Computational Methods

Approximation of Derivatives & Difference Notation

Forward Scheme:

$$\frac{\partial}{\partial S} \equiv D; \ \frac{\partial^2}{\partial S^2} \equiv D^2$$

If $\Delta-$ forward difference operator, consider a forward difference

$$\Delta V(S) = V(S+h) - V(S) \qquad \text{(a)}$$

Using Taylor on $S+h$, $V(S+h) =$

$$\begin{aligned}
& V(S) + hV_S + \tfrac{1}{2!}h^2 V_{SS} + O\left(h^3\right) \\
= \ & V(S) + hV_S + \tfrac{1}{2!}h^2 V_{SS} + ... \\
= \ & V(S) + hDV + \tfrac{1}{2!}h^2 D^2 V + ... \\
= \ & \underbrace{\left(1 + hD + \tfrac{1}{2!}h^2 D^2 + ...\right)}_{=\exp(hD)} V(S) \\
= \ & \underbrace{e^{hD}}_{\text{operator - acts on functions}} \times V(S)
\end{aligned}$$

Therefore $V(S+h) = e^{hD} V(S)$.

Then from (a) :

$$
\begin{aligned}
\Delta V(S) &= V(S+h) - V(S) \\
&= e^{hD} V(S) - V(S) \\
&= \left(e^{hD} - \mathbf{1}\right) V(S)
\end{aligned}
$$

where $\mathbf{1}$ is the identity operator.

So the *forward operator* becomes

$$
\Delta \equiv e^{hD} - \mathbf{1} \tag{b}
$$

Rearranging (b) gives:

$$
D = \frac{1}{h} \log(1 + \Delta). \tag{c}
$$

We know the TSE for $\log(1+x) =$

$$
x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + O\left(x^5\right)
$$

So

$$D=\frac{1}{h}\left(\Delta-\frac{\Delta^{2}}{2}+\frac{\Delta^{3}}{3}-\frac{\Delta^{4}}{4}+...\right) \tag{d}$$

This is the **first derivative** $\dfrac{\partial}{\partial S}$.
First Order Approximation:

$$D\approx\frac{\Delta}{h}$$

Hence

$$DV(S)\approx\frac{\Delta}{h}V(S)$$

which gives the Euler approximation for the first derivative

$$\frac{\partial V}{\partial S}\approx\frac{V(S+h)-V(S)}{h}.$$

Second Order Approximation:

Go back to (d) to get

$$D = \frac{1}{h}\left(\Delta - \frac{\Delta^2}{2}\right) \tag{e}$$

so operating on $V(S)$ gives

$$DV(S) = \frac{1}{h}\left(\Delta V - \frac{\Delta^2}{2}V\right)$$

where we can define $\Delta^2 V = \Delta(\Delta V)$

$$\begin{aligned} &= \Delta(V(S+h) - V(S)) \text{ from } (a) \\ &= \Delta V(S+h) - \Delta V(S) \end{aligned}$$

performing (a) on $V(S+h)$ and $V(S)$ gives:

$$\{V(S+2h) - V(S+h)\} - \{V(S+h) - V(S)\}$$

hence

$$\Delta^2 V(S) = V(S+2h) - 2V(S+h) + V(S). \tag{f}$$

So we can go back to (e) :

$$\begin{aligned}
\frac{\partial V}{\partial S} &= DV \approx \frac{1}{h}\left(\Delta V - \frac{\Delta^2 V}{2}\right) \\
&= \frac{1}{h}\left((V(S+h) - V(S)) - \frac{1}{2}(\mathsf{f})\right).
\end{aligned}$$

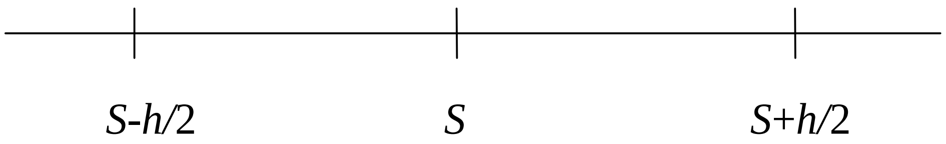
Simplifying gives:

$$\frac{\partial V}{\partial S} \approx \frac{1}{2h}\left(4V(S+h) - V(S+2h) - 3V(S)\right).$$

Note: Taking h to the l.h.s and setting $h = 0$ gives $4V(S) - V(S) - 3V(S) = 0$

Central Difference:

$$\delta V(S) = V(S + h/2) - V(S - h/2)$$



Using the earlier relationship $V(S+h) = e^{hD}V(S)$, the above becomes

$$\delta V(S) = \left(e^{\frac{h}{2}D} - e^{-\frac{h}{2}D}\right)V(S)$$

where the δ on the lhs and $\left(e^{\frac{h}{2}D} - e^{-\frac{h}{2}D}\right)$ on the rhs are operators, hence

$$\begin{aligned}\delta &\equiv e^{\frac{h}{2}D} - e^{-\frac{h}{2}D}\\ &= 2.\frac{e^{\frac{h}{2}D} - e^{-\frac{h}{2}D}}{2}\\ &= 2\sinh\left(\frac{h}{2}D\right)\end{aligned}$$

therefore

$$\begin{aligned}\sinh\left(\frac{h}{2}D\right) &= \delta/2\\ &\Longrightarrow D = \frac{2}{h}\sinh^{-1}(\delta/2)\end{aligned}$$

Using Taylor for $\sinh^{-1}$

$$D = \tfrac{1}{h}\left(\delta - \frac{\delta^3}{24} + \frac{3}{640}\delta^5 + O\left(\delta^7\right)\right) \tag{g}$$

1st Order Approximation:

$$D \approx \frac{\delta}{h}$$

3rd Order Approximation:

$$D \approx \frac{1}{h}\left(\delta - \frac{\delta^3}{24}\right)$$

Due to the gamma term in the BSE we need to approximate $\dfrac{\partial^2 V}{\partial S^2}$, so square

(g)

$$
\begin{aligned}
D^2 &= \frac{1}{h^2}\left(\delta^2 + \frac{\delta^6}{24^2} - \frac{\delta^4}{12} + \frac{6}{640}\delta^6 + O\left(\delta^8\right)\right) \\
&= \frac{1}{h^2}\left(\delta^2 - \frac{\delta^4}{12} + \frac{\delta^6}{90} + O\left(\delta^8\right)\right)
\end{aligned}
$$

So what is δ^2?

$$
\begin{aligned}
\delta^2 V(S) &= \delta\delta V(S) \\
&= \delta\left(\delta V(S)\right) \\
&= \delta\left(V(S + h/2) - V(S - h/2)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \delta V(S + h/2) - \delta V(S - h/2) \\
&= \left(V(S+h) - V(S)\right) - \left(V(S) - V(S-h)\right) \\
&= V(S+h) - 2V(S) + V(S-h)
\end{aligned}
$$

Therefore

$$
\frac{\partial^2 V}{\partial S^2} = \frac{1}{h^2}\left(V(S+h) - 2V(S) + V(S-h)\right)
$$

and we can write

$$D^2 \approx \frac{\delta^2}{h^2}$$

Finding Roots

A fundamental problem in numerical analysis consists of obtaining the zero of a function. Given a function $y = f(x)$ obtain the root of $f(x) = 0$, i.e. find the value of $x = c$ which satisfies $f(c) = 0$. e.g.

$$f(x) = x - \sin x$$

or another example:- zero's of the polynomial

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots.. + a_1 x + a_2 x_2^2 + \ldots.. + a_0$$

Here the most basic problem is solve $ax^2 + bx + c = 0$.

Consider the numerical solution of problems of this type in 4 sections:-

(i) Methods which do not use derivatives of the function

(ii) Methods which do use $f'(x)$

(iii) Methods for polynomials

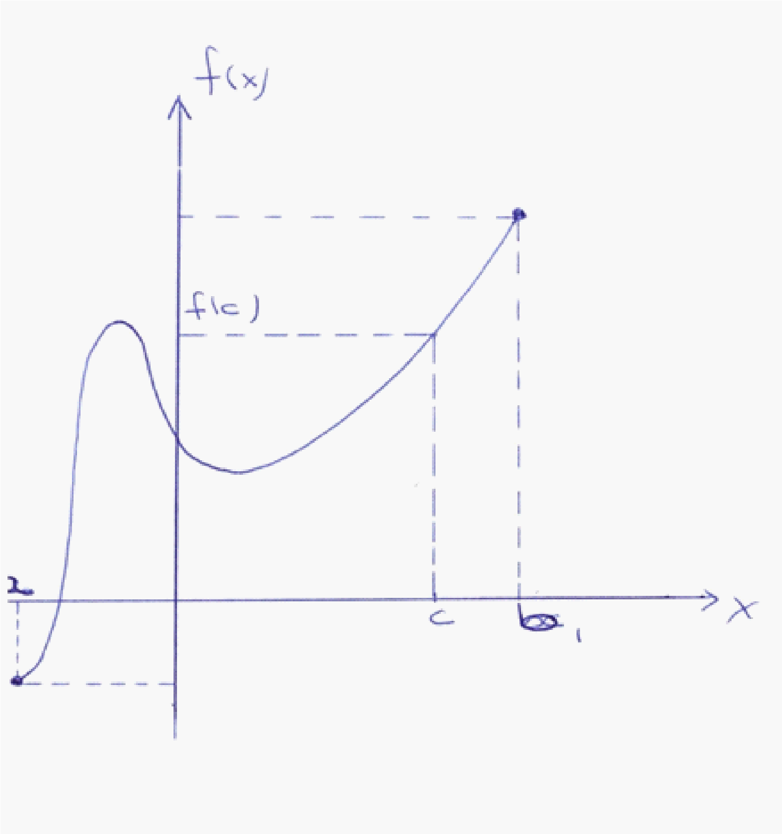
(iv) Methods which deal with complex roots

Bisection

The simplest method is that of bisection. The following theorem, from calculus class, insures the success of the method.

Intermediate Value Theorem (Theorem 1)

Suppose $f(x)$ is continuous on $[a,\ b]$ then for any y s.t y is between $f(a)$ and $f(b)$ there $\exists\ c \in [a,\ b]$ s.t $f(c) = y$.



Example 1

The function $f(x) = \dfrac{1}{x}$ is not continuous at 0. Thus if $0 \in [a,\ b]$, we *cannot* apply the IVT. In particular, if $0 \in [a,\ b]$ it happens to be the case that for *every* y between $f(a),\ f(b)$ there is no $c \in [a,\ b]$ such that $f(c) = y$.

In particular, the IVT tells us that if $f(x)$ is continuous and we know $a,\ b$ such that $f(a),\ f(b)$ have different sign, then there is some root in $[a,\ b]$. This is a fundamental test we can apply.

Example: Show that the function $g(x) = x^3 + 2x^2 + 5x - 1$ has a root lying between 0 and 1.

We note $f(0) = -1; f(1) = 7$. The sign change confirm that $\exists$ a root α s.t. $\alpha \in (0, 1)$.

Once location of a root α is established then a reasonable estimate of α is $c = \dfrac{a+b}{2}$. We can check whether $f(c) = 0$. If this does not hold then one and only one of the two following options holds:

1. $f(a),\ f(c)$ have different signs.

2. $f(c)$, $f(b)$ have different signs.

We now choose to recursively apply bisection to either $[a, c]$ or $[c, b]$, respectively, depending on which of these two options hold.

Modifications

It is not possible for a computer to test whether a given black box function is continuous. Thus malicious or incompetent users could cause a naively implemented bisection algorithm to fail. There are a number of easily conceivable problems:

1. The user might give f, a, b such that $f(a)$, $f(b)$ have the same sign. In this case the function f might be continuous, and might have a root in the interval $[a, b]$. If, taking $c = \dfrac{a+b}{2}$, $f(a)$, $f(b)$, $f(c)$ all have the

same sign, the algorithm would be at a stalemate!. Check on the input to make sure $f(a)$, $f(b)$ have different signs.

2. The user might give f, a, b such that f is not continuous on $[a, b]$, moreover has no root in the interval $[a, b]$. For a poorly implemented algorithm, this might lead to an infinite search on smaller and smaller intervals about some discontinuity of f. The algorithm might descend to intervals as small as machine precision, in which case the midpoint of the interval will, due to rounding, be the same as one of the endpoints, resulting in an infinite recursion.

Another common error occurs in the testing of the signs of $f(a)$, $f(b)$. A competent programmer would implement this with the following logical statement:

$$\text{if } (f(a)\, f(b) > 0) \text{ then....}$$

Note however, that $|f(a)|$, $|f(b)|$ might be very small, and that $f(a)\, f(b)$ might be too small to be representable in the computer; this calculation would be rounded to zero, and unpredictable behaviour would ensue. A wiser choice is

$$\text{if } (sign\,(f(a)) * sign\,(f(b)) > 0) \text{ then....}$$

where the function sign(x) returns $-1,\ 0,\ 1$ depending on whether x is negative, zero, or positive, respectively.

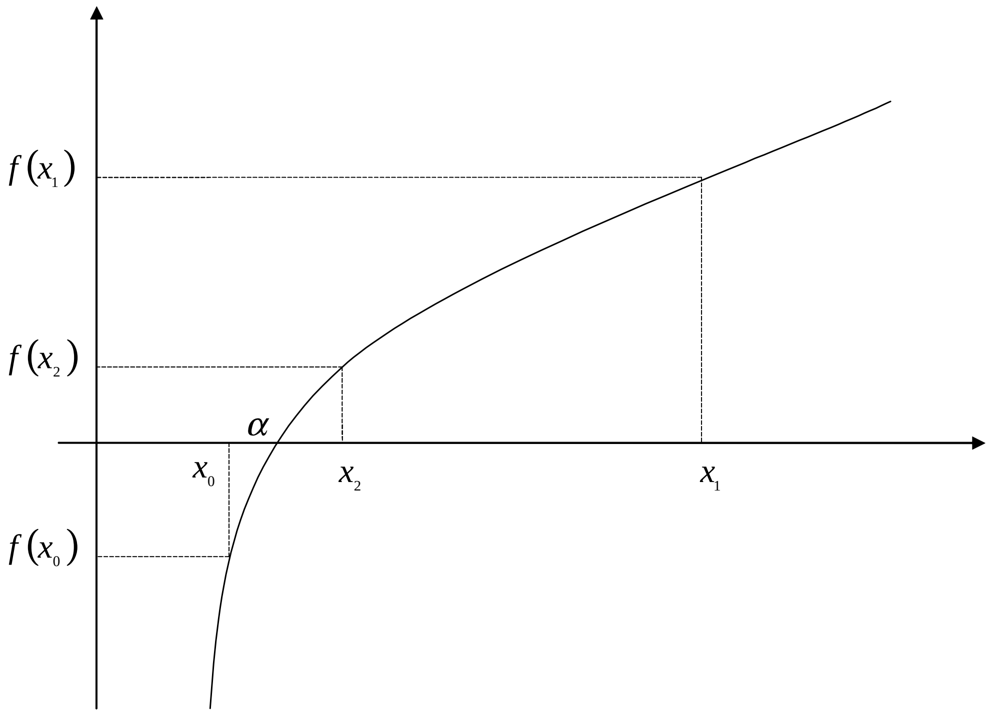
Whichever an interval is chosen,the new interval containing the root can be further subdivided. If the first interval is $|b-a|$, then the second is half the length and so on. After n steps of bisection the interval containing the root will be reduced in size to

$$\frac{|b-a|}{2^n}$$

where in the earlier example the value $b = 1$ and $a = 0$.

If the size of the interval becomes smaller than some specified tolerance, t, then the calculation stops and convergence has been attained.

Theorem 2 (Bisection Method Theorem)



If $f(x)$ is a continuous function on $[a,\ b]$ such that $f(a)f(b) < 0$, then after n steps, the method will return c such that

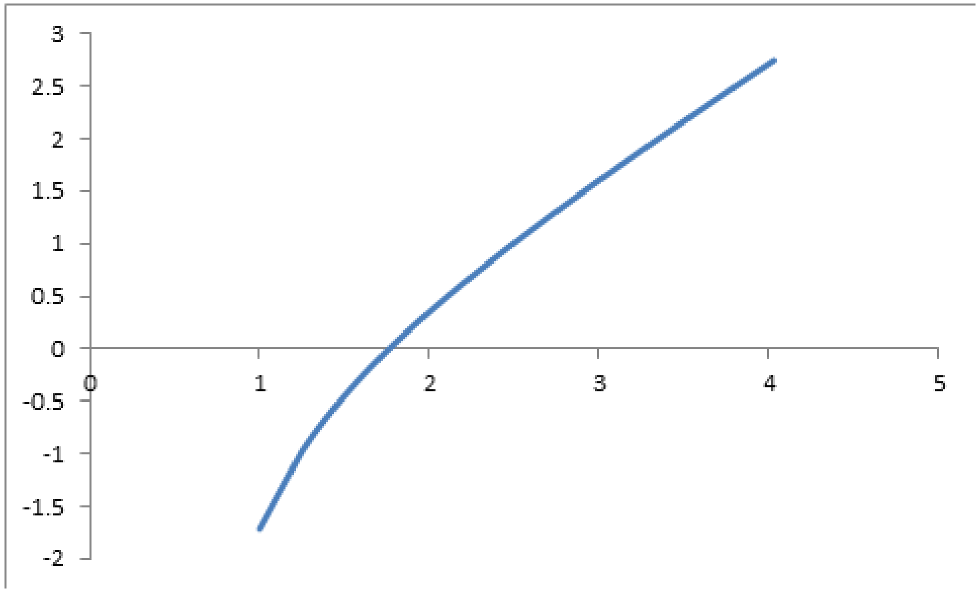
$$|c - \alpha| \leq \frac{|b - a|}{2^n}$$

where α is some approximate root of f.

Example 1 Consider

$$f(x) = x - e^{1/x}$$

There is a root in [1, 2]



(a) Use the bisection method to show that the root of $f(x) = x - e^{1/x}$ in the interval $[1,\ 2]$ is 1.763 (correct to 3 decimal places)

$$
\begin{aligned}
f(x_0) &= f(1) = 1 - e^1 < 0 \\
f(x_1) &= f(2) > 0
\end{aligned}
$$

$$
x_2 = \frac{x_0 + x_1}{2} = 1.5
$$

$$
f(1.5) = 1.5 - e^{2/3} = f(x_2) = -0.4477
$$

$$
f(x_0)\, f(x_2) > 0
$$

$\therefore$ root in $[x_2,\ x_1]$ i.e. in $[1.5,\ 2]$

$$
\begin{aligned}
x_3 &= \frac{x_2 + x_1}{2} = \frac{1.5 + 2}{2} = 1.75 \\
f(x_3) &= 1.75 - e^{1/1.75} = -0.0208 \\
f(x_3)\, f(x_2) &> 0
\end{aligned}
$$

$\therefore$ root in $[x_3,\ x_1]$ i.e. in $[1.75,\ 2]$

$$
\begin{aligned}
x_4 &= \frac{1.75 + 2}{2} = \frac{3.75}{2} = 1.875 \\
f(x_4) &= f(1.875) = 0.1704 \\
f(x_3)\, f(x_4) &< 0
\end{aligned}
$$

$\therefore$ root in $[x_3,\ x_4]$

i.e. in [1.75, 1.875]

$$\begin{aligned} x_5 &= \frac{x_3 + x_4}{2} = \frac{1.75 + 1.875}{2} = 1.8125 \\ f(x_5) &= f(1.8125) = 0.0763 \\ f(x_3)\, f(x_5) &< 0 \end{aligned}$$

$\therefore$ root in $[x_3,\ x_5]$

i.e. in [1.75, 1.8125]

$$\begin{aligned} x_6 &= \frac{x_3 + x_5}{2} = \frac{1.75 + 1.8125}{2} = 1.78125 \\ f(x_6) &= f(1.78125) > 0 \\ f(x_3)\, f(x_6) &< 0 \end{aligned}$$

$\therefore$ root in $[x_3,\ x_6]$

i.e. in [1.75, 1.78125]

$$\begin{aligned} x_7 &= \frac{x_3 + x_6}{2} = 1.765625 \\ f(x_7) &= f(1.765625) > 0 \\ f(x_3)\, f(x_7) &< 0 \end{aligned}$$

$\therefore$ root in $[x_3,\ x_7]$

i.e. in $[1.75,\ 1.765625]$

$$\begin{aligned} x_8 &= \frac{x_3 + x_7}{2} = 1.7578125 \\ f(x_8) &= f(1.7578125) < 0 \\ f(x_3)\, f(x_8) &> 0 \end{aligned}$$

$\therefore$ root in $[x_8,\ x_7]$

i.e. in [1.7578125, 1.765625]

$$
\begin{aligned}
x_9 &= \frac{x_8 + x_7}{2} = 1.76178 \\
f(x_9) &= f(1.76178) < 0 \\
f(x_8)\, f(x_9) &> 0
\end{aligned}
$$

$\therefore$ root in $[x_8,\ x_7]$

i.e. in [1.761718, 1.765625]

$$
\begin{aligned}
x_{10} &= \frac{x_9 + x_7}{2} = 1.7636715 \\
f(x_{10}) &= f(1.7636715) > 0 \\
f(x_9)\, f(x_{10}) &< 0
\end{aligned}
$$

$\therefore$ root in $[x_9,\ x_{10}]$

i.e. in [1.761718, 1.7636715]

$$\begin{aligned} x_{11} &= \frac{x_9 + x_{10}}{2} = 1.76269 \\ &= 1.763 \text{ to 3 decimal places} \end{aligned}$$

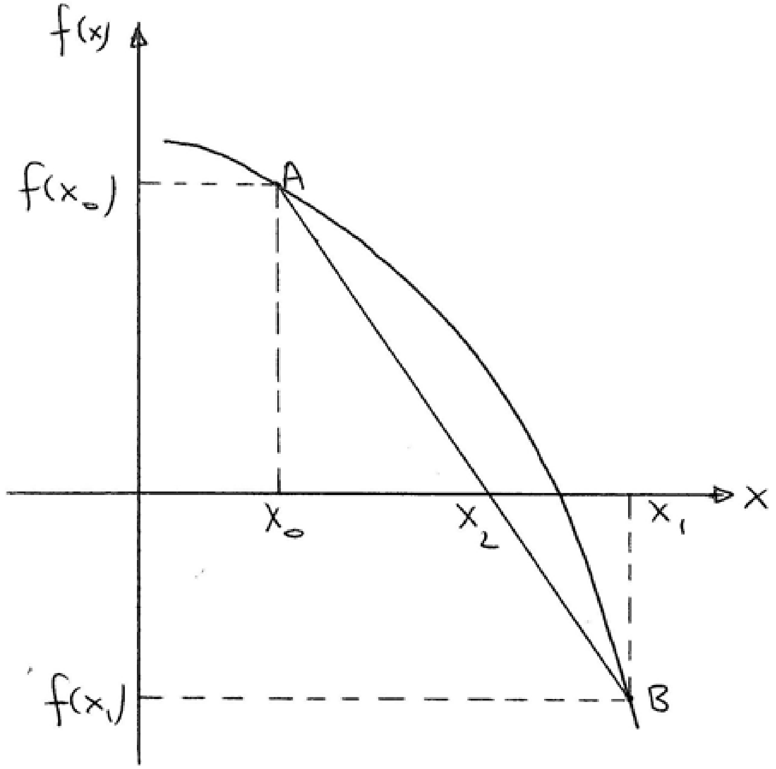
Exercise Calculate the largest root of the equation

$$e^x - 3x = 0.$$

Hint: Do a sketch first and show that a root lies between 0.4 and 0.7

The Method of False Position - Linear Interpolation

As with the bisection method assume an interval is given which contains the root



Point A is $(x_0,\ f(x_0))$; Point B is $(x_1,\ f(x_1))$

Let the line joining A and B cut the $x-$axis at x_2.

The equation for a suitable straight line joining the two points is in general

$$f(x) = mx + c \tag{1}$$

In this case we have

At A :

$$f(x_0) = mx_0 + c \tag{2}$$

At B :

$$f(x_1) = mx_1 + c \tag{3}$$

(3) − (2) gives

$$f(x_1) - f(x_0) = m(x_1 - x_0)$$

$$m = \frac{f(x_1) - f(x_0)}{(x_1 - x_0)} \tag{4}$$

From (2)

$$c = f(x_0) - \frac{(f(x_1) - f(x_0))}{(x_1 - x_0)}.x_0 \tag{5}$$

Put (4) and (5) into (1)

$$f(x) = \frac{f(x_1) - f(x_0)}{(x_1 - x_0)}x + f(x_0) - \frac{(f(x_1) - f(x_0))}{(x_1 - x_0)}.x_0$$

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{(x_1 - x_0)}(x - x_0) \tag{6}$$

To find x_2, note that $f(x) = 0$ when $x = x_2$ (i.e. that is our root)

From (6)

$$\begin{aligned} 0 &= f(x_0) + \frac{(f(x_1) - f(x_0))}{(x_1 - x_0)} \cdot (x_2 - x_0) \\ & -f(x_0)(x_1 - x_0) \\ &= (f(x_1) - f(x_0))(x_2 - x_0) \end{aligned}$$

$$x_2 = x_0 - \frac{f(x_0)(x_1 - x_0)}{f(x_1) - f(x_0)} \tag{7}$$

or

$$
\begin{aligned}
x_2 &= \frac{x_0 \left(f\left(x_1\right) - f\left(x_0\right)\right) - f\left(x_0\right)\left(x_1 - x_0\right)}{f\left(x_1\right) - f\left(x_0\right)} \\
x_2 &= \frac{x_0 f\left(x_1\right) - x_1 f\left(x_0\right)}{f\left(x_1\right) - f\left(x_0\right)}
\end{aligned}
$$

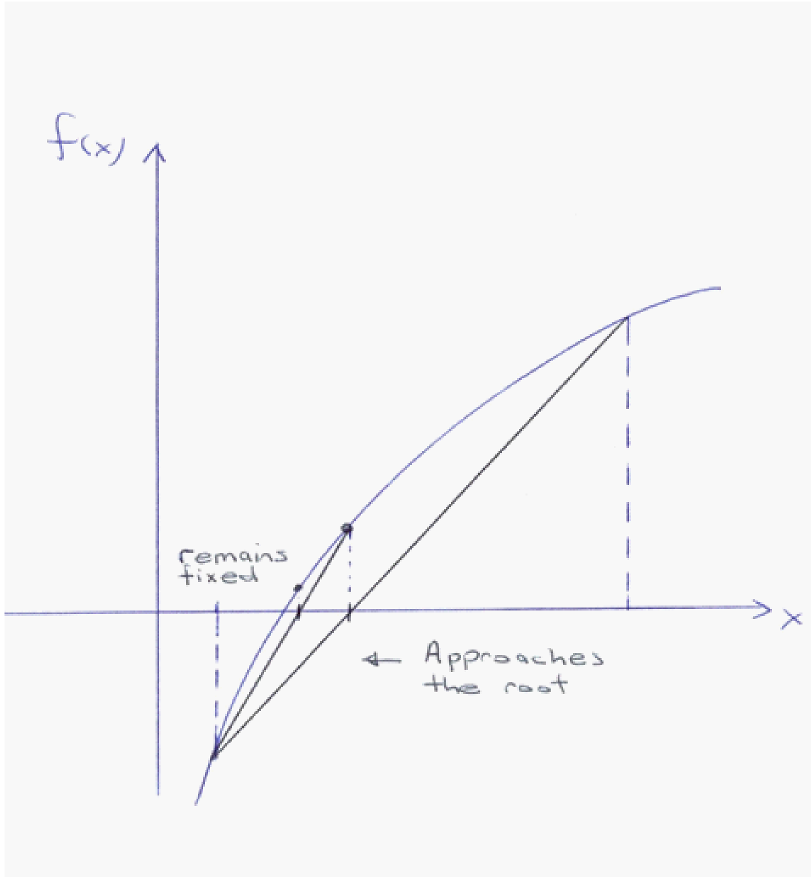
The program used for bisection can be used again for the method of false position.

Simply replace $x_2 := \dfrac{(x_0 + x_1)}{2}$ by x_2 from equation (7)

Difficulties arise with the implementation of false position as follows:-

1. $(f(x_1) - f(x_0))$ can become zero. Division by zero in (7) has to be avoided. A check is therefore required to print out a warning.

2. One end of the interval tends to remain fixed while the other end approaches the root required



As a consequence the size of the interval is not a reliable indication of the convergence. This can be overcome however by always choosing the smaller of $(x_2 - x_0)$ and $(x_1 - x_2)$ to compare with x tol in order to terminate the iterative process.

Newton's Method

Newton's method is an *iterative* method for root finding. That is, starting from some guess at the root, x_0, one iteration of the algorithm produces a number x_1, which is supposed to be closer to a root; guesses x_2, x_3, ..., x_n follow identically.

We know from Taylor that

$$f(x+h) = f(x) + f'(x)\,h + O\left(h^2\right)$$

This approximation is better when $f''(.)$ is "well-behaved" between x and $x+h$. Newton's method attempts to find some h such that

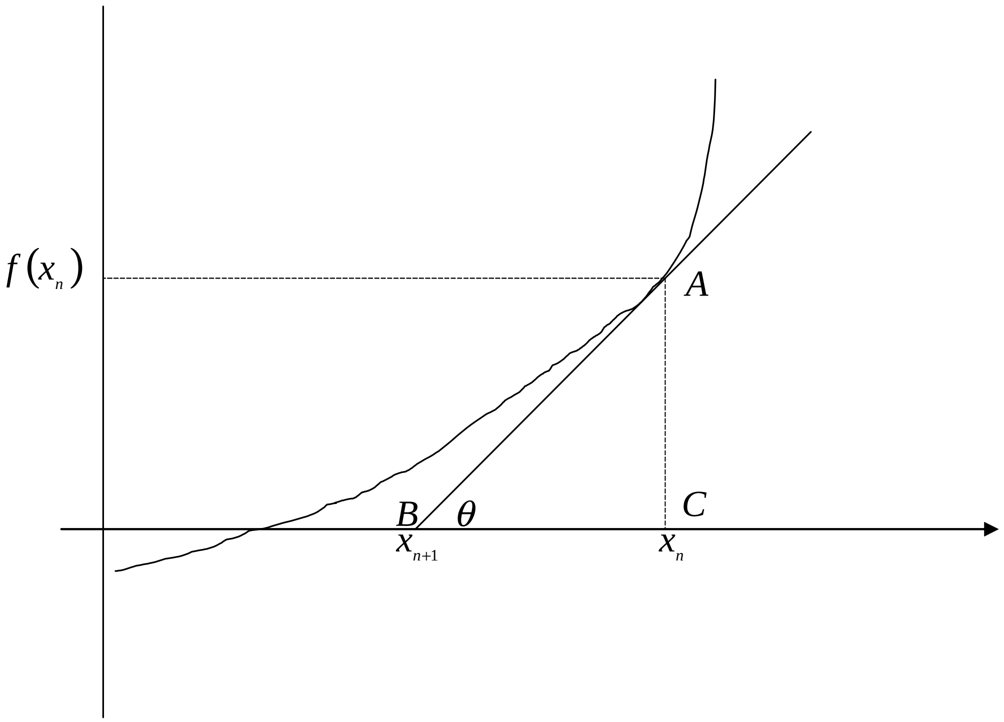
$$0 = f(x+h) = f(x) + f'(x)\,h$$

This is easily solved as

$$h = \frac{-f(x)}{f'(x)}$$

An iteration of Newton's method, then, takes some guess x_n and returns x_{n+1} defined by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$



From above we see that $\tan\theta = \dfrac{AC}{BC} = \dfrac{f(x_n)}{(x_n - x_{n+1})}$

But

$$\begin{aligned} \tan\theta &= f'(x_n) \\ f'(x_n) &= \frac{f(x_n)}{(x_n - x_{n+1})} \end{aligned}$$

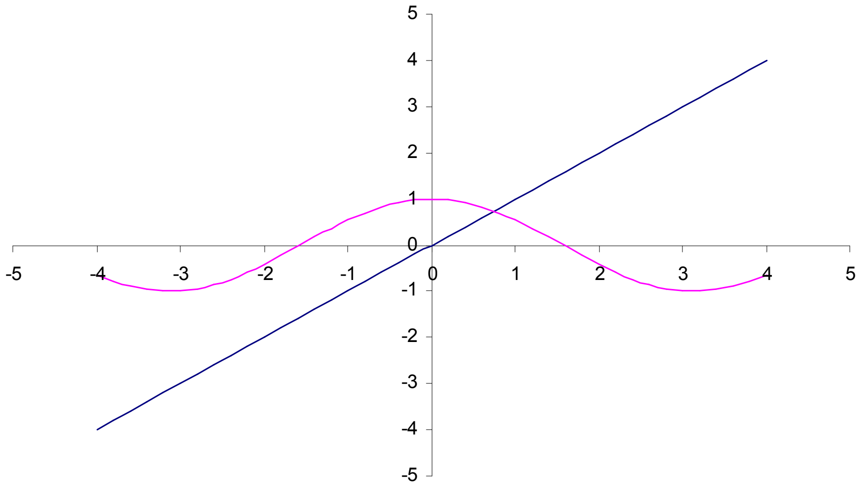
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

This is the *Newton-Raphson Technique.*

Example: Solve for roots, the function $f(x) = x - \cos x$.

Start by considering $x = \cos x$. That is draw $y = x$ and $y = \cos x$ to obtain an initial guess for the root(s).

Clearly the diagram above shows that there is only one root $\alpha \in (0, 1)$.



We use the Newton formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, \ldots.$$

where $n = 0$ is the initial guess. $f(x_n) = x_n - \cos x_n \longrightarrow f'(x_n) = 1 + \sin x_n$

x	0	1
$f(x)$	-1	0.75

so numerically we also see that $f(0)\, f(1) < 0 \Longrightarrow \alpha \in (0, 1)$. NR formula for this function becomes

$$x_{n+1} = x_n - \frac{x_n - \cos x_n}{1 + \sin x_n}, \quad x_0 = 1$$

$$\begin{aligned}
x_1 &= x_0 - \frac{x_0 - \cos x_0}{1 + \sin x_0} = 0.75036 \\
x_2 &= 0.75036 - \frac{0.75036 - \cos 0.75036}{1 + \sin 0.75036} = 0.73911 \\
x_3 &= 0.73911 - \frac{0.73911 - \cos 0.73911}{1 + \sin 0.73911} = 0.73909 \\
x_4 &= 0.73909 - \frac{00.73909 - \cos 0.73909}{1 + \sin 0.73909} = 0.73909
\end{aligned}$$

which gives the root $\alpha \approx 0.73909$

Problems

As mentioned above, convergence is dependent on $f(x)$, and the initial estimate x_0. A number of conceivable problems might come up. We illustrate them here.

Example

Consider Newton's method applied to the function $f(x) = \dfrac{\ln x}{x}$, with initial estimate $x_0 = 3$.

Note that $f(x)$ is continuous on $\mathbb{R}^+$. It has a single root at $x = 1$. Our initial guess is not too far from this root. However, consider the derivative:

$$f'(x) = \frac{x\frac{1}{x} - \ln x}{x^2} = \frac{1 - \ln x}{x^2}$$

If $x > e^1$, then $1 - \ln x < 0$, and so $f'(x) < 0$. However, for $x > 1$, we know $f(x) > 0$. Thus taking

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} > x_k$$

The estimates will diverge from the root $x = 1$.

Example

Consider Newton's method applied to the function $f(x) = \sin x$ for the initial estimate $x_0 \neq 0$, where x_0 has the property $2x_0 = \tan x_0$.

You should verify that there an infinite number of such x_0. Consider the identity of x_1 :

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = x_0 - \frac{\sin(x_0)}{\cos(x_0)} = x_0 - \tan x_0 = x_0 - 2x_0 = -x_0$$

Now consider x_2 :

$$\begin{aligned} x_2 &= x_1 - \frac{f(x_1)}{f'(x_1)} = -x_0 - \frac{\sin(-x_0)}{\cos(-x_0)} - x_0 + \frac{\sin(x_0)}{\cos(x_0)} = -x_0 + \tan x_0 \\ &= -x_0 + 2x_0 = x_0 \end{aligned}$$

Thus Newton's method "cycles" between the two values $x_0, \ -x_0$.

Of course, Newton's method may find some iterate x_k for which $f'(x_k) = 0$, in which case, there is no well-defined x_{k+1}.

Interpolation

Polynomial Interpolation

We consider the problem of finding a polynomial that interpolates a given set of values:

$$\begin{array}{ccccc} x & x_0 & x_1 & \cdots & x_n \\ y & y_0 & y_1 & \cdots & y_n \end{array}$$

where the x_i are all distinct. A polynomial $p(x)$ is said to interpolate these data if $p(x_i) = y_i$ for $i = 0, 1,$ ......, n. The x_i values are called "nodes".

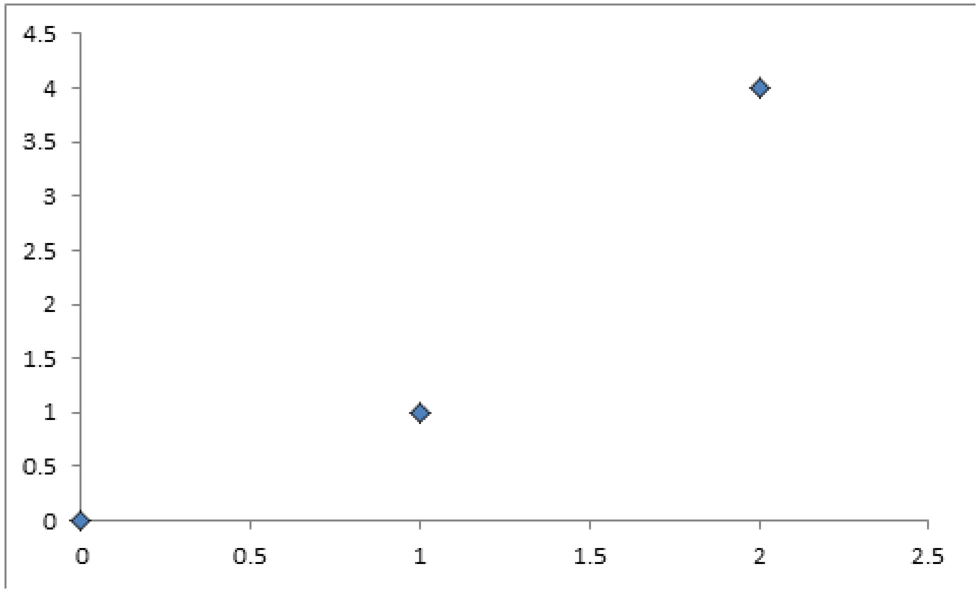
Sometimes, we will consider a variant of this problem: we have some black box function, $f(x)$, which we want to approximate with a polynomial $p(x)$. We do this by finding the polynomial interpolant to the data

$$\begin{array}{ccccc} x & x_0 & x_1 & \cdots & x_n \\ f(x) & f(x_0) & f(x_1) & \cdots & f(x_n) \end{array}$$

for some choice of distinct nodes x_i.

Lagrangian Interpolation

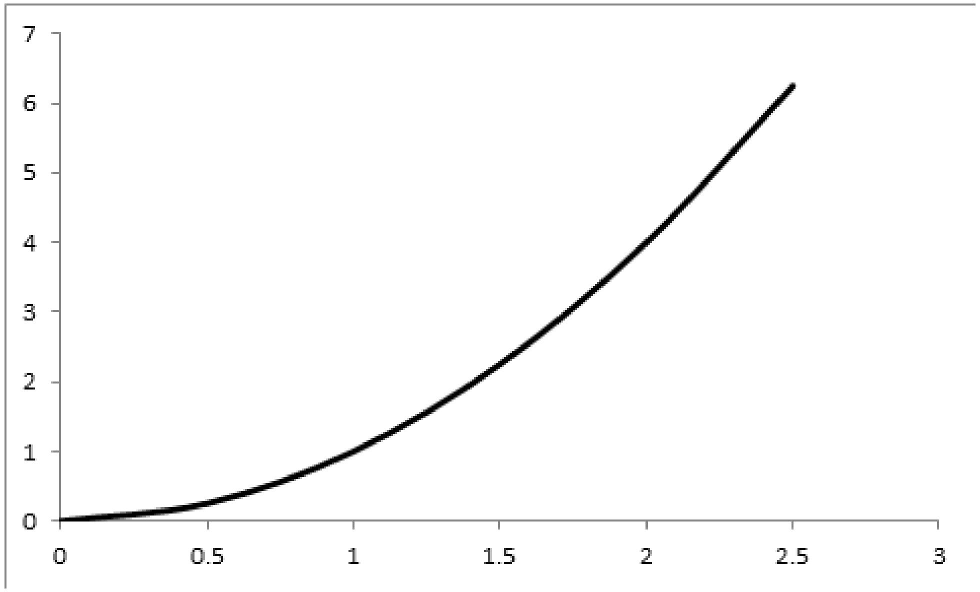
In general the term **interpolation** is the term used to describe the process in which a function, usually a polynomial function is constructed to pass through a given set of data points. As an example consider the following data points (x_i, y_i) as follows: $(0,0)$; $(1,1)$; $(2,4)$



In this example the function $y = x^2$ passes through these 3 data points.

Why is interpolation important and how is it used?

Suppose an estimate is required for the value of y at some x other than $x = 0, 1, 2$. In this case, the interpolating polynomial $y = x^2$ can be used to estimate such intermediate values, e.g $x = 1.5$, which gives $y = 2.25$


7
6
5
4
3
2
1
0
0
0.5
1
1.5
2
2.5
3

Given the data points $(0,0)$; $(1,1)$; $(2,4)$ the values of y corresponding to any value of x between $x = 0$ and $x = 2$ can be estimated. This is known as *interpolation*.

Lagrange Polynomials

For a given set of $n + 1$ nodes x_i,

i	0	1	2	....	n
x_i	x_0	x_1	x_2	....	x_i
y_i	y_0	y_1	y_2	....	y_n

the Lagrange polynomials are the $n + 1$ polynomials l_i defined by

$$l_i\left(x_j\right) = \delta_{ij} = \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j \end{cases}$$

Then we define the interpolating polynomial as

$$p_n(x) = \sum_{i=0}^{n} y_i l_i(x)$$

If each Lagrange Polynomial is of degree at most n, then p_n also has this property. The Lagrange Polynomials can be characterised as follows:

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

By evaluating this product for each x_j, we see that this is indeed a characterisation of the Lagrange Polynomials. Moreover, each polynomial is clearly the product of n "mononomials", and thus has degree no greater than n.

There is a useful check on the correctness of numerical calculations, i.e.

$$\sum_{i=0}^{n} l_i\,(x) = 0.$$

Example

Construct the polynomial interpolating the data

$$\begin{array}{cccc} x & 1 & \frac{1}{2} & 3 \\ y & 3 & -10 & 2 \end{array}$$

by using Lagrange Polynomials.

Solution:

$$\begin{array}{cccc} i & 0 & 1 & 2 \\ x_i & 1 & \frac{1}{2} & 3 \\ y_i & 3 & -10 & 2 \end{array}$$

We construct the Lagrange Polynomials:

$$\begin{aligned} l_0(x) &= \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} \\ &= \frac{\left(x - \frac{1}{2}\right)(x - 3)}{\left(1 - \frac{1}{2}\right)(1 - 3)} \\ &= -\left(x - \frac{1}{2}\right)(x - 3) \end{aligned}$$

$$
\begin{aligned}
l_1(x) &= \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} \\
&= \frac{(x-1)(x-3)}{\left(\frac{1}{2}-1\right)\left(\frac{1}{2}-3\right)} \\
&= \frac{4}{5}(x-1)(x-3)
\end{aligned}
$$

$$
\begin{aligned}
l_2(x) &= \frac{(x-x_0)(x-x_1)}{(x_2-x_1)(x_2-x_1)} \\
&= \frac{(x-1)\left(x-\frac{1}{2}\right)}{(3-1)\left(3-\frac{1}{2}\right)} \\
&= \frac{1}{5}(x-1)\left(x-\frac{1}{2}\right)
\end{aligned}
$$

Then the interpolating polynomial, in "Lagrange Form" is

$$\begin{aligned} p_2(x) &= \sum_{i=0}^{n} y_i l_i(x) = y_0 l_0(x) + y_1 l_1(x) + y_2 l_2(x) \\ &= 3 l_0(x) - 10 l_1(x) + 2 l_2(x) \\ &= -3\left(x - \frac{1}{2}\right)(x-3) - 8(x-1)(x-3) \\ &\quad + \frac{2}{5}(x-1)\left(x - \frac{1}{2}\right) \end{aligned}$$

Class Exercise: Using Lagrangian Interpolation find $f(0.14)$ from the given values of

i	0	1	2	3
x_i	0	0.1	0.3	0.6
y_i	1.0	1.10517	1.34986	1.82212

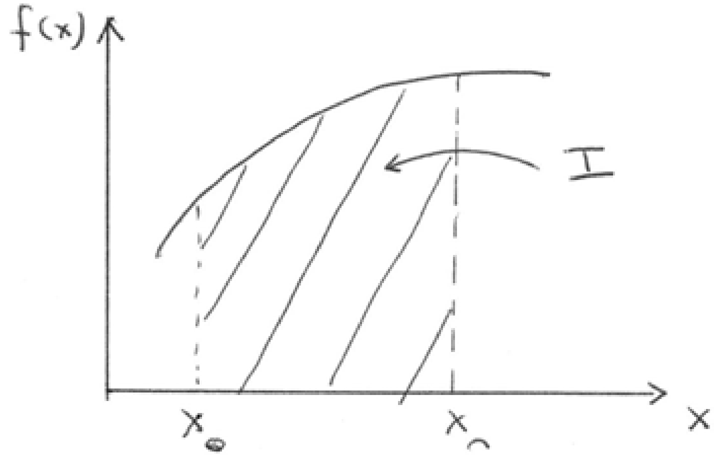
For the given set of data calculate sum of the Lagrange polynomials.

Integration by Numerical Methods

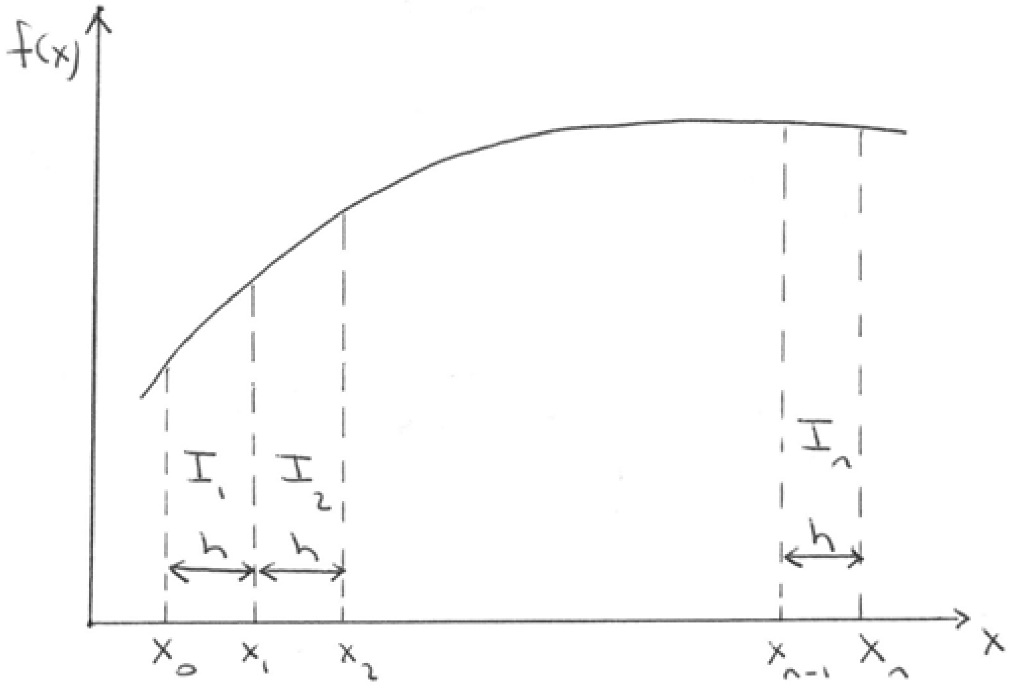
The problem is to find I where

$$I = \int_{x_0}^{x_n} f(x)\, dx$$

by numerical means, I maybe regarded as an area, in the following diagram:



The general idea of numerical integration is to breakdown the area under the curve $f\left(x\right)$ against x into small equally spaced strips as follows:



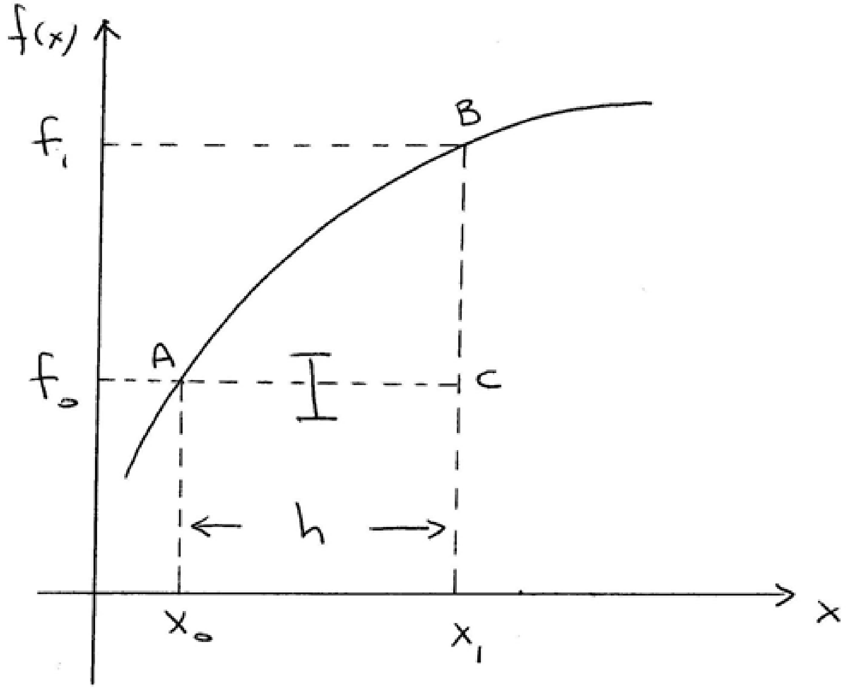
Then approximate the area of each strip (i.e $I_1,\ I_2, ...$) and add together the areas resulting to give

$$I = I_1 + I_2 + ... + I_n \tag{1}$$

There are many different ways of approximating the area of the strips. And some of the common rules for numerical integration will be covered in the subsection.

Trapezoidal Rule

In this case it is assumed that each area, I_1 for example, is a trapezium.



That is, points A and B are joined with a straight line and the area of the trapezium ABx_1x_0 is assumed to be a good approximation to the area under the curve.

The area ABx_1x_0 is calculated as follows:

$$\begin{aligned} I_1 &\approxeq \text{Area } ACx_1x_0 + \text{Area } ABC \\ I_1 &\approxeq f_0 h + \frac{1}{2}(f_1 - f_0)\,h \\ I_1 &\approxeq \frac{h}{2}(f_0 + f_1) \end{aligned}$$

An approximation to

$$I_1 = \int_{x_0}^{x_1} f(x)\,dx$$

is now given by

$$I_1 \approxeq \frac{h}{2}(f_0 + f_1) \tag{2}$$

This is the *trapezoidal rule*.

Consider the example

$$\int_0^{\pi/2} \sin x dx$$

Analytically the result is 1.

$$\int_0^{\pi/2} \sin x dx \approxeq \frac{h}{2}(f_0 + f_1) = \frac{\pi/2}{2}(0+1) = 0.785398$$

The global error is the difference between the exact solution and approximation $= 1 - 0.785398 = 0.214602$ which is a large error because $h = \pi/2$ is so large.

To obtain an analytic expression for the truncation error start by writing

$$\underbrace{\int_{x_0}^{x_1} f(x)\, dx}_{\text{Exact}} = \frac{h}{2}(f(x_0) + f(x_1)) + \text{Error}$$

LHS by Taylor

$$f(x) = f(x_0) + (x - x_0) f'(x_0) + (x - x_0)^2 \frac{f''(x_0)}{2!} + \ldots$$

$$\begin{aligned}
\int_{x_0}^{x_1} f(x)\,dx &= \int_{x_0}^{x_0+h} f(x)\,dx \\
&= \int_{x_0}^{x_0+h} \left(f(x_0) + (x - x_0) f'(x_0) + (x - x_0)^2 \frac{f''(x_0)}{2!} + \ldots \right) dx \\
&= f(x_0)\,[x]_{x_0}^{x_0+h} + f'(x_0) \left[\frac{(x - x_0)^2}{2} \right]_{x_0}^{x_0+h} + \\
&\quad \frac{f''(x_0)}{2!} \left[\frac{(x - x_0)^3}{3} \right]_{x_0}^{x_0+h} + \ldots \\
&= hf(x_0) + \frac{h^2}{2} f'(x_0) + \frac{h^3}{6} f''(x_0) + \ldots . \text{Exact}
\end{aligned}$$

Now the RHS $=$

$$\frac{h}{2}(f_0+f_1)=\frac{h}{2}\left(f(x_0)+\underbrace{f(x_0)+hf'(x_0)+\frac{h^2}{2}\frac{f''(x_0)}{2!}+\ldots}_{\text{Taylor for } f(x_1)}\right)$$

$$\begin{aligned}
&= \frac{h}{2}\left(2f(x_0)+hf'(x_0)+\frac{h^2}{2}\frac{f''(x_0)}{2!}+\ldots\right)\\
&= hf(x_0)+\frac{h^2}{2}f'(x_0)+\frac{h^3}{4}\frac{f''(x_0)}{2!}+\ldots\text{Approximate}
\end{aligned}$$

The error is LHS $-$ RHS

$$
\begin{aligned}
& \left(hf\left(x_0\right)+\frac{h^2}{2}f'\left(x_0\right)+\frac{h^3}{6}f''\left(x_0\right)+\ldots.\right)- \\
& \left(hf\left(x_0\right)+\frac{h^2}{2}f'\left(x_0\right)+\frac{h^3}{4}\frac{f''\left(x_0\right)}{2!}+\ldots\right) \\
= \quad & \underbrace{-\frac{h^3}{12}f''\left(x_0\right)}_{\textbf{Truncation Error}} \; + \; \text{H.O.T in } h.
\end{aligned}
$$

In the earlier example we had $\int_0^{\pi/2} \sin x dx \approxeq 0.785398$.

To estimate the truncation error using

$$E = -\frac{h^3}{12} f''(x_0)$$

Taking $h = \pi/2$

$$f(x) = \sin x; \;\; f'(x) = \cos x; \;\; f''(x) = -\sin x$$

To find a bound for the error

$$|E| \leq \frac{h^3 M}{12} \text{ where } M = \max \left| f''(x) \right| \quad x \in [0, \pi/2]$$

$$\therefore |E| \leq \frac{(\pi/2)^3 \times 1}{12} \;\; \text{since } |\sin(x)| \leq 1 = M$$

The Composite Trapezoidal Rule

This is simply an extension of the trapezoidal rule to include all strips. We have

$$I = I_1 + I_2 + ... + I_n$$

this is the Trapezoidal Rule

$$\begin{aligned}\int_{x_0}^{x_n} f(x)\,dx &= \int_{x_0}^{x_1} f(x)\,dx + \int_{x_1}^{x_2} f(x)\,dx + ... \\ &\quad + \int_{x_{n-1}}^{x_n} f(x)\,dx\end{aligned}$$

From (2)

$$I_1 = \frac{h}{2}(f_0 + f_1)$$

and by extension

$$\begin{aligned} I_2 &= \frac{h}{2}(f_1 + f_2) \\ I_3 &= \frac{h}{2}(f_2 + f_3) \end{aligned}$$

etc. So summing these quantities gives the following

$$I = \frac{h}{2}\left[(f_0 + f_1) + (f_1 + f_2) + (f_2 + f_3) + \ldots. (f_{n-1} + f_n)\right]$$

<u>OR</u>

$$I = \frac{h}{2}\left[f_0 + 2f_1 + 2f_2 + \ldots\ldots + 2f_{n-1} + f_n\right]$$

i.e.

$$I = \frac{h}{2}\left[f_0 + 2\sum_{i=1}^{n-1} f_i + f_n\right] \tag{3}$$

This is the *Composite Trapezoidal Rule*.

From (3) special cases maybe derived e.g, the composite trapezoidal rule for just two strips.

$$I = \int_{x_0}^{x_2} f(x)\,dx = \frac{h}{2}[f_0 + 2f_1 + f_2] \tag{4}$$

Now calculate the error in the Composite Trapezoidal Rule:

$$\begin{aligned}
\int_{x_0}^{x_n} f(x)\,dx &= \int_{x_0}^{x_1} f(x)\,dx + \int_{x_1}^{x_2} f(x)\,dx + \dots\dots\dots + \int_{x_{n-1}}^{x_n} f(x)\,dx \\
&= \frac{h}{2}(f_0 + f_1) + \frac{h}{2}(f_1 + f_2) + \frac{h}{2}(f_2 + f_3) + \dots. + \frac{h}{2}(f_{n-1} + f_n)
\end{aligned}$$

Error $=$

$$\begin{aligned}
&-\frac{h^3}{12}f''(x_0) - \frac{h^3}{12}f''(x_1) - \frac{h^3}{12}f''(x_2) - \dots.. - -\frac{h^3}{12}f''(x_{n-2}) - \frac{h^3}{12}f''(x_{n-1}) \\
= &-\frac{h^3}{12}\left(f''(x_0) + f''(x_1) + \dots. + f''(x_{n-1})\right)
\end{aligned}$$

Approximate

$$f''(x_i) = \frac{f'(x_{i+1}) - f'(x_i)}{h}$$

Error $=$

$$
\begin{aligned}
& -\frac{h^3}{12}\left(\frac{f_1' - f_0'}{h} + \frac{f_2' - f_1'}{h} + \frac{f_3' - f_2'}{h} + \ldots + \frac{f_{n-1}' - f_{n-2}'}{h} + \frac{f_n' - f_{n-1}'}{h}\right) \\
= \quad & -\frac{h^3}{12}\frac{\left(f_n' - f_0'\right)}{h}
\end{aligned}
$$

Error $= -\dfrac{h^2}{12}\left(f_n' - f_0'\right).$

An alternate form of the Mean Value Theorem:

$$f_b' - f_a' = (b-a)\, f''(\xi) \quad a < \xi < b$$

From above we know the error in $\int_a^b f(x)\, dx = -\frac{h^2}{12}\left(f_b' - f_a'\right)$

Hence error is

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

Putting

$$\begin{aligned} a &= x_0;\ b = x_n; \\ b - a &= x_n - x_0 = x_0 + nh - x_0 = nh \end{aligned}$$

Hence the truncation error is

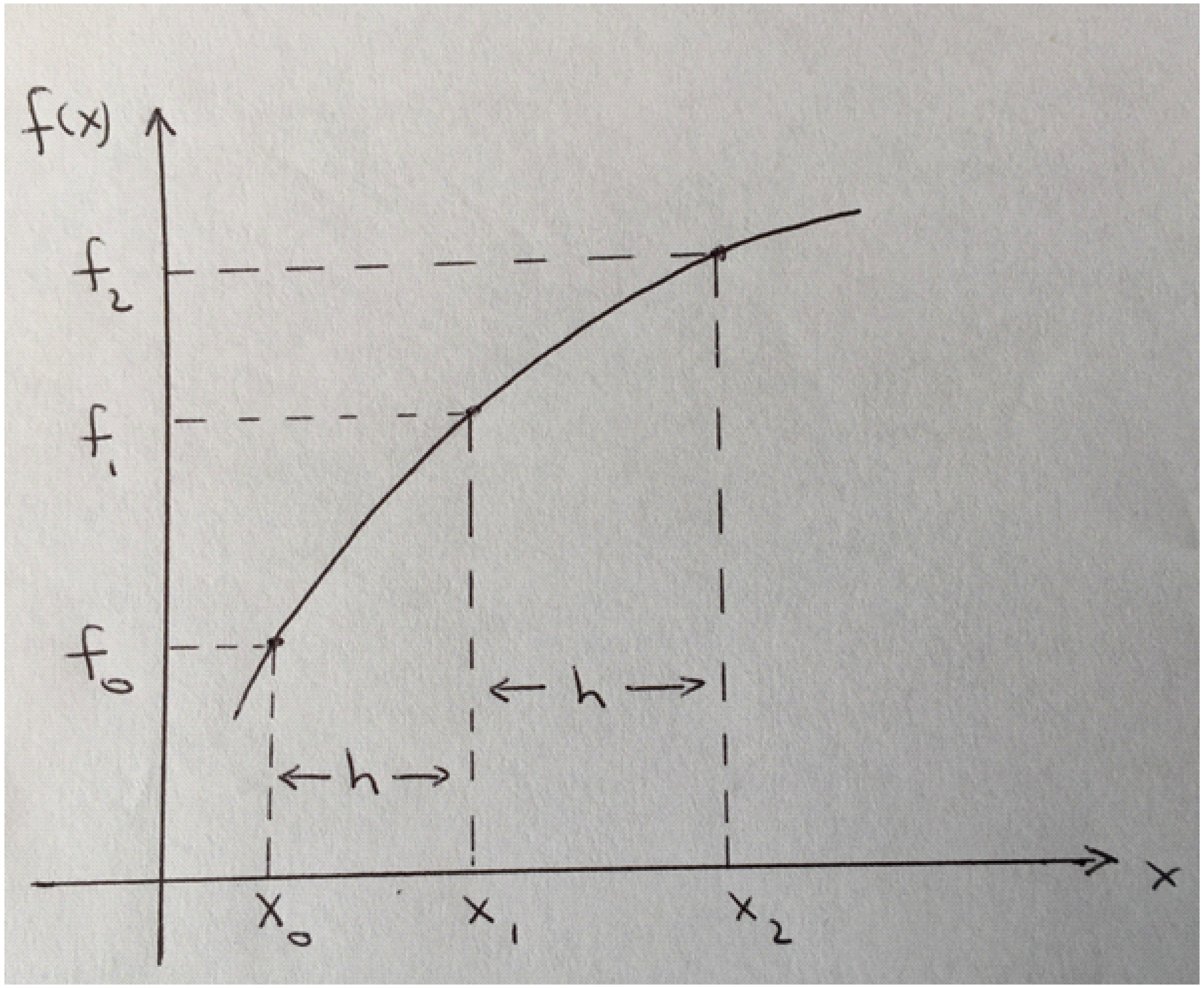
$$-\frac{h^2}{12}(nh)\, f''(\xi)$$

Simpson's Rule

Start by re-stating the earlier result

$$l_i(x) = \prod_{j=0,\ j\neq i}^{n} \frac{x - x_j}{x_i - x_j} \tag{5}$$

Simpson's Rule is obtained by considering two strips together, fitting a quadratic polynomial through the points on the curve and then integrating the quadratic function



Let

$$I = \int_{x_0}^{x_2} f\left(x\right) dx \tag{6}$$

To find I first fit a quadratic to the points ABC. By Lagrangian Interpolation of degree 2

$$f(x) \approxeq p_2(x) = f_0 l_0 + f_1 l_1 + f_2 l_2 \tag{7}$$

$$l_0 = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}$$

etc

Now

$$\begin{aligned}
(x_0 - x_1) &= -h \; ; \; (x_0 - x_2) = -2h \\
(x_1 - x_0) &= h \; ; \; (x_1 - x_2) = -h \\
(x_2 - x_1) &= h \; ; \; (x_2 - x_0) = 2h
\end{aligned}$$

So from (7)

$$\begin{aligned} f(x) &\approxeq p_2(x) = \frac{(x-x_1)(x-x_2)}{2h^2} f_0 - \\ &\frac{(x-x_0)(x-x_2)}{h^2} f_1 + \frac{(x-x_0)(x-x_1)}{2h^2} f_2 \end{aligned}$$

$f(x)$ is now substituted into (6) to give

$$\begin{aligned} I \approxeq\ & \frac{1}{2h^2} \int_{x_0}^{x_2} \{(x-x_1)(x-x_2) f_0 \\ & -2(x-x_0)(x-x_2) f_1 + (x-x_0)(x-x_1) f_2\} dx \end{aligned} \tag{8}$$

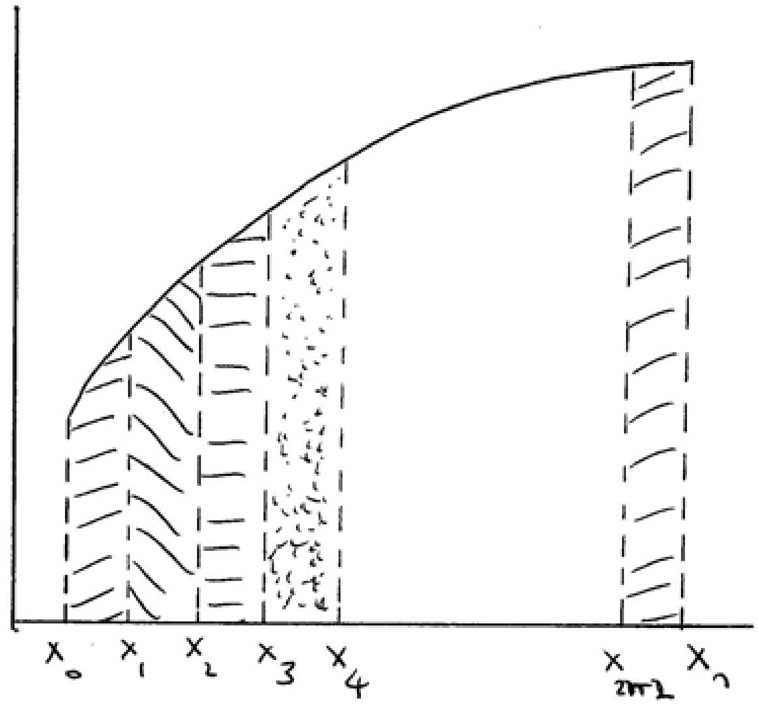
(8) is integrated by parts to give

$$I \approxeq \frac{h}{3}(f_0 + 4f_1 + f_2) \tag{9}$$

This is *Simpson's Rule*.

Composite Form of Simpson's Rule

This is similar in principle to the Composite Trapezoidal Rule but here the region of integration is split into pairs of strips as shown in the diagram



$$I = \int_{x_0}^{x_{2n}} f(x)\, dx$$

is row split into pairs of strips such that

$$I = \int_{x_0}^{x_2} f(x)\, dx + \int_{x_2}^{x_4} f(x)\, dx + ... + \int_{x_{2n-2}}^{x_{2n}} f(x)\, dx$$

Each integral on the right hand side of this expression above is now evaluated by Simpson's Rule.

This gives from (9) applied in the appropriate interval

$$I \approxeq \frac{h}{3} \begin{pmatrix} f_0 + 4f_1 + 2f_2 + 4f + 2f_4 + ... \\ +2f_{2n-2} + 4f_{2n-1} + f_{2n} \end{pmatrix}$$

so that

$$I \approxeq \frac{h}{3} \left(f_0 + f_{2n} + 4 \sum_{i=1}^{n} f_{2i-1} + 2 \sum_{i=1}^{n-1} f_{2i} \right) \qquad (10)$$

This is the composite form of Simpson's Rule.

Error Terms

When the function $f(x)$ can be integrated analytically it is possible to work out the actual error in the numerical approximation. Simply take the exact value, I, and subtract the approximate value giving, for example

$$E_T = I - I_T$$

for the trapezoidal rule or

$$E_S = I - I_S$$

for Simpson's Rule.

It is also possible to work out what is called the leading error term. This gives an indication of the order of magnitude of the error in the numerical approximations. This topic Error Terms will not be covered in detail. A formula for the leading error term for Simpson's Rule will be quoted (not derived). Simpson's Rule and its leading term is as follows

$$\int_{x_0}^{x_2} f(x)\,dx = \underbrace{\frac{h}{3}\left(f_0 + 4f_1 + f_2\right)}_{I_S} - \underbrace{\frac{h^5}{90} f^{(4)}(\xi)}_{\text{Leading Error Term}}$$

where $\xi \in (x_0, x_2)$

The trapezoidal rule can be written as

$$\int_{x_0}^{x_1} f(x)\,dx = \underbrace{\frac{h}{2}\left(f_0 + f_1\right)}_{I_T} - \underbrace{\frac{h^3}{12} f^{(2)}(\xi)}_{\text{Leading Error Term}}$$

where $\xi \in (x_0, x_1)$.

The Composite Trapezoidal and Simpson's rules become, in turn

$$\int_a^b f(x)\,dx = \frac{h}{2}\left[f(a) + 2\sum_{i=1}^{n-1} f(x_i) + f(b)\right] - \frac{(b-a)h^3}{12} f^{(2)}(\mu)$$

$$\int_a^b f(x)\,dx = \frac{h}{3}\left(f(a) + 2\sum_{i=1}^{n-1} f(x_{2i}) + 4\sum_{i=1}^{n} f(x_{2i-1}) + f(b)\right) - \frac{(b-a)h^4}{180} f^{(4)}(\mu)$$

where $\mu \in (a, b)$.

To compute the error in Simpson's Rule expand both sides in a Taylor series and then error = LHS − RHS. So starting with

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}\left(f_0 + 4f_1 + f_2\right)$$

$$\mathbf{LHS} = \int_{x_0}^{x_2} f(x)\,dx =$$

$$\int_{x_0}^{x_2}\left(f(x_0) + (x-x_0)f'(x_0) + (x-x_0)^2\frac{f''(x_0)}{2!} + \right.$$
$$\left.(x-x_0)^3\frac{f'''(x_0)}{3!} + (x-x_0)^4\frac{f^{(iv)}(x_0)}{4!} + \ldots dx\right)$$

$$= \quad f(x_0)\,[x]_{x_0}^{x_2} + f'(x_0)\left[\frac{(x-x_0)^2}{2}\right]_{x_0}^{x_2} + \frac{f''(x_0)}{2!}\left[\frac{(x-x_0)^3}{3}\right]_{x_0}^{x_2} +$$
$$\frac{f'''(x_0)}{3!}\left[\frac{(x-x_0)^4}{4}\right]_{x_0}^{x_2} + \frac{f^{(iv)}(x_0)}{4!}\left[\frac{(x-x_0)^5}{5}\right]_{x_0}^{x_2} + \ldots$$

$$= f_0 (x_2 - x_0) + \frac{f_0'}{2} (x_2 - x_0)^2 + \frac{f_0''}{6} (x_2 - x_0)^3 + \frac{f_0'''}{24} (x_2 - x_0)^4 + \frac{f_0^{(iv)}}{120} (x_2 - x_0)^5 + \ldots$$

Now $x_2 = x_0 + 2h$, so the expression above becomes

$$f_0 (2h) + \frac{f_0'}{2} (2h)^2 + \frac{f_0''}{6} (2h)^3 + \frac{f_0'''}{24} (2h)^4 + \frac{f_0^{(iv)}}{120} (2h)^5 + \ldots$$

$$= 2hf_0 + 2h^2 f_0' + \frac{4}{3} h^3 f_0'' + \frac{2}{3} h^4 f_0''' + \frac{4}{15} h^5 f_0^{(iv)} + \ldots = \mathbf{LHS}$$

Now consider the $\mathbf{RHS} = \frac{h}{3} (f_0 + 4f_1 + f_2)$

$$f_1 = f(x_1) = f(x_0 + h) = f_0 + hf_0' + \frac{h^2}{2} f_0'' + \frac{h^3}{6} f_0''' + \frac{h^4}{24} f_0^{(iv)} + \ldots$$

$$
\begin{aligned}
f_2 &= f\left(x_2\right) = f\left(x_0+2h\right) = f_0 + 2hf_0' + (2h)^2\,\frac{f_0''}{2} + (2h)^3\,\frac{f_0'''}{6} + (2h)^4\,\frac{f_0^{(iv)}}{24} + \ldots \\
&= f_0 + 2hf_0' + 2h^2 f_0'' + \frac{4}{3}h^3 f_0''' + \frac{2}{3}h^4 f_0^{(iv)} + \ldots
\end{aligned}
$$

So Simpson's Rule becomes

$$
\begin{aligned}
& \frac{h}{3}\left(f_0 + \left(4f_0 + 4hf_0' + 4h^2\frac{f_0''}{2} + 4h^3\frac{f_0'''}{6} + 4h^4\frac{f_0^{(iv)}}{24} + \ldots\right)\right. \\
& \left. + f_0 + 2hf_0' + 2h^2 f_0'' + \frac{4h^3}{3}f_0''' + \frac{2h^4}{3}f_0^{(iv)} + \ldots\right) \\
= \;& \frac{h}{3}\left(6f_0 + 6hf_0' + 4h^2 f_0'' + 2h^3 f_0''' + \frac{5h^4}{6}f_0^{(iv)} + \ldots\right)
\end{aligned}
$$

Further simplification gives

Simpson's Rule $= 2hf_0 + 2h^2 f_0' + \frac{4}{3}h^3 f_0'' + \frac{2}{3}h^4 f_0''' + \frac{5}{18}h^5 f_0^{(iv)} + \ldots. = \mathbf{RHS}$

Error $=$ $\mathbf{LHS}-$ Simpson's Rule $=$

$$\left(2hf_0 + 2h^2 f_0' + \frac{4}{3}h^3 f_0'' + \frac{2}{3}h^4 f_0''' + \frac{4}{15}h^5 f_0^{(iv)} + ...\right) -$$
$$\left(2hf_0 + 2h^2 f_0' + \frac{4}{3}h^3 f_0'' + \frac{2}{3}h^4 f_0''' + \frac{5}{18}h^5 f_0^{(iv)} + \right)$$
$$= \frac{4}{15}h^5 f_0^{(iv)} - \frac{5}{18}h^5 f_0^{(iv)} + O\left(h^6\right)$$
$$= -\frac{1}{90}h^5 f_0^{(iv)}$$

A similar working as with the Composite Trapezoidal Rule by applying the MVT gives an error for the Composite Simpson Rule as

$$-\frac{h^4}{180}(b-a)\, f^{(iv)}(\xi)\,; \quad \xi \in [a,b]$$
$$\text{Error} = -\frac{h^5}{180} n f^{(iv)}(\xi)\,.$$

Vector and Matrix Norms

In two and three dimensions, the size of vectors called the modulus is generally obtained by Pythagoras. So if

$$\underline{v} = (a_1,\ a_2,\ a_3)$$

then the modulus of this vector $|\underline{v}| = \sqrt{\left(a_1^2 + a_2^2 + a_3^2\right)}$ which gives us the distance from the origin.

Now consider $\underline{x} = (x_1,\ x_2, ...,\ x_n)^{\mathsf{T}}$. Let $\mathbb{R}^n$ denote the set of all $n-$ dimensional vectors (so all vector components are real). To define distance in $\mathbb{R}^n$ we use the notion of a *norm.* A pair of double vertical lines $\|\cdot\|$ is used to denote a norm, i.e. size of an $n-$ dimensional vector.

More formally a *vector norm* on $\mathbb{R}^n$ is a function, $\|\cdot\|$, such that

$$\|\cdot\| : \mathbb{R}^n \longrightarrow \mathbb{R}^+$$

with the following properties:

(i) $\|\underline{x}\| \geq 0 \quad \forall \underline{x} \in \mathbb{R}^n$

(ii) $\|\underline{x}\| = 0$ iff $\underline{x} = \underline{0}$

(iii) $\|\alpha \underline{x}\| = |\alpha| \, \|\underline{x}\| \quad \forall \alpha \in \mathbb{R}$ and $\underline{x} \in \mathbb{R}^n$

(iv) $\left\|\underline{x} + \underline{y}\right\| \leq \|\underline{x}\| + \left\|\underline{y}\right\| \quad \forall \underline{x}, \ \underline{y} \in \mathbb{R}^n$ (Triangle inequality)

If $\underline{x} = (x_1, \ x_2, ..., \ x_n)^{\mathsf{T}}$, then the vector norm $\|\underline{x}\|_p$ for $p = 1, 2, ...$ is defined as

$$\|\underline{x}\|_p = \left\{ \sum_{i=1}^{n} |x_i|^p \right\}^{1/p}$$

Then there are a number of ways of defining a *norm*, depending on the value of p.

The special case $\|\underline{x}\|_\infty$ is given by

$$\|\underline{x}\|_\infty = \max_{1 \leq i \leq n} |x_i| \qquad\qquad l_\infty \text{ norm.}$$

The most commonly encountered vector norm (also known as the modulus of a vector) is $\|\underline{x}\|_2$ and defined by

$$\|\underline{x}\|_2 = \left\{ \sum_{i=1}^{n} x_i^2 \right\}^{1/2} \qquad\qquad l_2 \text{ norm}$$

For obvious reasons this is also called the *Euclidean norm*.

Examples

1. $\underline{x} = \left(\sin k,\ \cos k,\ 2^k\right)^\mathsf{T}$; k is a positive integer

$$\|\underline{x}\|_2 = \sqrt{\sin^2 k + \cos^2 k + 4^k} = \sqrt{1 + 4^k}$$

$$\|\underline{x}\|_\infty = \max\left|\sin k,\ \cos k,\ 2^k\right| = 2^k$$

2. $\underline{x} = (2,\ 1,\ 3,\ -4)^\mathsf{T}$

$\|\underline{x}\|_2 = \sqrt{30}$ and $\|\underline{x}\|_\infty = 4$

These and other types of vector norms are further summarized in the following table, using a particular vector $\underline{v} = (1,\ 2,\ 3)$

Name	Symbol	value	numerical value
l_1 − norm	$\|\underline{x}\|_1$	6	6.000
l_2 − norm	$\|\underline{x}\|_2$	$\sqrt{14}$	3.742
l_3 − norm	$\|\underline{x}\|_3$	$6^{2/3}$	3.302
l_4 − norm	$\|\underline{x}\|_4$	$2^{1/4}\sqrt{7}$	3.146
l_∞ − norm	$\|\underline{x}\|_\infty$	3	3.00

If $\underline{x} = (x_1,\ x_2, ...,\ x_n)^{\mathsf{T}}$, $\underline{y} = (y_1,\ y_2, ...,\ y_n)^{\mathsf{T}} \in \mathbb{R}^n$, then the definitions for l_2 and l_∞ can be easily extended to consider distances between $\underline{x}$ and $\underline{y}$

$$\left\|\underline{x} - \underline{y}\right\|_2 = \left\{ \sum_{i=1}^{n} (x_i - y_i)^2 \right\}^{1/2}$$

$$\left\|\underline{x} - \underline{y}\right\|_{\infty} = \max_{1 \leq i \leq n} |x_i - y_i|$$

Having introduced the concept of a vector norm, we can now define the most appropriate form for our convergence criteria, by making use of the l_{∞} norm. One such relative condition makes use of the infinity norm l_{∞} as a test for convergence, where

$$\frac{\left\|\underline{x}^{(k)} - \underline{x}^{(k-1)}\right\|_{\infty}}{\left\|\underline{x}^{(k-1)}\right\|_{\infty}} < \varepsilon.$$

ε is the specified tolerance (> 0) and $\|\underline{x}\|_{\infty} = \max_{1 \leq i \leq n} |x_i|$.

Consider a square matrix. The *matrix norm* is a non-negative number such that

$$\|\cdot\| :^n \mathbb{R}^n \longrightarrow \mathbb{R}^+$$

and (as with vector norms) satisfies the following properties:

(i) $\|A\| \geq 0 \quad \forall A \in \mathbb{R}^n$

(ii) $\|A\| = 0$ iff $A = 0$

(iii) $\|\alpha A\| = |\alpha| \, \|\underline{x}\| \quad \forall \alpha \in \mathbb{R}$ and $A \in \mathbb{R}^n$

(iv) $\|A + B\| \leq \|A\| + \|B\| \quad \forall \underline{x},\ \underline{y} \in \mathbb{R}^n$ (Triangle inequality)

(v) $\|AB\| \leq \|A\| \, \|B\|$

Loosely speaking There are a number of ways to define a matrix norm

The 1-norm

$$\|A\|_1 = \max_j \sum_{i=1}^{n} \left| a_{ij} \right|$$

This is the maximum absolute column sum, i.e. sum the absolute values down each column and take the largest.

For example calculate the $\mathbf{1-norm}$ of

$$A = \begin{pmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{pmatrix}$$

$\|A\|_1 = \max [5 + 1 + 2, 4 + 2 + 1, 2 + 3 + 10] = \max [8, 7, 5] = 8.$

The infinity-norm

$$\|A\|_\infty = \max_{1\le i\le n}\sum_{j=1}^{n}\left|a_{ij}\right|$$

which is the absolute row sum, i.e. sum the absolute values along each row and take the largest.

Example: calculate the $\infty - \mathbf{norm}$ of

$$A = \begin{pmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{pmatrix}$$

$$\|A\|_\infty = \max\,[5+4+2, 1+2+3, 2+1+0] = \max\,[11, 6, 3] = 11.$$

The Euclidean-norm

$$\|A\|_E = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_{ij}\right)^2}$$

which is the square root of the sum of all the squares and is essentially the 'Pythagorean' length where the size of a vector is determined from the square root of the sum of squares of all the elements

Example: calculate $\|A\|_\infty$

$$A = \begin{pmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{pmatrix}$$

$$\|A\|_\infty = \sqrt{5^2 + 4^2 + 2^2 + ... + 2^2 + 1^2 + 0^2} = 8$$