Numerical Integration

KEY WORDS. Numerical integration, Mid-point rule, Trapezoidal rule, Simpson's rule, degree of precision, superconvergence.

GOAL.

- Understand why numerical methods are necessary for the evaluation of integrals.
- Master basic methods for numerical integration.
- Understand what is the degree of precision and its uses.
- Understand the basics of Gaussian quadrature.

1 Introduction

Consider the definite integral

$$I(f) \equiv \int_a^b f(x)dx,$$

where $f(x)$ is a continuous function on the closed interval $[a, b]$, so that the integral $I(f)$ exists. Approximating $I(f)$ numerically is called *numerical integration* or *quadrature*.

There are a number of reasons for studying numerical integration. Let us recall that the antiderivative of $f(x)$ is a function $F(x)$ such that

$$F'(x) = f(x).$$

From elementary calculus, we know that

$$I(f) = F(x)\big|_a^b = F(b) - F(a).$$

Therefore, if the antiderivative of $f(x)$ is easy to compute and has an elementary representation, then the evaluation of $I(f)$ becomes trivial. But in practice, the antiderivative of f may not be known or may not be elementary. Furthermore, the integral may not be available because the function $f(x)$ may be defined by values in a table or by a subprogram. Or definite integrals must be approximated as part of a more complicated numerical scheme, such as one for the solution of differential equations by finite element methods, an advanced numerical technique for solving partial differential equations frequently used in practice.

As a particular example of where the evaluation of an integral appears, consider the analysis of measurement errors in scientific experiments. Suppose that a surveyor is measuring mountain terrain as part of a highway construction project. The equipment is accurate to the nearest foot (say). What is the probability that a particular measurement overestimates the true value by less than 2 feet? If the measurement errors have a standard normal distribution, that is, they follow a "bell" curve, then (as you have learned in Probability and Statistics) the desired probability is given by the definite integral

$$\frac{1}{\sqrt{2\pi}} \int_0^2 e^{-x^2/2}\, dx.$$

There is no closed form expression for the value of this integral, and it must be estimated using numerical methods.

A basic principle in numerical analysis is that if we cannot do what we want with a given function $f(x)$, we approximate it with a function for which we can. Often the approximating function is an interpolating polynomial. Using this principle, we develop efficient methods for computing approximations to the integral $I(f)$ using only values of the integrand $f(x)$ at points $x \in [a, b]$, and study their errors. When approximating functions, we found that piecewise polynomial interpolants have advantages over polynomial interpolants, and the same is true in this context. In a way, piecewise polynomials are more natural for numerical integration because using such a function amounts to breaking up the interval of integration into pieces and approximating by a polynomial on each piece.

2 Integrals and Rules

To approximate the integral $I(f)$ numerically, we integrate piecewise polynomial approximations of $f(x)$ on the interval $[a, b]$. The very first thing to do in this approximation procedure is to partition the interval $[a, b]$ into n equal subintervals $[x_{i-1}, x_i]$, $i = 1, 2, \cdots, n$, where $x_i = a + ih$, $i = 0, 1, \cdots, n$, and $h = x_i - x_{i-1} \equiv (b - a)/n$, and use the **additive property of integrals** to obtain

$$I(f) = \int_{x_0}^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \cdots + \int_{x_{n-1}}^{x_n} f(x)dx. \tag{1}$$

As a first example, we consider a piecewise constant approximation on each subinterval $[x_{i-1}, x_i]$. On the interval $[x_{i-1}, x_i]$, we approximate $f(x)$ by its value at the midpoint of the interval, so that

$$\int_{x_{i-1}}^{x_i} f(x)dx \approx \int_{x_{i-1}}^{x_i} f(x_{i-1/2})dx = (x_i - x_{i-1})f(x_{i-1/2}) = hf(x_{i-1/2}) \equiv M_1(f),$$

which is called the **midpoint rule**; the quantity $hf(x_{i-1/2})$ is the area of the rectangle of width $h = x_i - x_{i-1}$ and height $f(x_{i-1/2})$. Hence, using the additive property (1), we obtain the **composite midpoint rule**:

$$\begin{aligned} I(f) &= \int_{x_0}^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \cdots + \int_{x_{n-1}}^{x_n} f(x)dx && (2) \\ & && (3) \\ &\approx hf(x_{1/2}) + hf(x_{3/2}) + \cdots + hf(x_{n-1/2}) && (4) \\ &= h\sum_{i=1}^{n} f(x_{i-1/2}) && (5) \\ &\equiv M_n(f). && (6) \end{aligned}$$

3 Quadrature Rules

3.1 A general quadrature rule

Generally, a **quadrature rule** (such as the midpoint rule) has the form

$$R(f) \equiv \sum_{i=0}^{n} w_i f(x_i),$$

for given **nodes** $x_0 < x_1 < \cdots < x_n$ and **weights** $w_0, w_1, \cdots, w_n$.

We need some properties of the definite integral $I(f)$ and of the quadrature rule $R(f)$. First, the integral $I(f)$ is a linear functional[1]; that is,

$$I(\alpha f + \beta g) = \alpha I(f) + \beta I(g), \tag{7}$$

for any constants α and β and any functions f and g for which the integrals $I(f)$ and $I(g)$ exist. In integral notation, this equation provides the standard linearity result for integrals,

$$\int_a^b [\alpha f(x) + \beta g(x)]dx = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx.$$

Similarly, a quadrature rule $R(f)$ is a linear functional, that is,

$$R(\alpha f + \beta g) = \alpha R(f) + \beta R(g),$$

or, in summation notation,

$$\sum_{i=0}^{n} \omega_i[\alpha f(x_i) + \beta g(x_i)] = \alpha \sum_{i=0}^{n} \omega_i f(x_i) + \beta \sum_{i=0}^{n} \omega_i g(x_i).$$

As we have already mentioned, to approximate a definite integral $I(f)$ where we do not know an antiderivative for $f(x)$, a good choice is to integrate a simpler

[1] A functional maps functions to numbers. A definite integral provides a classic example of a functional; this functional assigns to each function a number that is the definite integral of that function.

function $q(x)$ that approximates the function $f(x)$ well and whose antiderivative we do know. From the linearity of the functional $I(f)$, we have

$$I(q) = I(f + [q - f]) = I(f) + I(q - f).$$

Thus the error in approximating the true integral $I(f)$ by the integral of the approximation $I(q)$ can be expressed as

$$I(q) - I(f) = I(q - f);$$

that is, the error in approximating the definite integral of the function $f(x)$ by using the definite integral of the approximating function $q(x)$ is the definite integral of the error in approximating $f(x)$ by $q(x)$. If $q(x)$ approximates $f(x)$ well, that is, if the error $q(x)-f(x)$ is in some sense small, then the error $I(q-f)$ in the integral $I(q)$ approximating $I(f)$ will be small because the integral of a small function is always relatively small.

In the following subsections, particular examples of the general quadrature rule are considered.

3.2 Midpoint rule (Revisited)

Consider the definite integral

$$I(f) = \int_a^b f(x)dx.$$

Interpolation can be used to determine polynomials $q(x)$ that approximate the function $f(x)$ on $[a, b]$. As we have seen, the choice of the function $q_0(x)$ as a polynomial of degree $n = 0$ (that is, a constant approximation) interpolating the function $f(x)$ at the midpoint $x = (a + b)/2$ of the interval $[a, b]$ gives the midpoint rule.

EXAMPLE 1. Using the midpoint rule, approximate the integral

$$I = \int_0^1 \frac{dx}{1 + x}.$$

The true value is $I = \log(2) \approx 0.693147$. Using the midpoint rule, we obtain

$$M_1 = (1 - 0)\left[\frac{1}{1 + \frac{1}{2}}\right] = \frac{2}{3} = 0.666.....$$

This is in error by

$$M_1 - I \approx -0.0265.$$

3.3 Trapezoidal rule

Consider the function $q_1(x)$ which is the polynomial of degree one that interpolates the function $f(x)$ at the integration interval endpoints $x = a$ and $x = b$; that is, the polynomial $q_1(x)$ is chosen so that the interpolating conditions

$$q_1(a) = f(a), \quad q_1(b) = f(b)$$

are satisfied. The Lagrange form of this interpolating straight line is

$$q_1(x) = \ell_1^{(1)}(x)f(a) + \ell_2^{(1)}(x)f(b),$$

where the Lagrange basis functions are

$$\ell_1^{(1)}(x) = \frac{x-b}{a-b}, \quad \ell_2^{(1)}(x) = \frac{x-a}{b-a}. \tag{8}$$

Since the function values $f(a)$ and $f(b)$ are constants, from the linearity condition (7), we obtain

$$I(q_1) = I(\ell_1^{(1)})f(a) + I(\ell_2^{(1)})f(b),$$

and since

$$I(\ell_1^{(1)}) = I(\ell_2^{(1)}) = \frac{b-a}{2},$$

we have

$$I(q_1) = w_1 f(a) + w_2 f(b),$$

where

$$w_1 = w_2 = \frac{b-a}{2}.$$

This is the so-called **trapezoidal** rule

$$T_1(f) \equiv \frac{b-a}{2}f(a) + \frac{b-a}{2}f(b) = \frac{b-a}{2}\left[f(a) + f(b)\right].$$

EXAMPLE 2. Using the trapezoidal rule, approximate the integral

$$I = \int_0^1 \frac{dx}{1+x}.$$

We obtain

$$T_1 = \frac{1}{2}\left[1 + \frac{1}{2}\right] = \frac{3}{4} = 0.75.$$

This is in error by

$$T_1 - I \approx 0.0569.$$

3.4 Errors in the Trapezoidal and Composite Trapezoidal Rules

Next, we need to see precisely how the error in approximating a definite integral $I(f)$ depends on the integrand $f(x)$. We use the trapezoidal rule to develop this analysis.

If the second derivative $f''(x)$ exists and is continuous on $[a,b]$, the error in the trapezoidal rule is

$$\begin{aligned} T_1(f) - I(f) &= I(q_1) - I(f) && (9)\\ &= I(q_1 - f) && (10)\\ &= \int_a^b \{\text{the error in linear interpolation}\}\, dx && (11)\\ &= -\int_a^b \frac{(x-a)(x-b)}{2} f''(\xi_x)dx && (12)\\ &= \int_a^b \frac{(x-a)(b-x)}{2} f''(\xi_x)dx, && (13) \end{aligned}$$

where ξ_x is an unknown point in the interval $[a,b]$ whose location depends both on the integrand $f(x)$ and on the location x. Here we have used the formula for the error in polynomial interpolation developed in Chapter 3. Now, since

$$(x-a)(b-x) \geq 0, \quad x \in [a,b],$$

we can apply the Mean Value Theorem for Integrals[2] to obtain

$$\begin{aligned} T_1(f) - I(f) &= \int_a^b \frac{(x-a)(b-x)}{2} f''(\xi_x)dx && (14)\\ & && (15)\\ &= f''(\eta) \int_a^b \frac{(x-a)(b-x)}{2} dx && (16)\\ & && (17)\\ &= \frac{(b-a)^3}{12} f''(\eta), && (18) \end{aligned}$$

where η is an (unknown) point located in the open interval (a,b) that depends only on the integrand $f(x)$. Note that the point η is necessarily unknown and must depend on the integrand $f(x)$ otherwise the formula

$$I(f) = T_1(f) - \frac{(b-a)^3}{12} f''(\eta)$$

[2]**Mean Value Theorem for Integrals.** If the functions $f(x)$ and $w(x)$ are continuous on the closed interval $[a,b]$, and the function $w(x)$ is nonnegative on the open interval (a,b), then, for some point $\eta \in [a,b]$,

$$\int_a^b w(x)f(x)dx = \left\{\int_a^b w(x)dx\right\} f(\eta).$$

could be used to evaluate *any* integral $I(f)$ exactly from explicit evaluations of $f(x)$ and of its second derivative $f''(x)$.

Now we develop a **composite** quadrature formula for

$$\int_a^b f(x)dx$$

using the trapezoidal rule. Recall that $h = x_i - x_{i-1}$, and approximate

$$\int_{x_{i-1}}^{x_i} f(x)dx$$

by the trapezoidal rule:

$$\int_{x_{i-1}}^{x_i} f(x)dx \approx \frac{x_i - x_{i-1}}{2}[f(x_{i-1}) + f(x_i)] \quad (19)$$

$$= \frac{h}{2}[f(x_{i-1}) + f(x_i)] \quad (20)$$

$$= T_1(f). \quad (21)$$

Thus the error is

$$T_1(f) - I(f) = \frac{h}{2}[f(x_{i-1}) + f(x_i)] - \int_{x_{i-1}}^{x_i} f(x)dx \quad (22)$$

$$= \frac{h^3}{12} f''(\eta_i), \quad (23)$$

where η_i is an unknown point in (x_{i-1}, x_i), and hence in (a, b). Then

$$I(f) \equiv \int_a^b f(x)dx \quad (24)$$

$$= \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f(x)dx \quad (25)$$

$$\approx \sum_{i=1}^{n} \frac{h}{2}[f(x_{i-1}) + f(x_i)] \quad (26)$$

$$= \frac{h}{2}\left[f(x_0) + 2\sum_{i=1}^{n-1} f(x_i) + f(x_n)\right] \quad (27)$$

$$\equiv T_n(f), \quad (28)$$

which is the **composite trapezoidal rule**. Assuming that $f''(x)$ is continuous on the interval $[a, b]$, the error in the composite trapezoidal rule is

$$\begin{aligned}
T_n(f) - I(f) &= \sum_{i=1}^{n} \frac{h}{2}[f(x_{i-1}) + f(x_i)] - \int_a^b f(x)dx && (29)\\
&= \sum_{i=1}^{n} \frac{h}{2}[f(x_{i-1}) + f(x_i)] - \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f(x)dx && (30)\\
&= \sum_{i=1}^{n} \left(\frac{h}{2}[f(x_{i-1}) + f(x_i)] - \int_{x_{i-l}}^{x_i} f(x)dx \right) && (31)\\
&= \frac{h^3}{12} \sum_{i=1}^{n} f''(\eta_i) && (32)\\
&= \frac{h^3}{12} n f''(\eta), && (33)
\end{aligned}$$

for some unknown point $\eta \in (a, b)$, on using the Generalized Mean Value Theorem for Sums.[3] Since $nh = (b - a)$, this expression reduces to

$$T_n(f) - I(f) = \frac{h^2}{12}(b - a) f''(\eta). \tag{34}$$

Thus, as $n \to \infty$ and $h \to 0$ simultaneously in such a way that $nh = (b - a)$, the error in the composite trapezoidal rule decreases like $O(h^2)$; that is, the error is bounded by Ch^2, where C is a positive constant.

EXAMPLE 3. Using the composite trapezoidal rule with two subintervals, approximate the integral

$$I = \int_0^1 \frac{dx}{1 + x}.$$

Then

$$T_2 = \frac{1}{2}\left[\frac{1 + \frac{2}{3}}{2}\right] + \frac{1}{2}\left[\frac{\frac{2}{3} + \frac{1}{2}}{2}\right] = \frac{17}{24} \approx 0.70833.$$

and

$$T_2 - I \approx 0.0152.$$

[3]**Generalized Mean Value Theorem for Sums.** If the function $f(x)$ is continuous on the closed interval $[a, b]$, the weights $\{w_i\}_{i=0}^n$ are all nonnegative numbers with

$$\sum_{i=0}^{n} w_i > 0,$$

and the points $\{x_i\}_{i=0}^n$ all lie in $[a, b]$, then, for some point $\eta \in [a, b]$,

$$\sum_{i=0}^{n} w_i f(x_i) = \left\{ \sum_{i=0}^{n} w_i \right\} f(\eta).$$

.

The error in T_2 is about $\frac{1}{4}$ of that for T_1 in Example 2. This is the error behavior that one would expect in general, because the error in T_n is $O(h^2)$, so that when h is halved, the error is reduced by a factor of 4.

3.5 Errors in the Midpoint and Composite Midpoint Rules

It can be shown that

$$M_1(f) - I(f) = hf(x_{i-1/2}) - \int_{x_{i-1}}^{x_i} f(x)dx = -\frac{h^3}{24} f''(\eta_i),$$

where η_i is an unknown point in (x_{i-1}, x_i). (The proof of this result is more involved than the corresponding result for $T_1(f)$, and is omitted.) It then follows that

$$M_n(f) - I(f) = -\frac{h^2}{24}(b-a)f''(\eta), \tag{35}$$

for some unknown point $\eta \in (a, b)$. Note that if $f''(x)$ is approximately constant on the interval $[a, b]$, then, from (34) and (35),

$$|M_n(f) - I(f)| \approx \frac{1}{2}|T_n(f) - I(f)|. \tag{36}$$

EXAMPLE 4. Using the composite midpoint rule with two subintervals, approximate the integral

$$I = \int_0^1 \frac{dx}{1+x}.$$

Then

$$M_2(f) = \frac{1}{2}\left[\frac{1}{1+\frac{1}{4}} + \frac{1}{1+\frac{3}{4}}\right] = \frac{24}{35} = 0.685714.$$

This is in error by

$$M_2(f) - I(f) \approx -0.0074,$$

which, in absolute value, is approximately half of the error in $T_2(f)$, as one might expect from (36) with $n = 2$. (Compare the errors in Examples 1 and 2.)

3.6 Interpolatory Quadrature: Simpson's Rule

Rather than adding points in the interval $[a, b]$ by making composite versions of simple rules such as the midpoint and trapezoidal rules, we may also generalize these rules by adding more interpolation points hence using a higher degree interpolating polynomial. Let $x_0 < x_1 < \cdots < x_n$ and $q_n(x)$ be the polynomial of degree less than or equal to n interpolating the data $\{(x_i, f(x_i))\}_{i=0}^n$. The Lagrange form of the interpolating polynomial is given by

$$q_n(x) = \sum_{i=0}^{n} f(x_i)\ell_i^{(n)}(x),$$

where the Lagrange basis functions $\ell_i^{(n)}(x)$ are defined by

$$\ell_i^{(n)}(x) = \prod_{j=0, j\neq i}^{n} \frac{x - x_j}{x_i - x_j}, \quad i = 0, \ldots, n.$$

Exploiting linearity, we have

$$I(f) \approx I(q_n) = I(\sum_{i=0}^{n} f(x_i)\ell_i^{(n)}) = \sum_{i=0}^{n} I(\ell_i^{(n)}) f(x_i) = \sum_{i=0}^{n} w_i f(x_i) \equiv R(f),$$

where the weights

$$w_i = I(\ell_i^{(n)}) = \int_a^b \ell_i^{(n)}(x)dx.$$

$R(f)$ is called an **interpolatory quadrature rule**.

When the nodes x_i are equally spaced in $[a, b]$ so that $x_i = a + ih$ where $h \equiv (b - a)/n$, we obtain the so-called closed $(n + 1)$–point **Newton–Cotes rule**. Note that the closed $(n+1)$–point Newton-Cotes rule *includes* the interval endpoints $x_0 = a$ and $x_n = b$. In contrast, the **open** $(n - 1)$–point Newton–Cotes rule has the points $x_1, x_2, \cdots, x_{n-1}$ as nodal points; it does *not* include the endpoints $x_0 = a$ and $x_n = b$ in the nodal list. It is easy to see that the open 1–point Newton–Cotes rule is the midpoint rule, and the closed 2–point Newton–Cotes rule is the trapezoidal rule. The closed 3–point Newton–Cotes rule gives the well-known **Simpson's rule**:

$$\int_a^b f(x)dx \approx \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] \equiv S_2(f). \tag{37}$$

Recall that $R(f) = I(q)$, so $R(f) - I(f) = I(q) - I(f) = I(q - f)$. Hence, $R(f)$ does not accurately approximate $I(f)$ when $I(q - f)$ is large, which can occur only when $q - f$ is large. This can happen when using many equally spaced interpolation points. Integrating the polynomial interpolant used there to approximate

$$f(x) = \frac{1}{1 + x^2}$$

would correspond to using an 11–point closed Newton–Cotes rule to approximate

$$\int_{-5}^{5} f(x)dx,$$

with a resulting large error.

However, $q - f$ can be large when $I(q - f)$ is zero. Consider the errors in the midpoint rule and Simpson's rule. The midpoint rule is derived by integrating a constant interpolating the data

$$((a + b)/2, f((a + b)/2)).$$

This interpolant is exact only for constants, so we would anticipate that the error would be zero only for constant integrands. But, like the trapezoidal rule, the midpoint rule is exact for all straight lines. How can a polynomial approximation $q(x)$ of $f(x)$ that is exact for only constants yield a quadrature rule that is also exact for all straight lines?

Similarly, Simpson's rule is derived by integrating a quadratic interpolant of the data

$$(a, f(a)), ((a+b)/2, f((a+b)/2)), (b, f(b)).$$

This quadratic interpolant is exact for all functions $f(x)$ that are quadratic polynomials, yet the quadrature rule derived from integrating this interpolant is exact for all cubic polynomials. How can a polynomial approximation $q(x)$ of $f(x)$ that is exact for only quadratic polynomials yield a quadrature rule that is also exact for all cubic polynomials?

Notice that, in both of these cases, $I(q - f) = 0$ when $q(x) - f(x)$ is not identically zero and, indeed, $q(x) - f(x)$ is potentially large.

These quadrature rules exhibit *superconvergence*; that is, the rules integrate exactly all polynomials of a certain higher degree than is to be anticipated from their construction. Indeed, all Newton–Cotes rules (closed or open) with an odd number of points exhibit superconvergence; that is, they each integrate exactly all polynomials of degree one higher than the degree of the polynomial integrated to derive the rule. As we shall see later, Gaussian quadrature yields the ultimate in superconvergent quadrature rules.

EXAMPLE 5. Using Simpson's rule, approximate the integral

$$I = \int_0^1 \frac{dx}{1+x}.$$

Then

$$S_2 = \frac{1}{6}\left[1 + 4\frac{2}{3} + \frac{1}{2}\right] = \frac{25}{36} \approx 0.69444.$$

The error is

$$S_2 - I = S_2 - \log(2) \approx 0.00130.$$

To compare with the trapezoidal rule, we use T_2 from the previous example, since the number of function evaluations is the same for both T_2 and S_2. The error in S_2 is smaller than that for T_2 by a factor of about 12, a significant increase in accuracy.

4 Degree of Precision

In the following, we present an alternative way to derive quadrature rules.

4.1 Degree of Precision

Definition (*Degree of precision (DOP)*) The quadrature rule

$$R(f) = \sum_{i=0}^{n} w_i f(x_i)$$

approximating the definite integral

$$I(f) = \int_a^b f(x)dx$$

has DOP $= m$ if

$$\int_a^b f(x)dx = \sum_{i=0}^{n} w_i f(x_i)$$

whenever $f(x)$ is a polynomial of degree at most m, but

$$\int_a^b f(x)dx \neq \sum_{i=0}^{n} w_i f(x_i)$$

for some polynomial $f(x)$ of degree $m+1$.

An equivalent definition of DOP is given in the following.

Definition The quadrature rule

$$R(f) = \sum_{i=0}^{n} w_i f(x_i)$$

approximating the definite integral

$$I(f) = \int_a^b f(x)dx$$

has DOP $= m$ if

$$\int_a^b x^r dx = \sum_{i=0}^{n} w_i x_i^r,$$

for $r = 0, 1, \cdots, m$, but, for $r = m+1$,

$$\int_a^b x^r dx \neq \sum_{i=0}^{n} w_i x_i^r.$$

If a quadrature rule $R_{hi}(f)$ has a higher DOP than another rule $R_{lo}(f)$, then $R_{hi}(f)$ is generally considered more accurate than $R_{lo}(f)$ because it integrates exactly higher degree polynomials and hence potentially integrates exactly more accurate polynomial approximations to $f(x)$.

The DOP concept may be used to derive quadrature rules directly. With the points $x_i, i = 0, 1, \cdots, n$, given, consider the rule

$$I(f) = \int_{-1}^{1} f(x)dx \approx \sum_{i=0}^{n} w_i f(x_i) = R(f).$$

Note that we have chosen a special (*canonical*) interval $[-1, 1]$ here. The weights, $w_i, i = 0, 1, \cdots, n$, are chosen to maximize the DOP, by solving the following *equations of precision* (starting from the first and leaving out no equations):

$$\int_{-1}^{1} 1dx = \sum_{i=0}^{n} w_i, \tag{38}$$

$$\int_{-1}^{1} xdx = \sum_{i=0}^{n} w_i x_i, \tag{39}$$

$$\vdots \tag{40}$$

$$\int_{-1}^{1} x^m dx = \sum_{i=0}^{n} w_i x_i^m, \tag{41}$$

for the weights w_i. When we reach an equation that we cannot satisfy, for example, we have satisfied equations the first $(m+1)$ equations but we cannot satisfy the next equation so that

$$\int_{-1}^{1} x^{m+1} dx \neq \sum_{i=0}^{n} w_i x_i^{m+1},$$

then the DOP corresponds to the last power of x for which we succeeded in satisfying the corresponding equation of precision, so that DOP$= m$.

EXAMPLE 1. Suppose that the quadrature rule

$$R(f) = w_0 f(-1) + w_1 f(0) + w_2 f(1)$$

approximates the integral

$$I(f) = \int_{-1}^{1} f(x)dx.$$

What choice of the weights w_0, w_1 and w_2 maximizes the DOP of the rule?

Solution: Create a table listing the values of $I(x^j)$ and $R(x^j)$ for $j = 0, 1, 2, \ldots$.

j	$I(x^j)$	$R(x^j)$
0	2	$w_0 + w_1 + w_2$
1	0	$-w_0 + w_2$
2	$\frac{2}{3}$	$w_0 + w_2$
3	0	$-w_0 + w_2$
4	$\frac{2}{5}$	$w_0 + w_2$

To determine the three free parameters, w_0, w_1 and w_2, we solve the first three equations of precision (to give us DOP≥ 2). That is, we solve $I(x^m) = R(x^m)$ for $m = 0, 1, 2$:

$$\begin{array}{ll} 2 = w_0 + w_1 + w_2 & (I(1) = R(1)) \\ 0 = -w_0 + w_2 & (I(x) = R(x)) \\ \frac{2}{3} = w_0 + w_2 & (I(x^2) = R(x^2)) \end{array}$$

These three equations have the unique solution:

$$w_0 = w_2 = \frac{1}{3}, w_1 = \frac{4}{3},$$

which corresponds to Simpson's rule. However, this rule has DOP $= 3$ not DOP $= 2$ because, for this choice of weights, $I(x^3) = R(x^3)$ also; that is, the first four equations of precision are satisfied. (Indeed, $I(x^m) = R(x^m) = 0$ for all odd powers $m \geq 0$.) DOP$= 3$ because if DOP$= 4$ then the equations $w_0 + w_2 = \frac{2}{3}$ and $w_0 + w_2 = \frac{2}{5}$ would both be satisfied, a clear contradiction.

5 Transformation from a Canonical Interval

Question. How to transform a quadrature rule for a canonical interval of integration, chosen here as $[-1, 1]$, to a quadrature rule on a general interval of integration $[a, b]$?

Denote the variable $x^\star \in [-1, 1]$, and $x \in [a, b]$. A linear change of variable (mapping) from $x^\star$ to x is,

$$x = g(x^\star) = \alpha x^\star + \beta.$$

Since $x(-1) = a$ and $x(1) = b$

$$\alpha = \frac{b-a}{2}, \quad \beta = \frac{a+b}{2},$$

$$x = g(x^\star) = \frac{b-a}{2} x^\star + \frac{a+b}{2}.$$

so the transformation $g(x^\star)$ maps the canonical interval onto the interval of interest.

The standard change of variable formula of integral calculus using the transformation, $x = g(x^\star)$ is

$$\int_a^b f(x)dx = \int_{-1}^1 f(g(x^\star))dg(x^\star) = \int_{-1}^1 f(g(x^\star))g'(x^\star)dx^\star.$$

For this transformation, we have

$$g'(x^\star) = \frac{b-a}{2},$$

and the change of variable formula reads

$$\int_a^b f(x)dx = \frac{b-a}{2}\int_{-1}^{1} f\left(\frac{b-a}{2}x^\star + \frac{b+a}{2}\right)dx^\star.$$

Now, assume that for the canonical interval $[-1,1]$ we have a quadrature rule

$$I^*(u) = \int_{-1}^{1} u(x^\star)dx^\star \approx \sum_{i=0}^{n} w_i^* u(x_i^*) = R^*(u).$$

Then, substituting for $u(x^\star) = f(g(x^\star)) \equiv (f\circ g)(x^\star)$, we have

$$\begin{aligned}
I(f) \equiv \int_a^b f(x)dx &= \frac{b-a}{2}\int_{-1}^{1} f\left(\frac{b-a}{2}x^\star + \frac{b+a}{2}\right)dx^\star \\
&= \frac{b-a}{2} I^*(f\circ g) \\
&\approx \frac{b-a}{2} R^*(f\circ g) \\
&= \frac{b-a}{2}\sum_{i=0}^{n} w_i^* f\left(\frac{b-a}{2}x_i^* + \frac{b+a}{2}\right) \\
&= \sum_{i=0}^{n} w_i f(x_i) \\
&\equiv R(f),
\end{aligned}$$

where in $R(f)$ the weights

$$w_i = \frac{b-a}{2}w_i^*, \tag{42}$$

and the points

$$x_i = \frac{b-a}{2}t_i^* + \frac{b+a}{2}. \tag{43}$$

Remark 5.1. The DOP of the transformed quadrature rule $R(f)$ is the same as the DOP of the canonical quadrature rule $R^*(f)$. The error term for the canonical quadrature rule may be transformed using $g(t)$ to obtain the error term for the transformed quadrature rule.

Example 5.2. The simpson's rule on $[-1,1]$ is

$$\int_{-1}^{1} f(x)dx = \frac{1}{3}f(-1) + \frac{4}{3}f(0) + \frac{1}{3}f(1).$$

with $x_0^\star = -1$, $x_1^\star = 0$, $x_2^\star = 1$ and $w_0^\star = \frac{1}{3}$, $w_1^\star = \frac{4}{3}$, $w_2^\star = \frac{1}{3}$. The simpson's rule on general interval $[a,b]$,

$$\int_a^b f(x)dx = \sum_i w_i f(x_i).$$

What are the x_i and w_i?

Solution: The linear transformation is from $x^\star \in [-1,1]$ to $x \in [a,b]$ is

$$x = g(x^\star) = \frac{b-a}{2}x^\star + \frac{a+b}{2}.$$

Therefore, from equation (42) and (43), we have $x_0 = a$, $x_1 = \frac{a+b}{2}$, $x_2 = b$ and $w_0 = \frac{h}{6}$, $w_1 = \frac{2h}{3}$, $w_2 = \frac{h}{6}$, assuming that $b - a = h$.

6 Gaussian Quadrature

The most popular choices for rules for numerical integration are the Gaussian quadrature rules, for which the canonical interval is $[-1,1]$; here,

$$I(f) \equiv \int_{-1}^{1} f(x)dx \approx \sum_{i=0}^{n} w_i f(x_i) \equiv R(f),$$

where the all weights w_i, $i = 0, 1, \cdots, n$, and all the points x_i, $i = 0, 1, \cdots, n$, are chosen to maximize the DOP. In the problems at the end of this section, you are asked to find the weights and points in some simple Gaussian quadrature rules. In reality, Gaussian rules with much larger numbers of points than in these problems are used.

If the points $x_0, x_1, \cdots, x_n$ were fixed arbitrarily, then, by analogy with those rules that we have derived previously, with $n+1$ free weights we would expect DOP$= n$, or, in some cases, DOP$=n+1$. But in the Gaussian quadrature rules the points are chosen to increase the DOP to $2n+1$; that is, the Gaussian quadrature rules are highly superconvergent. This should not be a surprise: the Gaussian quadrature rule has a total of $2(n+1)$ unknowns, taking the weights and the points together, and it is plausible that they can be chosen to solve all the $2(n+1)$ equations of precision $R(x^k) = I(x^k)$ for $k = 0, 1, \cdots, 2n+1$.

Theorem 6.1. *Let $q(x)$ be a nontrivial polynomial of degree $n+1$ such that*

$$\int_{-1}^{1} x^k q(x)dx = 0, \quad 0 \le k \le n.$$

Let x_0, $x_1, \cdots, x_n$ be zeros of $q(x)$. Then the formula

$$\int_{-1}^{1} f(x)dx \approx \sum_i f(x_i)w_i,$$

where $w_i = \int_{-1}^{1} l_i(x)dx$ with these x_i's as nodes will be exact for all polynomial of degree at most $2n+1$. In other words, the DOP for the formula is $2n+1$.

Before we prove the Theorem, let us work through a couple of examples.

Example 6.2. (n=0). $q(x)$ is a linear function, $q(x) = ax + b$, and

$$\int_{-1}^{1} q(x)dx = 0,$$

therefore $b = 0$, and $q(x) = ax$, the zero of which is $x_0 = 0$. Then the formula

$$\int_{-1}^{1} f(x)dx \approx f(0)w_0 = 2f(0),$$

is the mid-point rule and the DOP for the mid-point rule is $1(= 2n + 1)$.

Example 6.3. (n=1). $q(x)$ is a quadratic function, $q(x) = ax^2 + bx + c$, and

$$\int_{-1}^{1} q(x)dx = \int_{-1}^{1} (ax^2 + bx + c)dx = \frac{ax^3}{3}|_{-1}^{1} + 2c = \frac{2a}{3} + 2c = 0$$

$$\int_{-1}^{1} q(x)xdx = \int_{-1}^{1} (ax^3 + bx^2 + cx)dx = \frac{2b}{3} = 0$$

therefore $q(x) = ax^2 - \frac{a}{3} = a(x^2 - \frac{1}{3})$, the zero of which is $x_0 = -\frac{1}{\sqrt{3}}$ and $x_1 = \frac{1}{\sqrt{3}}$. Then the formula

$$\int_{-1}^{1} f(x)dx \approx f(-\frac{1}{\sqrt{3}})w_0 + f(\frac{1}{\sqrt{3}})w_1.$$

To compute w_0 and w_1, as well as the DOP of the quadrature formula, we use the DOP approach. Create a table listing the values of $I(x^j)$ and $R(x^j)$ for $j = 0, 1, 2, \ldots$.

j	$I(x^j)$	$R(x^j)$
0	2	$w_0 + w_1$
1	0	$(-w_0 + w_1)\frac{1}{\sqrt{3}}$
2	$\frac{2}{3}$	$\frac{1}{3}(w_0 + w_1)$
3	0	$(-w_0 + w_1)\frac{1}{3\sqrt{3}}$
4	$\frac{2}{5}$	$\frac{1}{9}(w_0 + w_1)$

From the first four equations, we know that $w_0 = w_1 = 1$ and the DOP for the formula is $3(= 2n + 1)$. The order of approximation is $\mathcal{O}(h^4)$.

Proof of Theorem 5.1. (see Theorem 1 in Chap 6.2 of the textbook 'Numerical Mathematics and Computing' by Cheney and Kincaid)

Two point and three point Gaussian quadrature rules. The following table shows some cases of the Gaussian quadrature rules.

n	Nodes x_i	Weights w_i
1	$-\sqrt{\frac{1}{3}}$ $\sqrt{\frac{1}{3}}$	1 1
2	$-\sqrt{\frac{3}{5}}$ 0 $\sqrt{\frac{3}{5}}$	$\frac{5}{9}$ $\frac{8}{9}$ $\frac{5}{9}$

Example 6.4. Using the two point Gaussian quadrature rule, approximate the integral

$$I = \int_0^1 \frac{dx}{1+x}.$$

The true value is $I = \log(2) \approx 0.693147$.

Solution: From Example 6.3, the two point Gaussian quadrature rule is

$$R(f) = f(x_0^\star)w_0^\star + f(x_1^\star)w_1^\star \approx \int_{-1}^{1} f(x)dx,$$

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

By equation (43) and (42), the two point Gaussian quadrature rule for general interval $[a, b]$ is

$$R(f) = f(x_0)w_0 + f(x_1)w_1 \approx \int_a^b f(x)dx,$$

with $x_0 = -\frac{1}{\sqrt{3}}\frac{b-a}{2} + \frac{b+a}{2}$, $x_1 = \frac{1}{\sqrt{3}}\frac{b-a}{2} + \frac{b+a}{2}$, $w_0 = w_1 = \frac{b-a}{2}$. Especially when $a = 0$ and $b = 1$,

$$x_0 = \frac{1}{2}(1 - \frac{1}{\sqrt{3}}), \quad x_1 = \frac{1}{2}(1 + \frac{1}{\sqrt{3}}),$$

$$w_0 = w_1 = \frac{1}{2}.$$

Therefore,

$$R(f) = f(x_0)w_0 + f(x_1)w_1 = \frac{1}{2}(f(\frac{1}{2}(1 - \frac{1}{\sqrt{3}})) + f(\frac{1}{2}(1 + \frac{1}{\sqrt{3}})))$$

Since

$$f(\frac{1}{2}(1 - \frac{1}{\sqrt{3}})) = \frac{1}{1 + \frac{1}{2}(1 - \frac{1}{\sqrt{3}})} = \frac{1}{3 - \frac{1}{\sqrt{3}}}$$

$$f(\frac{1}{2}(1 + \frac{1}{\sqrt{3}})) = \frac{1}{1 + \frac{1}{2}(1 + \frac{1}{\sqrt{3}})} = \frac{1}{3 + \frac{1}{\sqrt{3}}},$$

$$R(f) = \frac{1}{2}(\frac{1}{3 - \frac{1}{\sqrt{3}}} + \frac{1}{3 + \frac{1}{\sqrt{3}}}) = 9/13$$

This is in error by

$$9/13 - \log(2) \approx 0.0008.$$

Compare with the error from the mid-point rule 0.0074 (one function evaluation), from the Trapezoid rule 0.0152 (two function evaluations) and the Simpson's rule 0.0013 (three function evaluations), the two point Gaussian quadrature rule has the smallest error with only two function evaluations, indicating the best efficiency.

7 Properties of Gaussian quadrature rule (optional reading-but very interesting)

Gauss rules have a number of interesting properties (some of which we will derive subsequently). For each value of $n \geq 0$, the $(n+1)$–point Gauss rule has the following properties:

(1) The DOP of the $(n+1)$–point Gauss rule is $2n+1$.

(2) All of the weights w_i are positive. (This guarantees the *stability* of the rules; cf., the comment earlier concerning the weights in Newton-Cotes rules for $n \geq 10$.) (Addressed in Section 7.2)

(3) All of the points x_i are real, distinct, and lie in the open interval $(-1, 1)$. (These points are the roots of a *Legendre polynomial.*) (Addressed in Section 7.1)

(4) The points x_i are placed symmetrically about the origin and the weights w_i are correspondingly symmetric. For n odd, the points satisfy $x_0 = -x_n, x_1 = -x_{n-1}$, etc., and the weights satisfy $w_0 = w_n, w_1 = w_{n-1}$, etc. For n even, the points and weights satisfy the same relations as for n odd plus we have $x_{n/2} = 0$.

(5) The points $\overline{x}_i$ of the n–point Gauss rule interlace the points x_i of the $(n+1)$–point Gauss rule: $-1 < x_0 < \overline{x}_0 < x_1 < \overline{x}_1 < x_2 < \cdots < x_{n-1} < \overline{x}_{n-1} < x_n < 1$.

(6) The Gauss rules are interpolatory quadrature rules; that is, after the points $x_0, x_1, \cdots, x_n$ have been determined, then the weights $w_0, w_1, \cdots, w_n$ may be computed by integrating over the interval $[-1, 1]$ the polynomial of degree n that interpolates the integrand $f(x)$ at the points $x_0, x_1, \cdots, x_n$.

7.1 Legendre Polynomials: the nodes in Gaussian quadrature rule.

From Theorem 6.1, it is known that the $x_i's$ are the roots for the polynomial $q(x)$, which is known as Legendre polynomials. The Legendre polynomial of order $k \geq 0$ is defined by

$$L_k(x) = \frac{(-1)^k k!}{(2k)!} \frac{d^k}{dx^k} \left[(1 - x^2)^k \right], \quad k = 0, 1, 2, \ldots \tag{44}$$

For example, for $k = 0, 1, 2, 3$, we have

$$L_0(x) = 1, \quad L_1(x) = x, \quad L_2(x) = x^2 - \frac{1}{3}, \quad L_3(x) = x^3 - \frac{3}{5}x. \tag{45}$$

Lemma 6.3.1 *For any non-negative integers n and m, we have*

$$\int_{-1}^{1} L_n(x) L_m(x) dx = 0, \tag{46}$$

provided that $n \neq m$.

Proof: Equation (46) is equivalent to

$$\int_{-1}^{1} \frac{d^n}{dx^n}\left[(1-x^2)^n\right] \frac{d^m}{dx^m}\left[(1-x^2)^m\right] dx = 0. \tag{47}$$

Since $n \neq m$, without loss of generality, we may assume that $n > m$. Using integration by parts n times, we can rewrite (47) as

$$(-1)^n \int_{-1}^{1} (1-x^2)^n \frac{d^{m+n}}{dx^{m+n}}\left[(1-x^2)^m\right] dx = 0, \tag{48}$$

which is clearly true since $n + m > 2m$ and $2m$ is the order of the polynomial $(1-x^2)^m$.

Lemma 6.3.1 indicates that the Legendre polynomial $L_n(x)$ is *orthogonal* to all polynomials of degree $\leq n-1$. Mathematically, this orthogonality can be expressed as

$$\int_{-1}^{1} L_n(x) x^s dx = 0, \quad s = 0, 1, \cdots, n-1. \tag{49}$$

Lemma 6.3.2 *The Legendre polynomial* $L_n(x)$ *has no complex roots.*

Proof: Suppose that $L_n(x)$ has a complex root $z_0 = a + ib$ for $b \neq 0$. Then, the conjugate of z_0 is also a root of $L_n(x)$. As a result, the Legendre polynomial of $L_n(x)$ can be written as

$$L_n(x) = q(x)[(x-a)^2 + b^2],$$

where $q(x)$ is a polynomial of degree $n-2$. The orthogonality property (49) then implies that

$$0 = \int_{-1}^{1} L_n(x) q(x) dx = \int_{-1}^{1} q^2(x)[(x-a)^2 + b^2] dx > 0,$$

which is a contradiction.

Lemma 6.3.3 *The Legendre polynomial* $L_n(x)$ *has* n *distinct roots* $x_0, x_1, \cdots, x_{n-1}$*, which lie in the open interval* $(-1, 1)$.

Proof: We first show that all the (real) roots of $L_n(x)$ lie in the open interval $(-1, 1)$. Suppose that x_* is a real root of $L_n(x)$ that does not belong to $(-1, 1)$. Without loss of generality, we may assume that $x_* \leq -1$. As a result, the function $g(x) = x - x_*$ is non-negative in the interval $[-1, 1]$. Let $p_{n-1}(x)$ be a polynomial of degree $n-1$ such that

$$L_n(x) = p_{n-1}(x)(x - x_*).$$

It follows that

$$\int_{-1}^{1} L_n(x) p_{n-1}(x) dx = \int_{-1}^{1} p_{n-1}^2(x)(x - x_*) dx > 0.$$

But the orthogonality property (49) claims that the left-hand side of the above is equal to 0. Therefore, the assumption that $L_n(x)$ has a root x_* outside the open interval $(-1, 1)$ is incorrect.

Next we show that all the roots are simple. Assume, on the contrary, that x_* is a multiple root. It follows that

$$L_n(x) = p_{n-2}(x)(x - x_*)^2$$

for some polynomial $p_{n-2}(x)$ of degree $n - 2$. Thus, we have from (49)

$$0 = \int_{-1}^{1} L_n(x)p_{n-2}(x)dx = \int_{-1}^{1} p_{n-2}^2(x)(x - x_*)^2 dx > 0,$$

which is a contradiction.

7.2 The weights in Gaussian quadrature rules

The roots of the Legendre polynomials $L_n(x)$ are called *Gauss nodes of degree* n. As was shown in the previous section, these roots are real, distinct and lie in the open interval $(-1, 1)$. From (45), we see that the Gauss nodes of degree 1 comprise only one point, $x_0 = 0$, the Gauss nodes of degree 2 are

$$x_0 = -\sqrt{\frac{1}{3}}, \quad x_1 = \sqrt{\frac{1}{3}},$$

and the Gauss nodes of degree 3 are

$$x_0 = -\sqrt{\frac{3}{5}}, \quad x_1 = 0, \quad x_2 = \sqrt{\frac{3}{5}}.$$

The Gauss quadrature rules are interpolatory quadrature rules based on the Gauss nodes (cf., Section 6.4.6). Let

$$x_0, x_1, \cdots, x_n$$

be the roots of $L_{n+1}(x)$ and let $f(x)$ be a continuous function defined on $[-1, 1]$. Let $q_n(x)$ be the polynomial interpolating $f(x)$ at the Gauss nodes of degree $n + 1$. Then,

$$\int_{-1}^{1} f(x)dx \approx \int_{-1}^{1} q_n(x)dx = \sum_{i=0}^{n} w_i f(x_i),$$

where

$$w_i = \int_{-1}^{1} \ell_i^{(n)}(x)dx, \tag{50}$$

and

$$\ell_i^{(n)}(x) = \prod_{j=0, j\neq i}^{n} \frac{x - x_j}{x_i - x_j},$$

the Lagrange basis function corresponding to the node x_i. Now,

$$\ell_i^{(n)}(x) = \frac{\omega_{n+1}(x)}{(x-x_i)\omega'_{n+1}(x_i)}, \tag{51}$$

where

$$\omega_{n+1}(x) = \prod_{j=0}^{n}(x-x_i).$$

Thus, substituting (51) into (50) gives

$$w_i = \frac{1}{\omega'_{n+1}(x_i)} \int_{-1}^{1} \frac{\omega_{n+1}(x)}{x-x_i} dx, \quad i = 0, 1, \cdots, n.$$

Using this result, we can prove that the weights are positive.

Lemma 6.3.4 *The weights, w_i, in the Gauss quadrature rules are positive for all $i = 0, 1, \cdots, n$, and all n.*

Proof: Since the Gauss quadrature rule with $n+1$ nodes has degree of precision $2n+1$, it yields the exact value for

$$\int_{-1}^{1} f(x)dx$$

when $f(x)$ is any polynomial of degree $2n+1$ or less. In particular, it is exact for

$$r_i(x) = \frac{\omega_{n+1}^2(x)}{(x-x_i)^2}, \quad i = 0, 1, \cdots, n,$$

which are polynomials of degree $2n$; that is,

$$\int_{-1}^{1} r_i(x)dx = \sum_{k=0}^{n} w_k r_i(x_k). \tag{52}$$

However, it is clear that

$$\begin{aligned} r_i(x_k) &= 0, \quad \text{for } k \neq i, \\ r_i(x_i) &= \prod_{j=0, j\neq i}^{n} (x_i - x_j)^2 = [\omega'_n(x_i)]^2 > 0. \end{aligned} \tag{53}$$

Using (53) in (52) , we find that

$$w_i = \frac{1}{r_i(x_i)} \int_{-1}^{1} r_i(x)dx = \frac{1}{[\omega'_{n+1}(x_i)]^2} \int_{-1}^{1} \frac{\omega_{n+1}^2(x)}{(x-x_i)^2} dx > 0,$$

which complete the proof.