_{n+1}. To this end, assume that $x_0, \dots, x_n$ and weights $w_0, \dots, w_n$ are such that the equations (3.25) are satisfied, and let

$$q(x) = (x - x_0)\dots(x - x_n).$$

Our goal is to show that $q(x)$ coincides with L_{n+1} up to a constant factor. In order to show this, it is sufficient to prove that $q(x)$ is orthogonal to x^d for all $d = 0, \dots, n$, because the only polynomial in $\mathbf{P}(n+1)$ that satisfies these orthogonality relations is the Legendre polynomial L_{n+1}, or a multiple thereof. For the values of d considered, the polynomial $q(x)x^d$ belongs to $\mathbf{P}(2n+1)$. Given that the integration rule with nodes $x_0, \dots, x_n$ and weights $w_0, \dots, w_n$ is exact for all the elements of $\mathbf{P}(2n+1)$ by assumption, we deduce that

$$\forall d \in \{0, \dots, n\}, \qquad \int_{-1}^{1} q(x)x^d\,\mathrm{d}\,x = \sum_{i=0}^{n} w_i q(x_i) x_i^d = 0.$$

Finally, we show that the weights are necessarily given by (3.26). Since the integration rule must be exact for any Lagrange polynomial ℓ_j, we have

$$\int_{-1}^{1} \ell_j(x) \,\mathrm{d}\, x = \sum_{i=0}^{n} w_i \ell_j(x_i) = w_j,$$

which concludes the proof of uniqueness.

Positivity of the weights. Since the integration rule is exact for all the polynomials in $\mathbf{P}(2n+1)$ and $\ell_j(x)^2 \in \mathbf{P}(2n+1)$, we deduce that

$$\int_{-1}^{-1} \left|\ell_j(x)\right|^2 \,\mathrm{d}\, x = \sum_{i=0}^{n} w_i \left|\ell_j(x_i)\right|^2 = w_j.$$

The left-hand side is positive, and so w_j must also be positive.

□

Since the integration weights are all positive, the Gauss–Legendre quadrature rules are less susceptible to roundoff errors than the Newton–Cotes methods. In addition, we have the following result.

Theorem 3.6. *Assume that $u \in C([-1,1])$, and let $\hat{I}_n(u)$ denote the approximation of $I(u)$ using the Gauss–Legendre quadrature. Then $\hat{I}_n(u) \to I(u)$ in the limit as $n \to \infty$.*

Proof. The positivity of the weights is crucial for the proof. By the Weierstrass approximation theorem, for all $\varepsilon > 0$ there exists polynomial p such that

$$E \coloneqq \max_{x \in [-1,1]} \left|p(x) - u(x)\right| \leqslant \varepsilon.$$

Since the degree of precision of the Gauss–Legendre quadrature with $n+1$ integration nodes is $2n+1$, there exists N sufficiently large such that $\hat{I}_n(p) = I(p)$ for all $n \geqslant N$. Thus

$$\begin{aligned} \forall n \geqslant N, \qquad \left|\hat{I}_n(u) - I(u)\right| &\leqslant \left|\hat{I}_n(u) - \hat{I}_n(p)\right| + \left|\hat{I}_n(p) - I(p)\right| + \left|I(p) - I(u)\right| \\ &= \left|\hat{I}_n(u) - \hat{I}_n(p)\right| + \left|I(p) - I(u)\right| \\ &\leqslant \sum_{i=0}^{n} |w_i| \; E + \int_{-1}^{1} E \,\mathrm{d}\, x = 2\varepsilon + 2\varepsilon = 4\varepsilon. \end{aligned}$$

Indeed, all the weights are positive and the Gauss–Legendre quadrature rule is exact for the constant function $u(x) = 1$, which implies that the weights add up to 2. Since $\varepsilon > 0$ was arbitrary, this concludes the proof. □

Computation of the integration nodes. We proved that the integration nodes are given by the roots of the Legendre polynomials. In practice, calculating these roots can be achieved by calculating the eigenvalues of a tridiagonal matrix, see Section 2.2.4. This is known as the Golub–Welsch algorithm.

Generalization to higher dimensions. Gauss–Legendre integration is ubiquitous in numerical methods for partial differential equations, in particular the *finite element method*. Its generalization to higher dimensions is immediate: for a function u: $[-1,1] \times [-1,1] \to \mathbf{R}$, we have

$$\int_0^1 \int_0^1 u(x,y) \,\mathrm{d}\, y \,\mathrm{d}\, x \approx \sum_{i=0}^{n} \sum_{j=0}^{n} w_i w_j u(x_i, y_i).$$

The degree of precision of this integration rule is the same as that of the corresponding one-dimensional rule, and this approach can be generalized to any dimension d. The associated computational cost, however, scales as n^d, and so it is not a good idea to use a deterministic method of this type in the high dimensional setting. The explosion of the computational cost as the dimension increases is known as *the curse of dimensionality*.

3.5 Introduction to probabilistic integration methods

So far in this chapter, we covered only *deterministic* integration formulas. Much of the research around the calculation of high-dimensional integrals today is concerned with *probabilistic* integration methods using probabilistic approaches. These methods are based on the connection between integrals and expectations. To illustrate the simplest probabilistic approach, called the *Monte Carlo method*, we consider the problem of approximating the integral

$$I = \int_0^1 u(x) \,\mathrm{d}\, x.$$

This integral may be expressed as the expectation $\mathbf{E}\left[u(X)\right]$, where $\mathbf{E}$ is the expectation operator and $X \sim \mathcal{U}(0,1)$ is a uniformly distributed random variable over the interval $[0,1]$. Therefore, in practice, the integral I may be approximated by generating a large number of *independent* samples $X_1, X_2, \ldots$ drawn from the distribution $\mathcal{U}(0,1)$ and averaging $f(X_i)$ over all of these samples:

$$\hat{I}_N = \frac{1}{N} \sum_{n=1}^{N} u(X_n).$$

The quantity $\hat{I}_N$, where the subscript N denotes the number of samples employed, is itself a random variable; it is called an *estimator* of the exact integral I. In Julia, the calculation of the integral using the Monte Carlo method can be achieved with the following code.

```
N = 1000
u(x) = x^2
X = rand(N)
Î = (1/N) * sum(u.(X))
```

Since the expectation operator is linear, we calculate that

$$\mathbf{E}\,[\hat{I}_N] = \frac{1}{N}\sum_{n=1}^{N}\mathbf{E}\,\Big[u(X_n)\Big] = \frac{1}{N}\sum_{n=1}^{N}\int_0^1 u(x)\,\mathrm{d}\,x = \frac{1}{N}\sum_{n=1}^{N} I = I.$$

The estimator $\hat{I}_N$ is therefore said to be *unbiased*, because the bias $\mathbf{E}\,[\hat{I}_N] - I$ is zero. Now assume that the function u is square integrable over the interval $[0,1]$ and let

$$\sigma^2 := \mathbf{V}\,\Big[u(X)\Big] := \mathbf{E}\,\left[\Big(u(X) - \mathbf{E}\,\Big[u(X)\Big]\Big)^2\right] = \int_0^1 |u(x) - I|^2\,\mathrm{d}\,x,$$

where $\mathbf{V}\,[X]$ denotes the variance of the random variable X. We can calculate explicitly the variance of the estimator $\hat{I}_N$:

$$\mathbf{V}\,[\hat{I}_N] = \mathbf{E}\,\left[\left(\frac{1}{N}\sum_{n=1}^{N} u(X_n) - I\right)^2\right] = \mathbf{E}\,\left[\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\Big(u(X_n) - I\Big)\Big(u(X_m) - I\Big)\right].$$

Since the samples $X_1, X_2, \ldots$ are independent, it holds that

$$\mathbf{E}\,\Big[\Big(u(X_n) - I\Big)\Big(u(X_m) - I\Big)\Big] = \delta_{nm}\sigma^2,$$

where δ_{mn} is the Kronecker delta. Consequently, we deduce that

$$\mathbf{V}\,[\hat{I}_N] = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\mathbf{E}\,\Big[\Big(u(X_n) - I\Big)\Big(u(X_m) - I\Big)\Big] = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\delta_{mn}\sigma^2 = \frac{\sigma^2}{N}.$$

Therefore, the variance of $\hat{I}_N$ decreases as $1/N$ when the number of samples N increases, indicating that the estimator becomes increasingly accurate. To state this more precisely, we will use Chebyshev's inequality.

Theorem 3.7 (Chebyshev's inequality)**.** *Let Z be a random variable with mean m and variance s^2. Then for any real $k > 0$,*

$$\mathbf{P}\,\Big[\big|Z - m\big| \geqslant ks\Big] \leqslant \frac{1}{k^2}.$$

Let $\varepsilon > 0$. Employing Chebyshev's inequality with $k = \varepsilon/\sqrt{\mathbf{V}\,[\hat{I}_N]}$, we obtain

$$\mathbf{P}\,\Big[\big|\hat{I}_N - I\big| \geqslant \varepsilon\Big] \leqslant \frac{\mathbf{V}\,[\hat{I}_N]}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2}. \tag{3.29}$$

Consequently, for any $\varepsilon > 0$, the probability that the integration error $\big|\hat{I}_N - I\big|$ is greater than or equal to ε tends to zero in the limit as $N \to \infty$. In probabilistic jargon, we say that $\hat{I}_N$ converges *in probability* to I. Equation (3.29) can also be employed in order to construct a *confidence interval* for the exact integral, as we demonstrate in the following paragraph.

Construction of a $(1-\alpha)$ confidence interval. By definition, a $(1-\alpha)$ confidence interval for I is an interval $[z_1, z_2]$, the endpoints of which being random variables, such that the

probability that $I \in [z_1, z_2]$ is greater than or equal to $1 - \alpha$. To construct such an interval, we begin by finding ε such that the right-hand side of (3.29) is equal to α, which gives $\varepsilon = \sqrt{\sigma^2/N\alpha}$. For this value of ε, we have

$$\mathbf{P}\left[\left[\hat{I}_N - I\right| \geqslant \varepsilon\right] \leqslant \alpha.$$

Since $|\hat{I}_N - I| \geqslant \varepsilon$ if and only if $I \notin (\hat{I}_N - \varepsilon, \hat{I}_N + \varepsilon)$, we conclude that

$$\mathbf{P}\left[I \in \left(\hat{I}_N - \varepsilon, \hat{I}_N + \varepsilon\right)\right] \geqslant 1 - \alpha.$$

We have thus shown that

$$\left(\hat{I}_N - \sqrt{\frac{\sigma^2}{N\alpha}}, \hat{I}_N + \sqrt{\frac{\sigma^2}{N\alpha}}\right)$$

is a $(1 - \alpha)$ confidence interval for I.

3.6 Exercises

Exercise 3.1. *Derive the Simpson's integration rule (3.6).*

Exercise 3.2. *Derive the composite Simpson integration rule (3.9).*

Exercise 3.3. *Consider the integration rule*

$$\int_0^1 u(x)\,\mathrm{d}\,x \approx w_1 u(0) + w_2 u(1) + w_3 u'(0).$$

Find w_1, w_2 and w_3 so that this integration rule has the highest possible degree of precision.

Exercise 3.4. *Consider the integration rule*

$$\int_{-1}^1 u(x)\,\mathrm{d}\,x \approx w_1 u(x_1) + w_2 u'(x_1).$$

Find w_1, w_2 and x_1 so that this integration rule has the highest possible degree of precision.

Exercise 3.5. *What is the degree of precision of the following quadrature rule?*

$$\int_{-1}^1 u(x)\,\mathrm{d}\,x \approx \frac{2}{3}\left(2u\left(-\frac{1}{2}\right) - u(0) + 2u\left(\frac{1}{2}\right)\right).$$

Exercise 3.6. *The Gauss–Hermite quadrature rule with $n + 1$ nodes is an approximation of the form*

$$\int_{-\infty}^{\infty} u(x)\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}\,x \approx \sum_{i=0}^{n} w_i u(x_i),$$

such that the rule is exact for all polynomials of degree less than or equal to $2n+1$*. Find the Gauss–Hermite rule with two nodes.*

Exercise 3.7. *Use Romberg's method to construct an integration rule with an error term scaling as* $\mathcal{O}(h^4)$*. Is there a link between the method you obtained and another integration rule seen in class?*

Exercise 3.8 (Improving the error bound for the composite trapezoidal rule)**.** *The notation used in this exercise is the same as in Section 3.2. In particular,* $\hat{I}_h$ *denotes the approximate integral obtained by using the composite trapezoidal rule (3.7), and* $\hat{u}_h$ *is the corresponding piecewise linear interpolant.*

A version of the mean value theorem states that, if $g\colon [a,b] \to \mathbf{R}$ *is a non-negative integrable function and* $f\colon [a,b] \to \mathbf{R}$ *is continuous, then there exists* $\xi \in (a,b)$ *such that*

$$\int_a^b f(x)g(x)\,\mathrm{d}\,x = f(\xi)\int_a^b g(x)\,\mathrm{d}\,x. \tag{3.30}$$

- *Using (3.30), show that, for all* $i \in \{0, \dots, n-1\}$*, there exists* $\xi_i \in (x_i, x_{i+1})$ *such that*

$$\int_{x_i}^{x_{i+1}} u(x) - \hat{u}_h(x)\,\mathrm{d}\,x = -u''(\xi_i)\frac{h^3}{12}.$$

- *Prove, by using the intermediate value theorem, that if* $f\colon [a,b] \to \mathbf{R}$ *is a continuous function, then for any set* $\xi_0, \dots, \xi_{n-1}$ *of points within the interval* (a,b)*, there exists* $c \in (a,b)$ *such that*

$$\frac{1}{n}\sum_{i=0}^{n-1} f(\xi_i) = f(c).$$

- *Combining the previous items, conclude that there exists* $\xi \in (a,b)$ *such that*

$$I - \hat{I}_h = -u''(\xi)(b-a)\frac{h^2}{12},$$

which is a more precise expression of the error than that obtained in (3.8).

Remark 3.4. One may convince oneself of (3.30) by rewriting this equation as

$$\frac{\int_a^b f(x)g(x)\,\mathrm{d}\,x}{\int_a^b g(x)\,\mathrm{d}\,x} = f(\xi).$$

The left-hand side is the average of $f(x)$ with respect to the probability measure with density given by

$$x \mapsto \frac{g(x)}{\int_a^b g(x)\,\mathrm{d}\,x}.$$

Exercise 3.9 (From the final exam of Spring 2022)**.** *Construct an integration rule of the form*

$$\int_{-1}^{1} u(x)\,\mathrm{d}\,x \approx w_1 u\Big(-\frac{1}{2}\Big) + w_2 u(0) + w_3 u\Big(\frac{1}{2}\Big)$$

with a degree of precision equal to at least 2. What is the degree of precision of the rule constructed?

Solution. The Lagrange polynomials associated with $-1/2$, 0 and $1/2$ are respectively

$$\begin{aligned} p_1(x) &= 2x\Big(x - \frac{1}{2}\Big), \\ p_2(x) &= -4\Big(x + \frac{1}{2}\Big)\Big(x - \frac{1}{2}\Big), \\ p_3(x) &= 2\Big(x + \frac{1}{2}\Big)x. \end{aligned}$$

We deduce that

$$w_1 = \int_{-1}^{1} p_1(x) = \frac{4}{3}, \qquad w_2 = \int_{-1}^{1} p_2(x) = -\frac{2}{3}, \qquad w_3 = \int_{-1}^{1} p_3(x) = \frac{4}{3}.$$

By construction, the degree of precision is at least 2. However, the integration rule is exact also when $u(x) = x^3$. Since the rule is not exact for $u(x) = x^4$, we conclude that the degree of precision is 3. △

Exercise 3.10 (From the final exam of Spring 2022)**.** *The Gauss–Laguerre quadrature rule with n nodes is an approximation of the form*

$$\int_0^{\infty} u(x)\,\mathrm{e}^{-x}\,\mathrm{d}\,x \approx \sum_{i=1}^{n} w_i u(x_i),$$

such that the rule is exact when u is a polynomial of degree less than or equal to $2n-1$.

- *Find the Gauss–Laguerre rule with one node ($n = 1$).*
- *Find the Gauss–Laguerre quadrature rule with two nodes ($n = 2$). You may find it useful to first calculate the Laguerre polynomial of degree 2.*

Solution. Below are the derivations of the Gauss–Laguerre rules with 1 and 2 nodes.

Gauss–Laguerre rule with 1 node. We are looking for w_1 and x_1 such that

$$\forall (a, b) \in \mathbf{R}^2, \qquad \int_0^{\infty} (a + bx)\,\mathrm{e}^{-x}\,\mathrm{d}\,x = w_1(a + bx_1).$$

The left-hand side is equal to

$$a\int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}\,x + b\int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}\,x = a + b\int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}\,x.$$

Using integration by parts, we find the value of the remaining integral on the right-hand side:

$$\begin{aligned}\int_0^\infty x\mathrm{e}^{-x} &= \int_0^\infty -(x\mathrm{e}^{-x})' + \mathrm{e}^{-x}\,\mathrm{d}\,x \\ &= -(x\mathrm{e}^{-x})\Big|_{x=\infty} + (x\mathrm{e}^{-x})\Big|_{x=0} + \int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}\,x \\ &= 0 + 0 + 1.\end{aligned}$$

(To be fully rigorous, we would need to write the first term on the second line as a limit.) Therefore, we obtain

$$a + b = w_1(a + bx_1),$$

which implies that $w_1 = x_1 = 1$.

Gauss–Laguerre rule with 2 nodes. The integration nodes are given by the roots of the Laguerre polynomials, which are the orthogonal polynomials for the inner product

$$\langle f, g\rangle \coloneqq \int_0^\infty f(x)g(x)\,\mathrm{e}^{-x}\,\mathrm{d}\,x.$$

The first polynomial is $\ell_0(x) = 1$. It is simple to check that the only linear monic polynomial orthogonal to ℓ_0 is given by $\ell_1(x) = x - 1$. Next, by integration by parts we calculate that

$$\int_0^\infty x^2\,\mathrm{e}^{-x}\,\mathrm{d}\,x = \int_0^\infty -(x^2\mathrm{e}^{-x})' + 2x\mathrm{e}^{-x}\,\mathrm{d}\,x = 2.$$

and, similarly,

$$\int_0^\infty x^3\,\mathrm{e}^{-x}\,\mathrm{d}\,x = \int_0^\infty -(x^3\mathrm{e}^{-x})' + 3x^2\mathrm{e}^{-x}\,\mathrm{d}\,x = 6.$$

Consider the ansatz $\ell_2(x) = x^2 + a\ell_1(x) + b$. In order for ℓ_2 to be orthogonal to ℓ_0 and ℓ_1, it is necessary that

$$\begin{aligned}0 &= \int_0^\infty \ell_2(x)\,\ell_0(x)\,\mathrm{e}^{-x}\,\mathrm{d}\,x = 2 + b, \\ 0 &= \int_0^\infty \ell_2(x)\,\ell_1(x)\,\mathrm{e}^{-x}\,\mathrm{d}\,x = 4 + a\int_0^\infty \ell_1(x)\ell_1(x)\;\,dx = 4 + a.\end{aligned}$$

Therefore, we conclude that $a = -4$ and $b = -2$, which gives

$$\ell_2(x) = x^2 - 4x + 2.$$

The roots are given by $2 \pm \sqrt{2}$, so we have $x_1 = 2 - \sqrt{2}$ and $x_2 = 2 + \sqrt{2}$. It remains to find the weights. To this end, we need only two additional equations; it is sufficient to require that, for any $(a, b) \in \mathbf{R}^2$,

$$a+b=\int_0^\infty (a+bx)\,\mathrm{e}^{-x}\,\mathrm{d}\,x = w_1(a+bx_1)+w_2(a+bx_2)$$
$$= a(w_1+w_2)+2b(w_1+w_2)+\sqrt{2}b(w_2-w_1).$$

Letting $a=1$ and $b=0$, we obtain $w_1+w_2=1$. Then, letting $a=0$ and $b=1$, we deduce

$$1=2+\sqrt{2}(w_2-w_1) \qquad \Leftrightarrow w_2-w_1=-\frac{\sqrt{2}}{2}.$$

Therefore

$$w_1=\frac{2+\sqrt{2}}{4}, \qquad w_2=\frac{2-\sqrt{2}}{4},$$

which concludes the exercise.

△

Exercise 3.11 (Calculating the volume of hyperballs)**.** *Let B_d denote the d-dimensional unit ball for the Euclidean norm:*

$$B_d=\Big\{\boldsymbol{x}\in\mathbf{R}^d : \|\boldsymbol{x}\|\leqslant 1\Big\}.$$

The volume of B_d is defined as the integral of the characteristic function over B_d:

$$\mathrm{vol}(B_d)=\underbrace{\int_{\mathbf{R}}\dots\int_{\mathbf{R}}}_{d\text{ times}}\chi(\boldsymbol{x})\,\mathrm{d}\,x_1\dots\,\mathrm{d}\,x_d, \qquad \chi(\boldsymbol{x}):=\begin{cases}1 \text{ if } \boldsymbol{x}\in B_d\\ 0 \text{ otherwise.}\end{cases}$$

Complete the following tasks:

- *Write a function*

```
function hyperball_volume(dim, n)
    # Your code comes here
    return vol, σ
end
```

  that calculates the volume of the unit ball B_d with $d=$ `dim` *using a Monte Carlo approach with* `n` *samples drawn from an appropriate distribution. Your function should return an estimation of the volume together with the standard deviation of the estimator (which you should estimate from the samples).*

- *Using the function* `hyperball_volume`*, plot the volumes for d going from 1 to 15, together with a 99% confidence interval. See Figure 3.3 for an example solution with $n=10^7$.*

You are allowed to use your knowledge of the fact that $\mathrm{vol}(B_2)=\pi$ and $\mathrm{vol}(B_3)=4\pi/3$, but do not use the general formula for the volume of B_d.

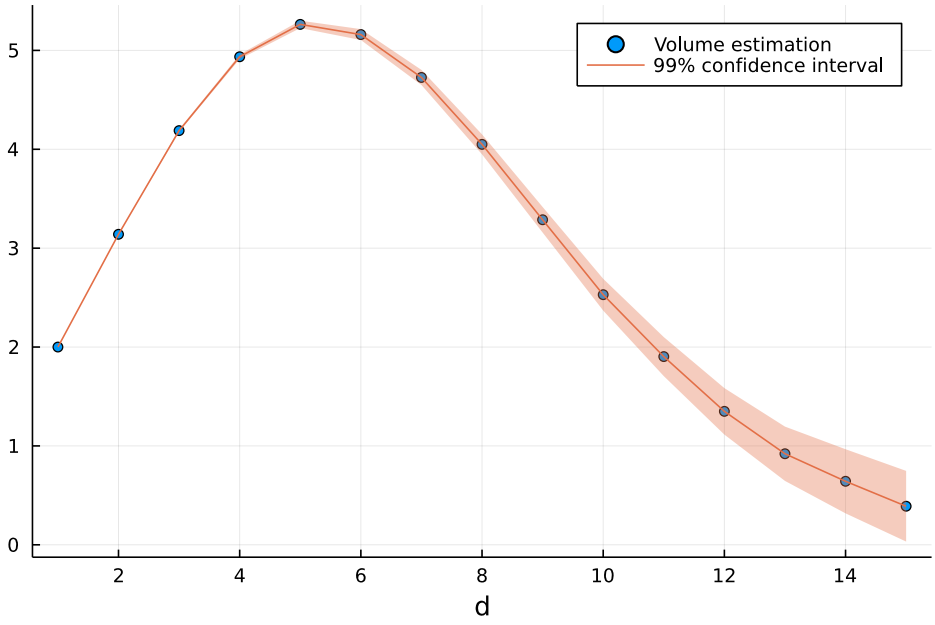

Figure 3.3: Example solution for Exercise 3.11.

Exercise 3.12. *Complete the following tasks:*

- *Write a function* `legendre(n)` *which returns the Legendre polynomial of degree* n*. To this end, you may use the* `Polynomials` *library and Rodrigues' formula:*

$$L_n(x) = \frac{1}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}\, x^n} (x^2 - 1)^n.$$

- *Write a function* `get_nodes_and_weights(n)` *which returns the nodes and weights of the Gauss–Legendre quadrature with* n *nodes. In order to construct Lagrange polynomials, you may find it useful to use the* `fromroots` *functions.*

- *Write a function* `composite_gauss_legendre(u, a, b, n, N)`*, which returns an approximation of the integral*

$$\int_a^b u(x) \,\mathrm{d}\, x$$

obtained by partitioning the integration interval $[a, b]$ *into* N *cells, and applying the Gauss–Legendre quadrature within each cell.*

- *Take* $u(x) = \cos(x)$*,* $a = -1$ *and* $b = 1$*. Illustrate on the same plot the error for the values* $n \in \{1, 2, 3\}$ *and* N *varying from 1 to 40. For each value of* n*, estimate the order of convergence with respect to* N*, i.e. find* $\alpha(n)$ *such that*

$$|\hat{I}_{n,N} - I| \propto C N^{-\alpha},$$

where I *denotes the exact value of the Integral and* $\hat{I}_{n,N}$ *denotes its approximation.*

3.7 Discussion and bibliography

The presentation of part of the material follows that in [7], and some exercises come from [10, Chapter 9]. The main advantage of probabilistic integration approaches is that they generalize naturally to high-dimensional and infinite-dimensional settings.