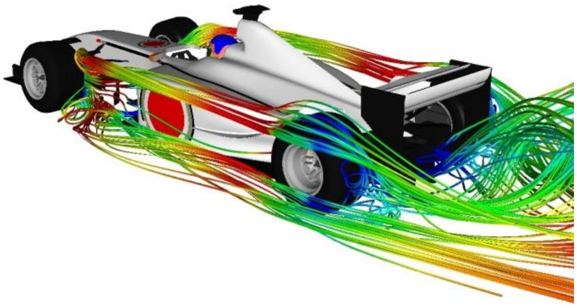
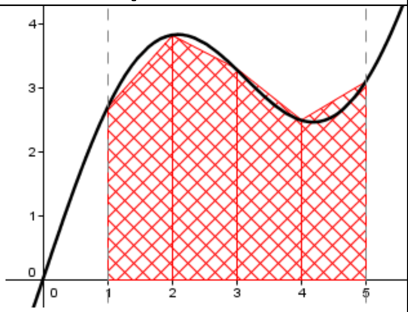
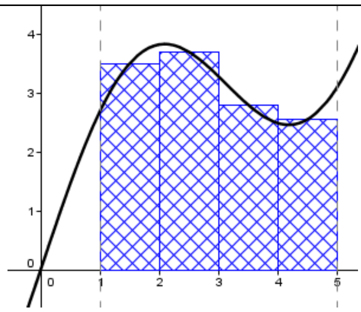
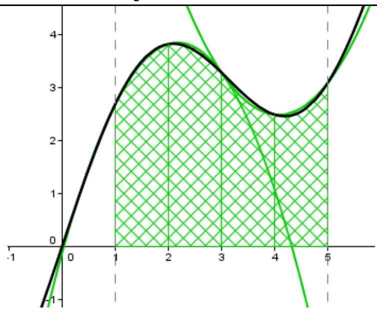
Numerical Integration

Some functions are very difficult to integrate, and some are actually impossible. In many cases (eg modelling airflow around a car), an exact answer may not even be strictly necessary. An approximation can be found using a numerical method:



The Rules

There are a great many methods for performing numerical integration, and since they each have advantages in terms of processing power required, or the type of functions they work well with, each has its place. The three rules you will be required to use are:

Rule:	**Trapezium Rule**	**Mid-ordinate Rule**	**Simpson's Rule**
			
Four strips:	$\int_1^5 x + 2\sin x \;\, dx \approx 12.4698$	$\int_1^5 x + 2\sin x \;\, dx \approx 12.5353$	$\int_1^5 x + 2\sin x \;\, dx \approx 12.5165$
Error:	*0.348%*	*0.176%*	*0.026%*
Eight strips:	$\int_1^5 x + 2\sin x \;\, dx \approx 12.5025$	$\int_1^5 x + 2\sin x \;\, dx \approx 12.5187$	$\int_1^5 x + 2\sin x \;\, dx \approx 12.5135$
Error:	*0.086%*	*0.043%*	*0.001%*

Accuracy

The accuracy of any of these methods depends to a certain extent on the function itself, but any of the methods can be improved by increasing the number of strips used. In general, for a given number of strips, Simpson's Rule is the most accurate of the three.

Notation

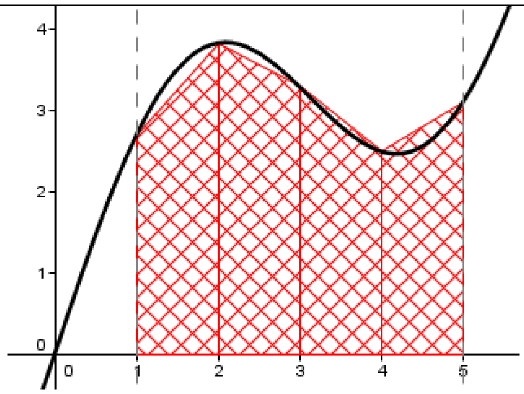
All these rules use the same basic notation:

y	The function, defined in terms of x, whose integral we are approximating.
a, b	The limits of the integration.
n	The number of strips we split the region into.
h	The 'height' of each strip (when viewed sideways – usually the horizontal width).
$x_0, x_1, \ldots, x_n$	The x ordinates: numbers between a and b at which we evaluate the function. Note: $x_0 = a$ and $x_n = b$. Also, since there are n strips, and there is an ordinate at the start and end of each, there are $n + 1$ ordinates in total.
$y_0, y_1, \ldots, y_n$	The y ordinates: the result of evaluating the function at each corresponding x ordinate. That is, $y_k = f(x_k)$.

Trapezium Rule

The concept:

Join a series of points on the curve with straight lines, and calculate the area of the trapezia formed.



Advantages/Limitations:

For a decreasing gradient (eg from $x = 1$ to $x = 3$ on the graph), the trapezium rule always gives an underestimate, while for an increasing gradient (eg from $x = 3$ to $x = 5$), it will give an overestimate.

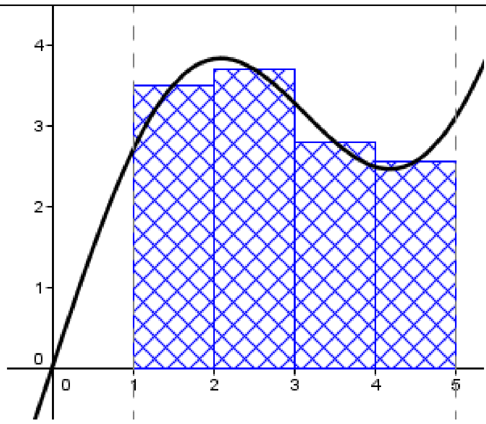
The formula:

$$\int_a^b y\,dx \approx \frac{1}{2}h\{(y_0 + y_n) + 2(y_1 + y_2 + \cdots + y_{n-1})\}, \text{ where } h = \frac{b-a}{n}$$

Mid-ordinate Rule

The concept:

Form rectangles whose height is equal to the height of the function in between each pair of x ordinates.



Advantages/Limitations:

While the rectangles give a good approximation for regions with roughly constant gradient (eg from $x = 3$ to $x = 4$, they are less accurate where the gradient changes rapidly (eg from $x = 4$ to $x = 5$).

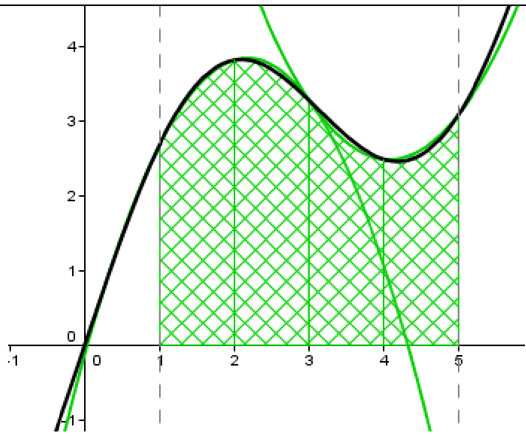
The formula:

$$\int_a^b y\,dx \approx h\left(y_{\frac{1}{2}} + y_{\frac{3}{2}} + \cdots + y_{n-\frac{3}{2}} + y_{n-\frac{1}{2}}\right), \text{ where } h = \frac{b-a}{n}$$

Simpson's Rule

The concept:

Approximate the curve as a series of quadratics, which, individually, are easy to integrate.



Advantages/Limitations:

This gives a better approximation than the previous two methods. However, since 3 points are required to define a parabola, an even number of strips is required for the formula to work.

The formula:

$$\int_a^b y\,dx \approx \frac{1}{3}h\{(y_0 + y_n) + 4(y_1 + y_3 + \cdots + y_{n-1}) + 2(y_2 + y_4 + \cdots + y_{n-2})\},$$

where $h = \frac{b-a}{n}$ and n is even

Derivation

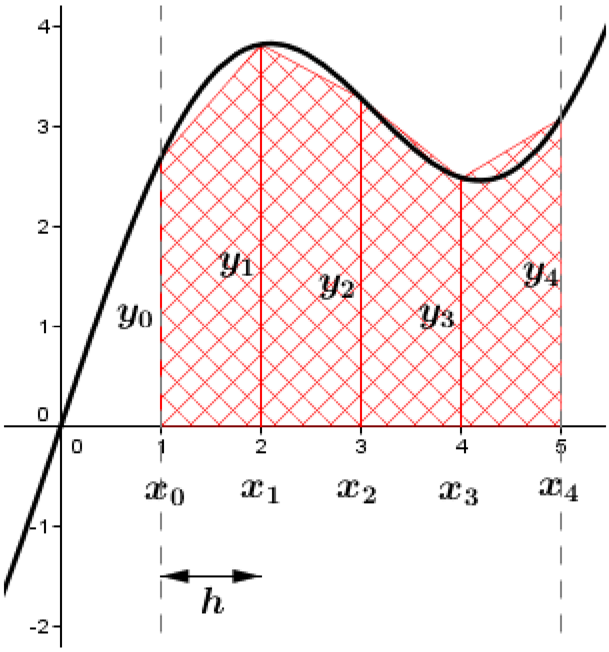
It is not a requirement of A level Mathematics to be able to derive these formulae, but the results below are included for reference.

Trapezium Rule

Consider an interval split into n strips.

It will have the following x ordinates: x_0, x_1, x_2, …, x_n, and corresponding y ordinates.

The 'height' of each trapezium is the horizontal width, given by h, where $h = \frac{b-a}{n}$ for $a = x_0$ and $b = x_n$ (the limits of integration).



Since the 'width' of each trapezium is given by the y ordinates at either side, the area of all trapezia will be:

$$\frac{y_0 + y_1}{2}h + \frac{y_1 + y_2}{2}h + \frac{y_2 + y_3}{2}h + \cdots + \frac{y_{n-1} + y_n}{2}h$$

Simplifying this, noting that, other than the first and last y ordinate, every one appears in the calculation of two trapezia, yields:

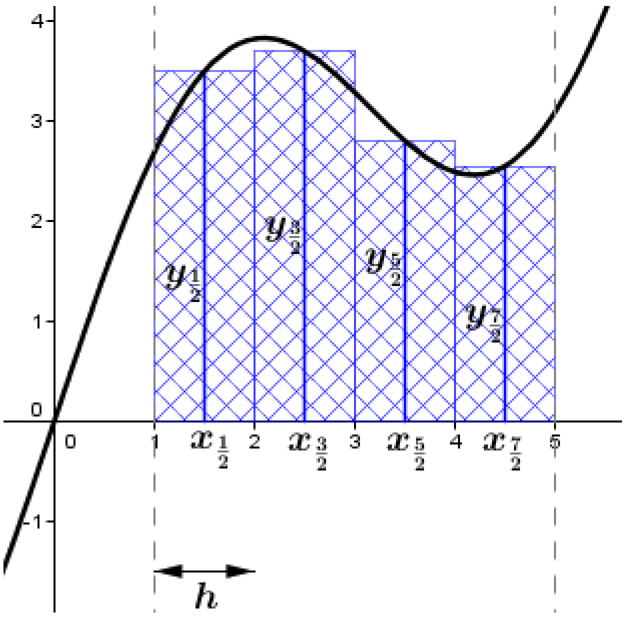
$$\frac{1}{2}h\{(y_0 + y_n) + 2(y_1 + y_2 + \cdots + y_{n-1})\}$$

Mid-ordinate Rule

Consider an interval split into n strips.

It will have the following x ordinates: x_0, x_1, x_2, …, x_n, and corresponding y ordinates.

The height of each rectangle is defined to be the y ordinate at the midpoint of each pair of x ordinates. The width is h, where $h = \frac{b-a}{n}$ for $a = x_0$ and $b = x_n$ (the limits of integration).



Totalling the area of all rectangles yields:

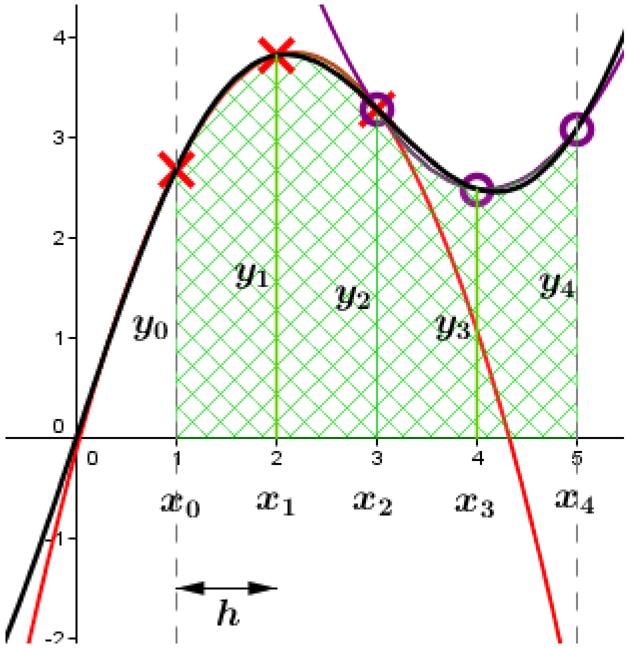
$$y_{\frac{1}{2}}h + y_{\frac{3}{2}}h + \cdots + y_{n-\frac{3}{2}}h + y_{n-\frac{1}{2}}h = h\left(y_{\frac{1}{2}} + y_{\frac{3}{2}} + \cdots + y_{n-\frac{3}{2}} + y_{n-\frac{1}{2}}\right)$$

Simpson's Rule

Consider an interval split into n strips.

It will have the following x ordinates: x_0, x_1, x_2, …, x_n, and corresponding y ordinates.

A parabola is mapped to pass through (x_0, y_0), (x_1, y_1) and (x_2, y_2), and another through (x_2, y_2), (x_3, y_3) and (x_4, y_4). This is repeated for the whole interval, mapping each subsequent ordinate pair with a parabolic curve. The process for doing this involves setting up simultaneous equations of the form $y = ax^2 + bx + c$, and solving for a, b and c. While tedious, this is achievable, and yields a series of quadratic functions which can readily be integrated to give a result, in terms of y_n, equal to:



$$\frac{y_0 h}{3} + \frac{4y_1 h}{3} + \frac{2y_2 h}{3} + \frac{4y_3 h}{3} + \cdots \frac{y_n h}{3}$$

Note that the odd terms represent the central point of each parabola, and the even terms represent the end points of each (including the first and last, which are only part of one parabola each). This simplifies to:

$$\frac{1}{3}h\{(y_0 + y_n) + 4(y_1 + y_3 + \cdots + y_{n-1}) + 2(y_2 + y_4 + \cdots + y_{n-2})\}$$

Alternative:

Another way of deriving Simpson's Rule involves a weighted average of the previous two rules. Essentially, the error terms in each one cancel out when combined in the right way:

$$S = \frac{2M + T}{3}$$

That is, Simpson's Rule can be found by combining the Mid-ordinate Rule and the Trapezium rule in the ratio $2:1$.

Error

As described earlier, the accuracy depends on a number of factors: the original function, the limits chosen, the number of strips and the choice of numerical method. A technical analysis of the three rules using Taylor Series yields the following information:

Rule:	**Trapezium Rule**	**Mid-ordinate Rule**	**Simpson's Rule**
Error:	$\frac{1}{12}(b-a)^3 f''(a) + O((b-a)^4)$	$-\frac{1}{24}(b-a)^3 f''(a) + O((b-a)^4)$	$\frac{1}{2880}(b-a)^5 \lvert f''''(\varepsilon)\rvert$

The precise meaning of these error terms is beyond the scope of this course, but it is worth noting that, in general, for a function with an increasing gradient, Trapezium Rule will tend to over-estimate and Mid-ordinate rule will tend to under-estimate (but by half as much). The appropriate combination of the two, in Simpson's Rule, eliminates this error term, giving a rule which will perfectly model anything up to a cubic, and have a proportionately lower error for any function of greater complexity.