Numerical Integration and Differentiation

There are two reasons for approximating derivatives and integrals of a function f(x). One is when the function is very difficult to differentiate or integrate, or only the tabular values are available for the function. Another reason is to obtain solution of a differential or integral equation. In this chapter we introduce some basic methods to approximate integral and derivative of a function given either explicitly or by tabulated values.

In Section 10.1, we obtain numerical methods for evaluating the integral of a given integrable function f defined on the interval [a, b]. Section 10.2 introduces various ways to obtain numerical formulas for approximating derivatives of a given differentiable function.

10.1 Numerical Integration

In this section we derive and analyze numerical methods for evaluating definite integrals. The problem is to evaluate the number

$$I(f) = \int_{a}^{b} f(x)dx. \tag{10.1}$$

Most such integrals cannot be evaluated explicitly, and with many others, it is faster to integrate numerically than explicitly. The process of approximating the value of I(f) is usually referred to as numerical integration or quadrature rule.

The idea behind numerical integration is to approximate the integrand f by a simpler function that can be integrated easily. One obvious approximation is the interpolation by polynomials. Thus, we approximate I(f) by $I(p_n)$, where $p_n(x)$ is the interpolating polynomial for the integrand f at some appropriately chosen nodes x_0, \dots, x_n . The general form of the approximation is

$$I(f) \approx I(p_n) = w_0 f(x_0) + w_1 f(x_1) + \dots + w_n f(x_n),$$

where the *weights* are given by

$$w_i = I(l_i),$$

with $l_i(x)$ the i^{th} Lagrange polynomial.

Without going through interpolation, we now propose a general formula

$$I(f) \approx w_0 f(x_0) + w_1 f(x_1) + \dots + w_n f(x_n),$$
 (10.2)

where x_0, \dots, x_n are distinct real numbers (called *quadrature points*) and w_0, \dots, w_n are real numbers (called *quadrature weights*). When the quadrature points are equally spaced, the quadrature formula of the form (10.2) is called the *Newton-Cotes formula*.

The Newton-Cotes formula (10.2) gives rise to different quadrature formulas depending on the degree of the interpolating polynomial n and also on the choice of the nodes. We now study few simple quadrature formulas.

10.1.1 Rectangle Rule

We now consider the case when n = 0 in (10.2). Then, the corresponding interpolating polynomial is the constant function $p_0(x) = f(x_0)$, and therefore

$$I(p_0) = (b-a)f(x_0).$$

From this, we can obtain two quadrature rules depending on the choice of x_0 .

• If $x_0 = a$, then this approximation becomes

$$I(f) \approx I_R(f) := (b - a)f(a) \tag{10.3}$$

and is called the *rectangle rule*. The geometrical interpretation of the rectangle rule is illustrated in Figure 10.1.

• If $x_0 = (a+b)/2$, we get

$$I(f) \approx I_M(f) := (b - a)f\left(\frac{a + b}{2}\right) \tag{10.4}$$

and is called the *mid-point rule*.

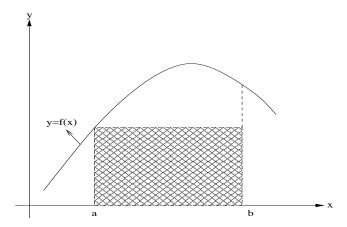


Figure 10.1: Geometric interpretation of the rectangle Rule.

We now obtain the *mathematical error* in rectangle rule, given by,

$$ME_R(f) := I(f) - I(p_0).$$

Theorem 10.1.1 [Error in Rectangle Rule].

Let $f \in C^1[a, b]$. The mathematical error $ME_R(f)$ of the rectangle rule takes the form

$$ME_R(f) = \frac{f'(\eta)(b-a)^2}{2},$$
 (10.5)

for some $\eta \in (a, b)$.

Proof.

Let $p_0(x)$ be the polynomial interpolating the function f at the node a. For each $x \in [a, b]$, we have

$$f(x) = p_0(x) + f[a, x](x - a).$$

Therefore, the mathematical error in the rectangle rule is given by

$$ME_R(f) = I(f) - I_R(f) = \int_a^b f[a, x](x - a) dx.$$
 (10.6)

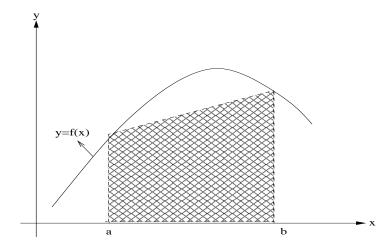


Figure 10.2: Geometric interpretation of the trapezoidal Rule.

From Corollary ?? (Conclusion 1), we see that the function $x \mapsto f[a, x]$ is continuous. Therefore, from the mean value theorem for integrals (after noting that (x - a) is negative for all $x \in [a, b]$), the expression (10.6) for the mathematical error takes the form

$$ME_{R}(f) = f[a, \xi] \int_{a}^{b} (x - a) dx,$$

for some $\xi \in (a, b)$. By mean value theorem, $f[a, \xi] = f'(\eta)$ for some $\eta \in (a, \xi)$. Thus, we get

$$ME_{R}(f) = \frac{f'(\eta)(b-a)^{2}}{2},$$

for some $\eta \in (a, b)$.

10.1.2 Trapezoidal Rule

We now consider the case when n = 1. Then

$$p_1(x) = f(x_0) + f[x_0, x_1](x - x_0),$$

and therefore

$$I(f) \approx I_T(f) := \int_a^b (f(x_0) + f[x_0, x_1](x - x_0)) dx.$$

Taking $x_0 = a$ and $x_1 = b$, we get

$$I_T(f) = (b-a)\left(\frac{f(a) + f(b)}{2}\right)$$
 (10.7)

and is called the *trapezoidal Rule*. The Trapezoidal rule is illustrated in Figure 10.2.

We now obtain the *mathematical error* in trapezoidal rule, given by

$$ME_T(f) := I(f) - I(p_1).$$

Theorem 10.1.2 [Error in Trapezoidal Rule].

Let $f \in C^2[a, b]$. The mathematical error $ME_T(f)$ of the trapezoidal rule takes the form

$$ME_{T}(f) = -\frac{f''(\eta)(b-a)^{3}}{12},$$
(10.8)

for some $\eta \in (a, b)$.

Proof.

We have for $x \in [a, b]$

$$f(x) = f(a) + f[a, b](x - a) + f[a, b, x](x - a)(x - b).$$

Integrating over the interval [a, b], we get

$$I(f) = I_T(f) + \int_a^b f[a, b, x](x - a)(x - b)dx.$$

Therefore the mathematical error is given by

$$ME_{T}(f) = I(f) - I_{T}(f) = \int_{a}^{b} f[a, b, x](x - a)(x - b)dx.$$
 (10.9)

From Corollary ?? (Conclusion 1), we see that the function $x \mapsto f[a, b, x]$ is continuous. Therefore, from the mean value theorem for integrals (after noting that (x-a)(x-b) is negative for all $x \in [a,b]$), the expression (10.9) for the mathematical error takes the form

$$ME_{T}(f) = f[a, b, \eta] \int_{a}^{b} (x - a)(x - b)dx,$$

for some $\eta \in (a, b)$. The formula (10.8) now follows from (8.21) and a direct evaluation of the above integral.

Example 10.1.3.

For the function f(x) = 1/(x+1), we approximate the integral

$$I = \int_{0}^{1} f(x)dx,$$

using trapezoidal rule to get

$$I_T(f) = \frac{1}{2} \left(1 + \frac{1}{2} \right) = \frac{3}{4} = 0.75.$$

The true value is $I(f) = \log(2) \approx 0.693147$. Therefore, the error is $ME_T(f) \approx -0.0569$. Using the formula (10.8), we get the bounds for $ME_T(f)$ as

$$-\frac{1}{6} < ME_{T}(f) < -\frac{1}{48}$$

which clearly holds in the present case.

Composite Trapezoidal Rule

We can improve the approximation of trapezoidal rule by breaking the interval [a, b] into smaller subintervals and apply the trapezoidal rule (10.7) on each subinterval. We will derive a general formula for this.

Let us subdivide the interval [a, b] into n equal subintervals of length

$$h = \frac{b - a}{n}$$

with endpoints of the subintervals as

$$x_j = a + jh, \quad j = 0, 1, \cdots, n.$$

Then, we get

$$I(f) = \int_{a}^{b} f(x)dx = \int_{x_0}^{x_n} f(x)dx = \sum_{j=0}^{n-1} \int_{x_j}^{x_{j+1}} f(x)dx.$$

Using trapezoidal rule (10.7) on the subinterval $[x_j, x_{j+1}]$, we get

$$\int_{x_j}^{x_{j+1}} f(x)dx \approx h\left(\frac{f(x_j) + f(x_{j+1})}{2}\right), \ j = 0, 1, \dots, n-1.$$

Substituting this in the above equation, we get

$$I(f) \approx h \left[\frac{f(x_0) + f(x_1)}{2} \right] + h \left[\frac{f(x_1) + f(x_2)}{2} \right] + \dots + h \left[\frac{f(x_{n-1}) + f(x_n)}{2} \right].$$

The terms on the right hand side can be combined to give the simpler formula

$$I_T^n(f) := h \left[\frac{1}{2} f(x_0) + f(x_1) + f(x_2) + \dots + f(x_{n-1}) + \frac{1}{2} f(x_n) \right]. \tag{10.10}$$

This rule is called the *composite trapezoidal rule*.

Example 10.1.4.

Using composite trapezoidal rule with n=2, let us approximate the integral

$$I = \int_{0}^{1} f(x)dx,$$

where

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

As we have seen in Example 10.1.3, the true value is $I(f) = \log(2) \approx 0.693147$. Now, the composite trapezoidal rule with $x_0 = 0$, $x_1 = 1/2$ and $x_2 = 1$ gives

$$I_T^2(f) \approx 0.70833.$$

Thus the error is -0.0152. Recall from Example 10.1.3 that with n = 1, the trapezoidal rule gave an error of -0.0569.

10.1.3 Simpson's Rule

We now calculate $I(p_2(x))$ to obtain the formula for the case when n=2. Let us choose $x_0=a, x_1=(a+b)/2$ and $x_2=b$. The Lagrange form of interpolating polynomial is

$$p_2(x) = f(x_0)l_0(x) + f(x_1)l_1(x) + f(x_2)l_2(x).$$

Then

$$\int_{a}^{b} p_{2}(x)dx = f(x_{0}) \int_{a}^{b} l_{0}(x) dx + f(x_{1}) \int_{a}^{b} l_{1}(x) dx + f(x_{2}) \int_{a}^{b} l_{2}(x) dx.$$

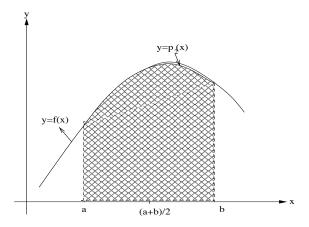


Figure 10.3: Geometric interpretation of the Simpson's Rule.

Using the change of variable

$$x = \frac{b-a}{2}t + \frac{b+a}{2},$$

we get

$$\int_{a}^{b} l_{0}(x) dx = \int_{a}^{b} \frac{(x - x_{1})(x - x_{2})}{(x_{0} - x_{1})(x_{0} - x_{2})} dx$$
$$= \int_{-1}^{1} \frac{t(t - 1)}{2} \frac{b - a}{2} dt = \frac{b - a}{6}.$$

Similarly, we can see

$$\int_{a}^{b} l_{1}(x) dx = \frac{4}{6}(b-a),$$

$$\int_{a}^{b} l_{2}(x) dx = \frac{b-a}{6}.$$

We thus arrive at the formula

$$I(f) \approx I_S(f) := \int_a^b p_2(x)dx = \frac{b-a}{6} \left\{ f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right\}$$
 (10.11)

which is the famous *Simpson's Rule*. The Simpson's rule is illustrated in Figure 10.3.

We now obtain the *mathematical error* in Simpson's rule, given by,

$$ME_S(f) := I(f) - I(p_2).$$

Theorem 10.1.5 [Error in Simpson's Rule].

Let $f \in C^4[a, b]$. The mathematical error $ME_S(f)$ of the Simpson's rule takes the form

$$ME_S(f) = -\frac{f^{(4)}(\eta)(b-a)^5}{2880},$$
 (10.12)

for some $\eta \in (a, b)$.

Proof of the above theorem is omitted for this course.

Example 10.1.6.

We now use the Simpson's rule to approximate the integral

$$I(f) = \int_{0}^{1} f(x) dx$$
, where $f(x) = \frac{1}{1+x}$.

The true value is $I(f) = \log(2) \approx 0.693147$. Using the Simpson's rule (10.11), we get

$$I_S(f) = \frac{1}{6} \left(1 + \frac{8}{3} + \frac{1}{2} \right) = \frac{25}{36} \approx 0.694444.$$

Therefore, the error is $ME_S(f) \approx 0.001297$.

Composite Simpson's Rule

Let us now derive the *composite Simpson's rule*. First let us subdivide the interval [a, b] into 2n equal parts, where $n \in \mathbb{N}$. For $k = 0, 1, \dots, 2n$, define $x_k := a + kh$, where h = (b - a)/2n. Applying Simpson's rule (10.11) on the interval $[x_{2i}, x_{2i+2}]$, we get

$$\int_{x_{2i}}^{x_{2i+2}} f(x)dx \approx \frac{2h}{6} \left\{ f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2}) \right\}$$

Summing over $i = 0, \dots, n-1$, we get

$$\int_{a}^{b} f(x)dx = \sum_{i=0}^{n-1} \int_{x_{2i}}^{x_{2i+2}} f(x)dx$$

$$\approx \frac{2h}{6} \sum_{i=0}^{n-1} \left\{ f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2}) \right\}.$$

Therefore, the *composite Simpson's rule* takes the form

$$I_S^n(f) = \frac{h}{3} \left[f(x_0) + f(x_{2n}) + 2 \sum_{i=1}^{n-1} f(x_{2i}) + 4 \sum_{i=0}^{n-1} f(x_{2i+1}) \right].$$
 (6.12)

10.1.4 Method of Undetermined Coefficients

All the rules so far derived are of the form

$$I(f) = \int_{a}^{b} f(x)dx \approx w_0 f(x_0) + w_1 f(x_1) + \dots + w_n f(x_n)$$
 (10.13)

where w_i 's are weights. In deriving those rules, we have fixed the nodes x_0, x_1, \dots, x_n (n = 0 for rectangle rule, n = 1 for trapezoidal rule and n = 2 for Simpson's rule), and used interpolating polynomials to obtain the corresponding weights. Instead, we may use another approach in which for a fixed set of nodes weights are determined by imposing the condition that the resulting rule is exact for polynomials of degree less than or equal to n. Such a method is called the *method of undetermined coefficients*.

Example 10.1.7.

Let us find w_0 , w_1 , and w_2 such that the approximate formula

$$\int_{a}^{b} f(x) dx \approx w_0 f(a) + w_1 f\left(\frac{a+b}{2}\right) + w_2 f(b)$$
 (10.14)

is exact for all polynomials of degree less than or equal to 2.

Since integral of sum of functions is the sum of the integrals of the respective functions, the formula (10.14) is exact for all polynomials of degree less than or equal to 2 if and only if the formula (10.14) is exact for the polynomials 1, x, and x^2 .

• The condition that the formula (10.14) is exact for the polynomial p(x) = 1 yields

$$b - a = \int_{a}^{b} 1 \, dx = w_0 + w_1 + w_2$$

• The condition that the formula (10.14) is exact for the polynomial p(x) = x yields

$$\frac{b^2 - a^2}{2} = \int_{a}^{b} x \, dx = aw_0 + \left(\frac{a+b}{2}\right) w_1 + bw_2.$$

• The condition that the formula (10.14) is exact for the polynomial $p(x) = x^2$ yields

$$\frac{b^3 - a^3}{3} = \int_a^b x^2 dx = a^2 w_0 + \left(\frac{a+b}{2}\right)^2 w_1 + b^2 w_2$$

Thus, we have a linear system of three equations satisfied by w_0 , w_1 , and w_2 . By solving this system, we get

$$w_0 = \frac{1}{6}(b-a), \quad w_1 = \frac{2}{3}(b-a), \quad w_2 = \frac{1}{6}(b-a),$$

which gives us the familiar Simpson's rule.

This motivates the following definition.

Definition 10.1.8 [Degree of Precision]. The $degree\ of\ precision$ (also called $order\ of\ exactness$) of a quadrature formula is the largest positive integer n such that the formula is exact for all polynomials of

Example 10.1.9.

Let us determine the degree of precision of Simpson's rule. It will suffice to apply the rule over the interval [0, 2] (in fact any interval is good enough and we chose this interval for the sake of having easy computation).

$$\int_{0}^{2} dx = 2 = \frac{2}{6}(1+4+1),$$

$$\int_{0}^{2} x dx = 2 = \frac{2}{6}(0+4+2),$$

$$\int_{0}^{2} x^{2} dx = \frac{8}{3} = \frac{2}{6}(0+4+4),$$

$$\int_{0}^{2} x^{3} dx = 4 = \frac{2}{6}(0+4+8),$$

$$\int_{0}^{2} x^{4} dx = \frac{32}{5} \neq \frac{2}{6}(0+4+16) = \frac{20}{3}.$$

Therefore, the degree of precision of Simpson's rule is 3.

Remark 10.1.10.

In Example 10.1.7 we have obtained the Simpson's rule using the method of undetermined coefficients by requiring the exactness of the rule for polynomials of degree less than or equal to 2, the above example shows that the rule is exact for polynomials of degree three as well.

10.1.5 Gaussian Rules

In Example 10.1.7 we have fixed the nodes and obtained the weights in the quadrature rule (10.14) such that the rule is exact for polynomials of degree less than or equal to 2. In general, by fixing the nodes, we can obtain the weights in (10.13) such that the rule is exact for polynomials of degree less than or equal to n. But it is also possible to derive a quadrature rule such that the rule is exact for polynomials of degree less than or equal to 2n + 1 by choosing the n + 1 nodes and the weights appropriately. This is the basic idea of $Gaussian\ rules$.

Let us consider the special case

$$\int_{-1}^{1} f(x)dx \approx \sum_{i=0}^{n} w_i f(x_i).$$
 (10.15)

The weights w_i and the nodes x_i ($i = 0, \dots, n$) are to be chosen in such a way that the rule (10.15) is exact, that is

$$\int_{-1}^{1} f(x)dx = \sum_{i=0}^{n} w_i f(x_i), \qquad (10.16)$$

whenever f(x) is a polynomial of degree less than or equal to 2n+1. Note that (10.16) holds for every polynomial f(x) of degree less than or equal to 2n+1 if and only if (10.16) holds for $f(x) = 1, x, x^2, \dots, x^{2n+1}$.

Case 1: (n = 0). In this case, the quadrature formula (10.15) takes the form

$$\int_{-1}^{1} f(x)dx \approx w_0 f(x_0).$$

The condition (10.16) gives

$$\int_{-1}^{1} 1 \, dx = w_0 \text{ and } \int_{-1}^{1} x \, dx = w_0 x_0.$$

These conditions give $w_0 = 2$ and $x_0 = 0$. Thus, we have the formula

$$\int_{-1}^{1} f(x)dx \approx 2f(0) =: I_{G_0}(f), \tag{10.17}$$

which is the required Gaussian rule for n = 0.

Case 2: (n = 1). In this case, the quadrature formula (10.15) takes the form

$$\int_{-1}^{1} f(x)dx \approx w_0 f(x_0) + w_1 f(x_1).$$

The condition (10.16) gives

$$w_0 + w_1 = 2,$$

$$w_0 x_0 + w_1 x_1 = 0,$$

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

$$w_0 x_0^3 + w_1 x_1^3 = 0.$$

A solution of this nonlinear system is $w_0 = w_1 = 1$, $x_0 = -1/\sqrt{3}$ and $x_1 = 1/\sqrt{3}$. This lead to the formula

$$\int_{-1}^{1} f(x)dx \approx f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right) =: I_{G1}(f), \tag{10.18}$$

which is the required Gaussian rule for n = 1.

Case 3: (General). In general, the quadrature formula is given by (10.15), where there are 2(n+1) free parameters x_i and w_i for $i=0,1,\dots,n$. The condition (10.16) leads to the nonlinear system

$$\sum_{j=0}^{n} w_j x_j^i = \begin{cases} 0, & i = 1, 3, \dots, 2n+1 \\ \frac{2}{i+1}, & i = 0, 2, \dots, 2n \end{cases}.$$

These are nonlinear equations and their solvability is not at all obvious and therefore the discussion is outside the scope of this course.

So far, we derived Gaussian rule for integrals over [-1,1]. But this is not a limitation as any integral on the interval [a,b] can easily be transformed to an integral on [-1,1]

by using the linear change of variable

$$x = \frac{b+a+t(b-a)}{2}, \quad -1 \le t \le 1. \tag{10.19}$$

Thus, we have

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

Example 10.1.11.

We now use the Gaussian rule to approximate the integral

$$I(f) = \int_{0}^{1} f(x) \, dx,$$

where

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

Note that the true value is $I(f) = \log(2) \approx 0.693147$.

To use the Gaussian quadrature, we first need to make the linear change of variable (10.19) with a = 0 and b = 1 and we get

$$x = \frac{t+1}{2}, -1 \le t \le 1.$$

Thus the required integral is

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

We need to take f(t) = 1/(3+t) in the Gaussian quadrature formula (10.18) and we get

$$\int_{0}^{1} \frac{dx}{1+x} = \int_{-1}^{1} \frac{dt}{3+t} \approx f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right) \approx 0.692308 \approx I_{G1}(f).$$

Therefore, the error is $I(f) - I_{G1}(f) \approx 0.000839$.

Theorem 10.1.12.

Let f(x) be continous for $a \le x \le b$, and let $n \ge 1$. Then the absolute (mathematical) error $|ME_n(f)|$ in using Gaussian numerical integration rule to obtain I(f) satisfies

$$|ME_n(f)| \le 2(b-a)\rho_{2n+1}(f)$$
 (10.20)

where

$$\rho_{2n+1}(f) = \inf_{\deg q \le 2n+1} ||f - q||_{\infty}$$

is the minmax error of order 2n + 1 for f(x) on [a, b].

Proof.

 $\mathrm{ME}_n(p) = 0$ for any polynomial p(x) of degree $\leq 2n+1$. Also, the error function ME_n satisfies $\mathrm{ME}_n(F+G) = \mathrm{ME}_n(F) + \mathrm{ME}_n(G)$ for all $F, G \in C[a, b]$. Let $p(x) = q_{2n+1}^*(x)$, the minimax approximation of degree $\leq 2n+1$ to f(x) on [a, b]. Then

$$ME_n(f) = ME_n(f) - ME_n(q_{2n+1}^*)
= ME_n(f - q_{2n+1}^*)
= \int_a^b (f(x) - q_{2n+1}^*(x)) dx - \sum_{j=0}^n w_j (f(x_j) - q_{2n+1}^*(x_j)).$$

Therefore, we have

$$|\mathrm{ME}_n(f)| \le ||f - q_{2n+1}^*||_{\infty} [(b-a) + \sum_{j=0}^n |w_j|].$$

But we have $\sum_{j=0}^{n} |w_j| = b - a$ and therefore, we get the desired result.

10.2 Numerical Differentiation

The aim of this section is to obtain formulas to approximate the values of derivatives of a given function at a given point. Such a formula for the first derivative of a function can be obtained directly using the definition of the derivative, namely, the difference quotients of the function. This will be discussed in Subsection 10.2.1. But this idea cannot be adopted for higher order derivatives. Approximating formulas for derivatives can be obtained in at least two ways, namely,

- 1. Methods based on Interpolation
- 2. Methods based on Undetermined Coefficients

These methods are discussed in Subsection 10.2.2 and 10.2.3, respectively.

10.2.1 Approximations of First Derivative

Forward Difference Formula

The most simple way to obtain a numerical method for approximating the derivative of a C^1 function f is to use the definition of derivative

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

The approximating formula can therefore be taken as

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} =: D_h^+ f(x)$$
 (10.21)

for a sufficiently small value of h > 0. The formula $D_h^+ f(x)$ is called the *forward difference formula* for the derivative of f at the point x.

Theorem 10.2.1.

Let $f \in C^2[a, b]$. The mathematical error in the forward difference formula is given by

$$f'(x) - D_h^+ f(x) = -\frac{h}{2} f''(\eta)$$
 (10.22)

for some $\eta \in (x, x + h)$.

Proof.

By Taylor's theorem, we have

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(\eta)$$
 (10.23)

for some $\eta \in (x, x + h)$. Using (10.21) and (10.23), we obtain

$$D_h^+ f(x) = \frac{1}{h} \left\{ \left[f(x) + hf'(x) + \frac{h^2}{2} f''(\eta) \right] - f(x) \right\} = f'(x) + \frac{h}{2} f''(\eta).$$

This completes the proof.

Remark 10.2.2.

If we consider the left hand side of (10.22) as a function of h, *i.e.*, if

$$g(h) = f'(x) - D_h f(x),$$

then we see that

$$\left| \frac{g(h)}{h} \right| = \frac{1}{2} |f''(\eta)|.$$

Let M>0 be such that $|f''(x)|\leq M$ for all $x\in[a,b]$. Then we see that

$$\left| \frac{g(h)}{h} \right| \le \frac{M}{2}.$$

That is, g = O(h) as $h \to 0$. We say that the forward difference formula $D_h^+ f(x)$ is of order 1 (order of accuracy).

Backward Difference Formula

The derivative of a function f is also given by

$$f'(x) = \lim_{h \to 0} \frac{f(x) - f(x - h)}{h}.$$

Therefore, the approximating formula for the first derivative of f can also be taken as

$$f'(x) \approx \frac{f(x) - f(x - h)}{h} =: D_h^- f(x)$$
 (10.24)

The formula $D_h^-f(x)$ is called the backward difference formula for the derivative of f at the point x.

Deriving the mathematical error for backward difference formula is similar to that of the forward difference formula. It can be shown that the backward difference formula is of order 1.

Central Difference Formula

The derivative of a function f is also given by

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}.$$

Therefore, the approximating formula for the first derivative of f can also be taken as

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h} =: D_h^0 f(x),$$
 (10.25)

for a sufficiently small value of h > 0. The formula $D_h^0 f(x)$ is called the *central difference* formula for the derivative of f at the point x.

The central difference formula is of order 2 as shown in the following theorem.

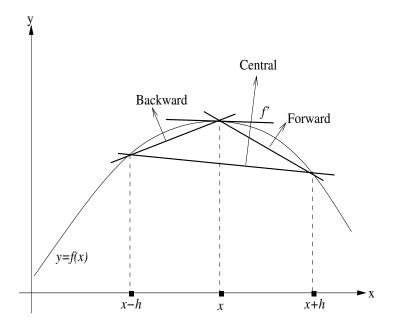


Figure 10.4: Geometrical interpretation of difference formulae.

Theorem 10.2.3.

Let $f \in C^3[a,b]$. The mathematical error in the central difference formula is given by

$$f'(x) - D_h^0 f(x) = -\frac{h^2}{6} f'''(\eta), \qquad (10.26)$$

where $\eta \in (x - h, x + h)$.

Proof.

Using Taylor's theorem, we have

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(\eta_1)$$

and

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2!}f''(x) - \frac{h^3}{3!}f'''(\eta_2),$$

where $\eta_1 \in (x, x + h)$ and $\eta_2 \in (x - h, x)$.

Therefore, we have

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{h^3}{3!}(f'''(\eta_1) + f'''(\eta_2)).$$

Since f'''(x) is continuous, by intermediate value theorem applied to f'', we have

$$\frac{f'''(\eta_1) + f'''(\eta_2)}{2} = f'''(\eta)$$

for some $\eta \in (x - h, x + h)$. This completes the proof.

Geometric interpretation of the three primitive difference formulae (forward, backward, and central) is shown in Figure 10.4.

Example 10.2.4.

To find the value of the derivative of the function given by $f(x) = \sin x$ at x = 1 with h = 0.003906, we use the three primitive difference formulas. We have

$$f(x-h) = f(0.996094) = 0.839354,$$

 $f(x) = f(1) = 0.841471,$
 $f(x+h) = f(1.003906) = 0.843575.$

- 1. Backward difference: $D_h^- f(x) = \frac{f(x) f(x-h)}{h} = 0.541935$.
- 2. Central Difference: $D_h^0 f(x) = \frac{f(x+h) f(x-h)}{2h} = 0.540303.$
- 3. Forward Difference: $D_h^+ f(x) = \frac{f(x+h) f(x)}{h} = 0.538670.$

Note that the exact value is $f'(1) = \cos 1 = 0.540302$.

10.2.2 Methods based on Interpolation

Using the polynomial interpolation we can obtain formula for derivatives of any order for a given function. For instance, to calculate f'(x) at some point x, we use the approximate formula

$$f'(x) \approx p'_n(x),$$

where $p_n(x)$ denotes the interpolating polynomial for f(x). Many formulas can be obtained by varying n and by varying the placement of the nodes x_0, \dots, x_n relative to the point x of interest.

Let us take n = 1. The linear interpolating polynomial is given by

$$p_1(x) = f(x_0) + f[x_0, x_1](x - x_0).$$

Hence, we have the formula

$$f'(x) \approx p'_1(x) = f[x_0, x_1].$$
 (10.27)

In particular,

- if we take $x_0 = x$ and $x_1 = x + h$ for a small value h > 0, we obtain the forward difference formula $D_h^+ f(x)$.
- if we take $x_0 = x h$ and $x_1 = x$ for small value of h > 0, we obtain the backward difference formula $D_h^- f(x)$.
- if we take $x_0 = x h$ and $x_1 = x + h$ for small value of h > 0, we get the central difference formula $D_h^0 f(x)$.

We next prove the formula for mathematical error in approximating the first derivative using interpolating polynomial.

Theorem 10.2.5 [Mathematical Error].

Hypothesis:

- 1. Let f be an (n+2)-times continuously differentiable function on the interval [a,b].
- 2. Let x_0, x_1, \dots, x_n be n+1 distinct nodes in [a, b].
- 3. Let $p_n(x)$ denote the polynomial that interpolates f at the nodes x_0, x_1, \dots, x_n .
- 4. Let x be any point in [a, b] such that $x \notin \{x_0, x_1 \cdots, x_n\}$.

Conclusion: Then

$$f'(x) - p'_n(x) = w_n(x) \frac{f^{(n+2)}(\eta_x)}{(n+2)!} + w'_n(x) \frac{f^{(n+1)}(\xi_x)}{(n+1)!}$$
(10.28)

with $w_n(x) = \prod_{i=0}^n (x - x_i)$ and ξ_x and η_x are points in between the maximum and minimum of x_0, x_1, \dots, x_n and x, that depend on x.

The proof of the above theorem is omitted for this course.

Difference formulas for higher order derivatives and their mathematical error can be obtained similarly. The derivation of the mathematical error for the formulas of higher order derivatives are omitted for this course.

Example 10.2.6.

Let x_0 , x_1 , and x_2 be the given nodes. Then, the Newton's form of interpolating polynomial for f is given by

$$p_2(x) = f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x^2 - (x_0 + x_1)x + x_0x_1).$$

Therefore, we take the first derivative of f as

$$f'(x) \approx p_2'(x) = f[x_0, x_1] + f[x_0, x_1, x_2](2x - x_0 - x_1).$$

- If we take $x_0 = x h$, $x_1 = x$, and $x_2 = x + h$ for any given $x \in [a, b]$, we obtain the central difference formula $D_h^0(f)$ and the corresponding error obtained from (10.28) is precisely the error given in (10.26).
- If we take $x_0 = x$, $x_1 = x + h$ and $x_2 = x + 2h$ for any given $x \in [a, b]$, we obtain the difference formula

$$f'(x) \approx \frac{-3f(x) + 4f(x+h) - f(x+2h)}{2h}$$

with mathematical error obtained using (10.28) as

$$f'(x) - p_2'(x) = \frac{h^2}{3}f'''(\xi),$$

for some $\xi \in (x, x + 2h)$.

10.2.3 Methods based on Undetermined Coefficients

Another method to derive formulas for numerical differentiation is called the *method of undetermined coefficients*. The idea behind this method is similar to the one discussed in deriving quadrature formulas.

Suppose we seek a formula for $f^{(k)}(x)$ that involves the nodes x_0, x_1, \dots, x_n . Then, write the formula in the form

$$f^{(k)}(x) \approx w_0 f(x_0) + w_1 f(x_1) + \dots + w_n f(x_n)$$
(10.29)

where w_i , $i = 0, 1, \dots, n$ are free variables that are obtained by imposing the condition that this formula is exact for polynomials of degree less than or equal to n.

Example 10.2.7.

We will illustrate the method by deriving the formula for f''(x) at nodes $x_0 = x - h$, $x_1 = x$ and $x_2 = x + h$ for a small value of h > 0.

For a small value of h > 0, let

$$f''(x) \approx D_h^{(2)} f(x) := w_0 f(x - h) + w_1 f(x) + w_2 f(x + h)$$
(10.30)

where w_0 , w_1 and w_2 are to be obtained so that this formula is exact when f(x) is a polynomial of degree less than or equal to 2. This condition is equivalent to the exactness for the three polynomials 1, x and x^2 .

Step 1: When f(x) = 1 for all x. Then the formula of the form (10.30) is assumed to be exact and we get

$$w_0 + w_1 + w_2 = 0. (10.31)$$

Step 2: When f(x) = x for all x. Then the formula of the form (10.30) is assumed to be exact and we get

$$w_0(x-h) + w_1x + w_2(x+h) = 0.$$

Using (10.31), we get

$$w_2 - w_0 = 0. (10.32)$$

Step 3: When $f(x) = x^2$ for all x. Then the formula of the form (10.30) is assumed to be exact and we get

$$w_0(x-h)^2 + w_1x^2 + w_2(x+h)^2 = 2.$$

Using (10.31) and (10.32), we get

$$w_0 + w_2 = \frac{2}{h^2}. (10.33)$$

Solving the linear system of equations (10.31), (10.32), and (10.33), we get

$$w_0 = w_2 = \frac{1}{h^2}, \quad w_1 = -\frac{2}{h^2}.$$

Substituting these into (10.30), we get

$$D_h^{(2)}f(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2},$$
(10.34)

which is the required formula.

Let us now derive the mathematical error involved in this formula. For this, we use the Taylor's series

$$f(x \pm h) = f(x) \pm hf'(x) + \frac{h^2}{2!}f''(x) \pm \frac{h^3}{3!}f^{(3)}(x) + \cdots$$

in (10.34) to get

$$D_h^{(2)}f(x) = \frac{1}{h^2} \left(f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f^{(3)}(x) + \frac{h^4}{4!}f^{(4)}(x) + \cdots \right) - \frac{2}{h^2}f(x) + \frac{1}{h^2} \left(f(x) - hf'(x) + \frac{h^2}{2!}f''(x) - \frac{h^3}{3!}f^{(3)}(x) + \frac{h^4}{4!}f^{(4)}(x) - \cdots \right).$$

After simplification, we get

$$D_h^{(2)}f(x) = f''(x) + \frac{h^2}{24}[(f^{(4)}(x) + \cdots) + (f^{(4)}(x) - \cdots)].$$

Now treating the fourth order terms on the right hand side as remainders in Taylor's series, we get

$$D_h^{(2)}f(x) = f''(x) + \frac{h^2}{24}[f^{(4)}(\xi_1) + f^{(4)}(\xi_2)],$$

for some $\xi_1, \xi_2 \in (x - h, x + h)$. Using intermediate value theorem for the function $f^{(4)}$, we get the mathematical error as

$$f''(x) - D_h^{(2)}f(x) = -\frac{h^2}{12}f^{(4)}(\xi)$$
 (10.35)

for some $\xi \in (x - h, x + h)$, which is the required mathematical error.

10.2.4 Arithmetic Error in Numerical Differentiation

Difference formulas are useful when deriving methods for solving differential equations. But they can lead to serious errors when applied to function values that are subjected to floating-point approximations. Let

$$f(x_i) = f_i + \epsilon_i, \quad i = 0, 1, 2.$$

To illustrate the effect of such errors, we choose the approximation $D_h^{(2)}f(x)$ given by (10.34) for the second derivative of f with $x_0 = x - h$, $x_1 = x$ and $x_2 = x + h$. Instead of using the exact values $f(x_i)$, we use the approximate values f_i in the difference formula (10.34). That is,

$$\bar{D}_h^{(2)}f(x_1) = \frac{f_2 - 2f_1 + f_0}{h^2}.$$

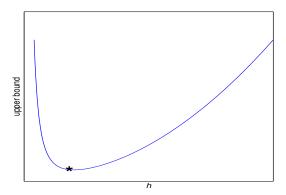


Figure 10.5: A sketch of the upper bound in total error as given in (10.36) as a function of h. The black star indicates the optimal value of h.

The total error committed is

$$f''(x_1) - \bar{D}_h^{(2)} f(x_1) = f''(x_1) - \frac{f(x_2) - 2f(x_1) + f(x_0)}{h^2} + \frac{\epsilon_2 - 2\epsilon_1 + \epsilon_0}{h^2}$$
$$= -\frac{h^2}{12} f^{(4)}(\xi) + \frac{\epsilon_2 - 2\epsilon_1 + \epsilon_0}{h^2}.$$

Using the notation $\epsilon_{\infty} := \max\{|\epsilon_0|, |\epsilon_1|, |\epsilon_3|\}$, we have

$$|f''(x_1) - \bar{D}_h^{(2)}f(x_1)| \le \frac{h^2}{12}|f^{(4)}(\xi)| + \frac{4\epsilon_\infty}{h^2}.$$
 (10.36)

The error bound in (10.36) clearly shows that, although the first term (bound of mathematical error) tends to zero as $h \to 0$, the second term (bound of arithmetic error) can tend to infinity as $h \to 0$. This gives a possibility for the total error to be as large as possible when $h \to 0$. In fact, there is an optimal value of h to minimize the right side of (10.36) (as shown in Figure 10.5), which we will illustrate in the following example.

Example 10.2.8.

In finding $f''(\pi/6)$ for the function $f(x) = \cos x$, if we use the function values f_i that has six significant digits when compared to $f(x_i)$, then

$$\frac{|f(x_i) - f_i|}{|f(x_i)|} \le 0.5 \times 10^{-5}.$$

Since $|f(x_i)| = |\cos(x_i)| \le 1$, we have $|f(x_i) - f_i| \le 0.5 \times 10^{-5}$.

We now use the formula $\bar{D}_h^{(2)}f(x)$ to approximate f''(x). Assume that other than the approximation in the function values, the formula $\bar{D}_h^{(2)}f(x)$ is calculated exactly.

Then the bound for the absolute value of the total error given by (10.36) takes the form

$$|f''(\pi/6) - \bar{D}_h^{(2)}f(\pi/6)| \le \frac{h^2}{12}|f^{(4)}(\xi)| + \frac{4\epsilon_\infty}{h^2},$$

where $\epsilon_{\infty} \leq 0.5 \times 10^{-5}$ and $\xi \approx \pi/6$. Thus, we have

$$|f''(\pi/6) - \bar{D}_h^{(2)}f(\pi/6)| \leq \frac{h^2}{12}\cos\left(\frac{\pi}{6}\right) + \frac{4}{h^2}(0.5 \times 10^{-5}) \approx 0.0722h^2 + \frac{2 \times 10^{-5}}{h^2} =: E(h).$$

The bound E(h) indicates that there is a smallest value of h, call it h^* , such that the bound increases rapidly for $0 < h < h^*$ when $h \to 0$. To find it, let E'(h) = 0, with its root being h^* . This leads to $h^* \approx 0.129$. Thus, for close values of $h > h^* \approx 0.129$, we have less error bound than the values $0 < h < h^*$. This behavior of E(h) is observed in the following table. Note that the true value is $f''(\pi/6) = -\cos(\pi/6) \approx -0.86603$.

h	$\bar{D}_h^{(2)} f(\pi/6)$	Total Error	E(h)
0.2	-0.86313	-0.0029	0.0034
0.129	-0.86479	-0.0012	0.0024
0.005	-0.80000	-0.0660	0.8000
0.001	0.00000	-0.8660	20

When h is very small, f(x - h), f(x) and f(x + h) are very close numbers and therefore their difference in the numerator of the formula (10.34) tend to have loss of significance. This is clearly observed in the values of $\bar{D}_h^{(2)} f(\pi/6)$ when compared to the true value where, we are not loosing much number of significant digits for h > 0.129, whereas for h < 0.129, there is a drastic loss of significant digits.

10.3 Exercises

Numerical Integration

1. Apply Rectangle, Trapezoidal, Simpson and Gaussian methods to evaluate

i)
$$I = \int_{0}^{\pi/2} \frac{\cos x}{1 + \cos^{2} x} dx$$
 (exact value ≈ 0.623225)

ii)
$$I = \int_{0}^{\pi} \frac{dx}{5 + 4\cos x} dx$$
 (exact value ≈ 1.047198)

iii)
$$I = \int_{0}^{1} e^{-x^{2}} dx$$
 (exact value ≈ 0.746824),

iv)
$$I = \int_{0}^{\pi} \sin^3 x \cos^4 x \, dx$$
 (exact value ≈ 0.114286)

v)
$$I = \int_{0}^{1} (1 + e^{-x} \sin(4x)) dx$$
. (exact value ≈ 1.308250)

Compute the relative error (when compared to the given exact values) in each method.

2. Write down the errors in the approximation of

$$\int_{0}^{1} x^{4} dx \quad \text{and} \quad \int_{0}^{1} x^{5} dx$$

by the Trapezoidal rule and Simpson's rule. Find the value of the constant C for which the Trapezoidal rule gives the exact result for the calculation of the integral

$$\int_{0}^{1} (x^5 - Cx^4) dx.$$

3. A function f has the values shown below:

- i) Use trapezoidal rule to approximate $\int_1^2 f(x) dx$.
- ii) Use Simpson's rule to approximate $\int_1^2 f(x) dx$.
- iii) Use composite Simpson's rule to approximate $\int_1^2 f(x) dx$.
- 4. Obtain expressions for the arithmetic error in approximating the integral $\int_a^b f(x) dx$ using the trapezoidal and the Simpson's rules. Also obtain upper bounds.
- 5. Let $a = x_0 < x_1 < \cdots < x_n = b$ be equally spaced nodes (i.e., $x_k = x_0 + kh$ for $k = 1, 2, \dots, n$) in the interval [a, b]. Note that h = (b a)/n. Let f be a twice continuously differentiable function on [a, b].
 - i) Show that the expression for the mathematical error in approximating the integral $\int_a^b f(x) dx$ using the composite trapezoidal rule, denoted by $E_T^n(f)$, is given by

$$E_T^n(f) = -\frac{(b-a)h^2}{12}f''(\xi),$$

for some $\xi \in (a, b)$.

ii) Show that the mathematical error $E_T^n(f)$ tends to zero as $n \to \infty$ (one uses the terminology *composite trapezoidal rule is convergent* in such a case).

- 6. Determine the minimum number of subintervals and the corresponding step size h so that the error for the composite trapezoidal rule is less than 5×10^{-9} for approximating the integral $\int_2^7 dx/x$.
- 7. Let $a = x_0 < x_1 < \cdots < x_n = b$ be equally spaced nodes (i.e., $x_k = x_0 + kh$ for $k = 1, 2, \dots, n$) in the interval [a, b], and n is an even natural number. Note that h = (b-a)/n. Let f be a four times continuously differentiable function on [a, b].
 - i) Show that the expression for the mathematical error in approximating the integral $\int_a^b f(x) dx$ using the composite Simpson rule, denoted by $E_S^n(f)$, is given by

$$E_S^n(f) = -\frac{(b-a)h^4}{180}f^{(4)}(\xi),$$

for some $\xi \in (a, b)$.

- ii) Show that the mathematical error $E_S^n(f)$ tends to zero as $n \to \infty$ (one uses the terminology *composite Simpson rule is convergent* in such a case).
- 8. Use composite Simpson's and composite Trapezoidal rules to obtain an approximate value for the improper integral

$$\int_{1}^{\infty} \frac{1}{x^2 + 9} dx, \text{ with } n = 4.$$

9. Determine the coefficients in the quadrature formula

$$\int_{0}^{2h} x^{-1/2} f(x) dx \approx (2h)^{1/2} (w_0 f(0) + w_1 f(h) + w_2 f(2h))$$

such that the formula is exact for all polynomials of degree as high as possible. What is the degree of precision?

- 10. Use the two-point Gaussian quadrature rule to approximate $\int_{-1}^{1} dx/(x+2)$ and compare the result with the trapezoidal and Simpson's rules.
- 11. Assume that $x_k = x_0 + kh$ are equally spaced nodes. The quadrature formula

$$\int_{x_0}^{x_3} f(x)dx \approx \frac{3h}{8} \left(f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3) \right)$$

is called the Simpson's $\frac{3}{8}$ rule. Determine the degree of precision of Simpson's $\frac{3}{8}$ rule.

Numerical Differentiation

12. In this problem, perform the calculations using 6-digit rounding arithmetic.

- i) Find the value of the derivative of the function $f(x) = \sin x$ at x = 1 using the forward, backward, and central difference formulae with $h_1 = 0.015625$, and $h_2 = 0.000015$.
- ii) Find f'(1) directly and compare with the values obtained for each h_i (i =1, 2).
- 13. Obtain the central difference formula for f'(x) using polynomial interpolation with nodes at x - h, x, x + h, where h > 0.
- 14. Given the values of the function $f(x) = \ln x$ at $x_0 = 2.0$, $x_1 = 2.2$ and $x_2 =$ 2.6, find the approximate value of f'(2.0) using the method based on quadratic interpolation. Obtain an error bound.
- 15. The following data corresponds to the function $f(x) = \sin x$.

$$\begin{array}{c|cccc} x & 0.5 & 0.6 & 0.7 \\ \hline f(x) & 0.4794 & 0.5646 & 0.6442 \end{array}$$

Obtain the approximate value of f'(0.5), f'(0.6), and f'(0.7) using forward, backward, and central difference formulae whichever are applicable. Compute the relative error in all the three cases.

16. The following data corresponds to the function $f(x) = e^x - 2x^2 + 3x + 1$. $\frac{x \mid 0.0 \mid 0.2 \mid 0.4}{f(x) \mid 0.0 \mid 0.7414 \mid 1.3718}$

Obtain the approximate value of f'(0.0), f'(0.2), and f'(0.4) using forward, backward, and central difference formulae whichever are applicable. Compute the relative error in all the three cases.

- 17. Obtain expressions for the arithmetic error in approximating the first derivative of a function using the forward, backward, and central difference formulae.
- 18. Find an approximation to f'(x) as a formula involving f(x), f(x+h), and f(x+2h). Obtain an expression for the mathematical error involved in this approximation.
- 19. Let h > 0. Use the method of undetermined coefficients to find a numerical differentiation formula for approximating f''(x) such that the formula uses values of the function f at each of the following sets of points:
 - i) x + 2h, x + h and x.
 - ii) x + 3h, x + 2h x + h and x.

Obtain expressions for mathematical error in both the cases.

20. Show that the formula

$$D^{(2)}f(x) := \frac{f(x) - 2f(x-h) + f(x-2h)}{h^2}$$

gives approximate value for f''(x). Find the order of accuracy of this formula.

21. For the method

$$f'(x) \approx \frac{4f(x+h) - f(x+2h) - 3f(x)}{2h},$$

obtain an expression for mathematical error, arithmetic error, and hence total error. Find a bound on the absolute value of the total error as function of h. Determine the optimal value of h for which the bound obtained is minimum.

- 22. Repeat the previous problem when central difference formula is used for numerical differentiation.
- 23. Let $f(x) = \ln(x)$ (here ln denotes the natural logarithm). Give the formula for approximating f'(x) using central difference formula. When we use this formula to get an approximate value of f'(1.5) with the assumption that the function values f(1.5-h) and f(1.5+h) are rounded to 2 digits after decimal point, find the value of h such that the total error is minimized. (Final Exam, Autumn 2010 (M.Sc.))
- 24. The voltage E = E(t) in an electrical circuit obeys the equation

$$E(t) = L\left(\frac{dI}{dt}\right) + RI(t),$$

where R is resistance and L is inductance. Use L = 0.05 and R = 2 and values for I(t) in the table following.

Find I'(1.2) using (i) central difference formula, and (ii) the formula given in Problem (21) and use it to compute E(1.2). Compare your answer with $I(t) = 10e^{-t/10}\sin(2t)$.