Supporting Australian Mathematics Project

A guide for teachers – Years 11 and 12

Calculus: Module 15

Integration

Education Services Australia

AMSI
AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Integration – A guide for teachers (Years 11–12)

Principal author: Dr Daniel Mathews, Monash University

Peter Brown, University of NSW
Dr Michael Evans AMSI
Associate Professor David Hunt, University of NSW

Editor: Dr Jane Pitkethly, La Trobe University

Illustrations and web design: Catherine Tan, Michael Shaw

Full bibliographic details are available from Education Services Australia.

Published by Education Services Australia
PO Box 177
Carlton South Vic 3053
Australia

Tel: (03) 9207 9600
Fax: (03) 9910 9800
Email: info@esa.edu.au
Website: www.esa.edu.au

This publication is funded by the Australian Government Department of Education, Employment and Workplace Relations.

Supporting Australian Mathematics Project

Australian Mathematical Sciences Institute
Building 161
The University of Melbourne
VIC 3010
Email: enquiries@amsi.org.au
Website: www.amsi.org.au

Integration

Assumed knowledge

- The content of the module *Introduction to differential calculus*.
- The content of the module *Applications of differentiation*.

Motivation

From little things, big things grow.

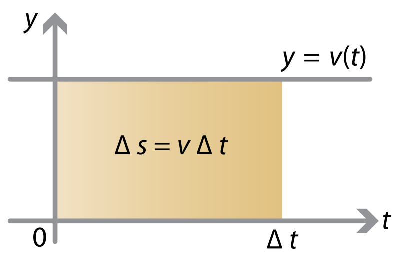
— Written and composed by Paul Kelly and Kev Carmody

Adding up a function in small pieces

You ride your bicycle for 5 minutes. Your bike is equipped with a fancy speedometer, and you keep your eye on it — so you know precisely how fast you are going at every instant for those 5 minutes. How far did you go in those 5 minutes?

This is not a very realistic situation, but let's suspend disbelief for now and try to imagine how you might answer this question.

The first thing you recall is $\Delta s = v\,\Delta t$. The change in your position, or **displacement**, Δs is given by your velocity v multiplied by the time taken Δt. However, this equation only holds if you travel at constant velocity — but your speedometer said otherwise!



When travelling at constant velocity, $\Delta s = v\,\Delta t$.

Trying to figure out how far you went on your bicycle, you recall how fast you were going at the 1-minute mark, at the 2-minute mark, and so on. Using this information and $\Delta s = v\,\Delta t$, you estimate how far you went in each minute. Adding up these estimates, you have an estimate for how far you went over the whole 5 minutes.

But this might not be a very accurate estimate — your speed might change a lot over the course of a minute! What about all those hills in the third minute?

Trying to be more accurate, you break your trip down into 30-second intervals, and recall your speed after 30 seconds, 60 seconds, 90 seconds, and so on. Then you can estimate how far you went in each 30-second interval, and add them up to estimate the total distance you travelled. This may be more accurate than the previous estimate, but your speed still varies a lot over 30-second intervals.

Trying to be ever more accurate, you break your journey down into 10-second intervals ... then 5-second intervals ... then 1-second intervals ... and further. If you really knew your speed at every instant of the journey, you could perform all these calculations, and see how far you travelled. Your calculations would lead you to ever more accurate estimates of how far you really went. Your better and better estimates will approach, in the limit, the exact distance travelled. You might imagine in this limit that you were 'adding up your velocity at every instant' to give the precise answer.

Of course, this whole situation is wildly unrealistic.[1] However, in principle, we have given a mathematical technique to answer the question of how far you went. Given the function $v(t)$, your velocity at time t, you could use these ideas to calculate exactly how far you went. This idea of 'adding up' the function at more and more closer and closer points is essentially the idea of *integration*.

We've previously seen *differentiation*, and you might recall that differentiation can answer a related question. If you know where you are on your bike (i.e., your displacement) at every instant, then you can work out your velocity at every instant. If $f(t)$ is how far you've travelled at time t, then the derivative $f'(t)$ gives your instantaneous velocity at time t. This is the opposite problem: given position, we used differentiation to figure out velocity; now, given velocity, we use integration to figure out position.

---

[1] If you tried to perform these calculations while on your bike, you likely would have crashed it. If you kept your eye on your speedometer at every instant, you certainly would have crashed it!

We could go on even more pedantically. The speedometer, like any measuring device, has uncertainties. Your memory cannot hold the full data of your speed at every instant in 10 minutes — there are infinitely many instants in that time! You are not really moving on a smooth surface; your path was really chaotically bumpy. Points on the Earth's surface are not fixed in place but are always moving about from climatic and geological phenomena. As it turns out, in the real world there is not really an objective notion of the 'time elapsed' or 'distance travelled': relativity theory tells us that time measurements depend on the observer. And to get an exactly accurate answer you would have to go down to the atomic level, where quantum mechanics precludes you from measuring distance and time accurately even in principle.

The real world is complicated! Luckily mathematics is much simpler.

It is an important fact that integrating a function — adding up the values over smaller and smaller intervals and seeing what the total is — turns out to be the *opposite of differentiation*. This is, as we will see, the idea of the *fundamental theorem of calculus*, the central idea of calculus.

Calculus is a crucial area of mathematics, necessary for understanding how quantities change in relation to each other, and for understanding almost every aspect of the physical world. Differentiation deals with how one quantity changes with respect to another, in the limit of small changes. But it is only half the story. Integration deals with the sum of these small changes, in the limit, adding them up and putting them back together again. Whether it's your speed on your bike, or biological populations, or chemical concentrations, or economic indicators, or environmental conditions, or physical phenomena — to understand how quantities in the world are related, we need to understand both differential and integral calculus.

Other questions

In the same vein as your bicycle with a speedometer, we present some more situations and questions below. These are not questions with exact answers, nor even the types of questions asked in secondary school mathematics; but they raise issues about how you might approach them mathematically. How might they also require us to add something up over many short intervals to get close to the exact answer — that is, by using integration? Some of them are contrived and unrealistic, but perhaps they will be useful to think about.

- Today is partly cloudy. Sometimes the sun is shining fully, sometimes the sun is partially blocked by clouds, and sometimes the clouds are thicker or thinner. You look at the sun and you know how much light is getting through at every instant. How hot is it? Do you need to look down to see if the ground is still wet from the rain last night? Will you get sunburnt?
- You are in control of a spacecraft heading towards Mars at 21 000 kilometres per hour. To land the spacecraft safely you need to slow down to zero. The spacecraft has rockets, and you know, when the rockets are turned on, the force they exert on the spacecraft. How long do you need to fire the rockets to stop the spacecraft?
- You are observing an ecological system in the forest. There are predators and prey. The number of predators increases when there is prey for them to eat, and the more prey around, the faster their numbers increase. The number of prey increases naturally but decreases when they are eaten by predators. How will the ecosystem evolve over time?

- You are at a power station and you can see, at every instant, the power that the station is producing. How much energy has been produced over the course of the day?
- You are lucky enough to have a smartphone with an accelerometer in it. That is, the phone can tell its acceleration at any instant. By using only its accelerometer, can the phone calculate which way it is going at any instant and how fast? Can it know at any instant where it is?
- You have a function $f(x)$ and its graph $y = f(x)$. What is the area under the graph?

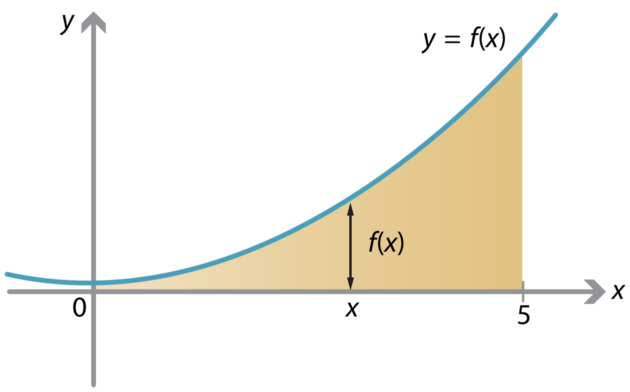
It is the last problem — an area under a graph — which is a purely mathematical one, and the most abstract. However, all of the other situations, it turns out, can be related to this type of mathematical problem. Integration is intimately connected to the area under a graph.

Content

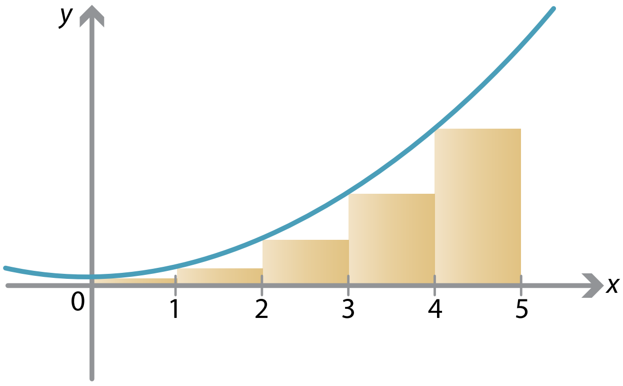
The area under a graph

Suppose you have a function $f: \mathbb{R} \to \mathbb{R}$ and its graph $y = f(x)$. You want to find the area under the graph. For now we'll assume that the graph $y = f(x)$ is always *above the x-axis*, and we'll estimate the area between the graph $y = f(x)$ and the x-axis. We set left and right endpoints and estimate the area between those endpoints.

Below is the graph of $f(x) = x^2 + 1$. We'll try to find the area under the graph $y = f(x)$ between $x = 0$ and $x = 5$, which is shaded.



As a first approximation, we can divide the interval $[0,5]$ into five subintervals of width 1, i.e., $[0,1]$, $[1,2]$, $\ldots$, $[4,5]$, and consider *rectangles* as shown on the following graph.

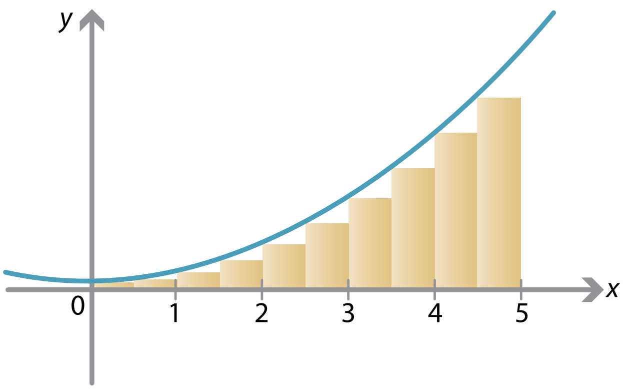


As we can see, the total area of these rectangles is a rough approximation to the area we are looking for, but a clear underestimate. The total area of the rectangles is calculated in the following table.

Approximating the area under the graph with 5 rectangles

Interval	Width of rectangle	Height of rectangle	Area of rectangle
$[0,1]$	1	$f(0) = 0^2 + 1 = 1$	1
$[1,2]$	1	$f(1) = 1^2 + 1 = 2$	2
$[2,3]$	1	$f(2) = 2^2 + 1 = 5$	5
$[3,4]$	1	$f(3) = 3^2 + 1 = 10$	10
$[4,5]$	1	$f(4) = 4^2 + 1 = 17$	17
		Total area	35

For a slightly better approximation, we can split the interval into 10 equal subintervals, each of length $\frac{1}{2}$. Using the same idea, we have the rectangles shown on the following graph. We clearly have a better estimate, but still an underestimate.



Approximating the area under the graph with 10 rectangles

Interval	Width of rectangle	Height of rectangle	Area of rectangle
$[0,\frac{1}{2}]$	$\frac{1}{2}$	$f(0)=0+1=1$	$\frac{1}{2}$
$[\frac{1}{2},1]$	$\frac{1}{2}$	$f(\frac{1}{2})=\frac{1}{4}+1=\frac{5}{4}$	$\frac{5}{8}$
$[1,\frac{3}{2}]$	$\frac{1}{2}$	$f(1)=1+1=2$	1
$[\frac{3}{2},2]$	$\frac{1}{2}$	$f(\frac{3}{2})=\frac{9}{4}+1=\frac{13}{4}$	$\frac{13}{8}$
$[2,\frac{5}{2}]$	$\frac{1}{2}$	$f(2)=4+1=5$	$\frac{5}{2}$
$[\frac{5}{2},3]$	$\frac{1}{2}$	$f(\frac{5}{2})=\frac{25}{4}+1=\frac{29}{4}$	$\frac{29}{8}$
$[3,\frac{7}{2}]$	$\frac{1}{2}$	$f(3)=9+1=10$	5
$[\frac{7}{2},4]$	$\frac{1}{2}$	$f(\frac{7}{2})=\frac{49}{4}+1=\frac{53}{4}$	$\frac{53}{8}$
$[4,\frac{9}{2}]$	$\frac{1}{2}$	$f(4)=16+1=17$	$\frac{17}{2}$
$[\frac{9}{2},5]$	$\frac{1}{2}$	$f(\frac{9}{2})=\frac{81}{4}+1=\frac{85}{4}$	$\frac{85}{8}$
		Total area	$\frac{325}{8}=40.625$

Splitting $[0,5]$ into more subintervals of smaller width, we can perform the same calculation and estimate the area. With more rectangles, the calculations become more tedious, but some results are given in the following table.

Approximating the area under the graph with more rectangles

Number of intervals	Width of each interval	Total area of rectangles
5	1	35
10	0.5	40.625
50	0.1	45.425
100	0.05	46.04375
1 000	0.005	46.6041875
10 000	0.0005	46.66041875
100 000	0.00005	46.66604166875
1 000 000	0.000005	46.6666041666875

With more and more thinner and thinner rectangles, the areas appear to be converging towards a limit. It turns out, as we will see, that the limit is $46\frac{2}{3}$.

The area under a curve is *defined* to be this limit. Although we have an intuitive notion of what area is, for a mathematically rigorous definition we need to use integration.

Estimates of area

Left-endpoint estimate

In the previous section, we estimated the area under the graph by splitting the interval $[0,5]$ into equal subintervals, and considering rectangles built on these subintervals. These rectangles had their *top-left corner* touching the curve $y = f(x)$. In other words, the height of the rectangle over a subinterval was the value of f at the *left endpoint* of that subinterval. For this reason, this method is known as the **left-endpoint estimate**.

To obtain a general formula for this estimate, suppose we have a real-valued function $f(x)$ defined on an interval $[a,b]$. In order to estimate the area under the graph $y = f(x)$ between $x = a$ and $x = b$, we split $[a,b]$ into n equal subintervals. Let these be

$$[x_0, x_1], \qquad [x_1, x_2], \qquad [x_2, x_3], \qquad \dots, \qquad [x_{n-1}, x_n].$$

So $x_0 = a$, $x_n = b$ and the real numbers $x_1, \dots, x_{n-1}$ are equally spaced between a and b. As the numbers are equally spaced, the width of each interval is $\frac{1}{n}(b-a)$. That is,

$$x_j - x_{j-1} = \frac{b-a}{n}, \qquad \text{for all } j = 1, 2, \dots, n.$$

We also write Δx for this width, the 'change in x' between successive endpoints.

Exercise 1

Show that x_j is given by

$$x_j = a + j\left(\frac{b-a}{n}\right) = a + j\,\Delta x, \qquad \text{for every } j = 0, 1, \dots, n.$$

Above the interval $[x_0, x_1]$, we have a rectangle whose width is Δx and whose height is $f(x_0)$. In general, above the interval $[x_{j-1}, x_j]$, we have a rectangle of width Δx and height $f(x_{j-1})$. The area of this rectangle is then $f(x_{j-1})\,\Delta x$. The total area estimate is therefore

$$\big[f(x_0) + f(x_1) + \cdots + f(x_{n-1})\big]\,\Delta x = \sum_{j=1}^{n} f(x_{j-1})\,\Delta x.$$

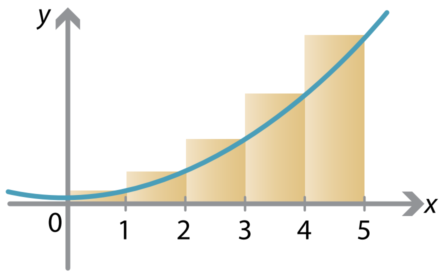
Right-endpoint estimate

Instead of setting the height of each rectangle to the value of f at the left endpoint of each subinterval, we can use the *right endpoint* and obtain the **right-endpoint estimate**.

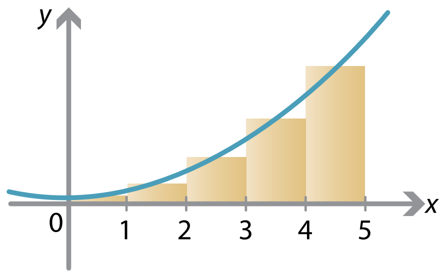
We again have a rectangle above each interval $[x_{j-1}, x_j]$ of width $\frac{1}{n}(b-a) = \Delta x$, but the height is now given at the right endpoint, i.e., the height is $f(x_j)$. So the total area estimate is

$$\big[f(x_1) + f(x_2) + \cdots + f(x_n)\big]\,\Delta x = \sum_{j=1}^{n} f(x_j)\,\Delta x.$$

Note that this formula only differs slightly from that for the left-endpoint estimate; the heights are obtained at x-values x_1 to x_n, rather than x_0 to x_{n-1}.



Right-endpoint estimate.



Midpoint estimate.

Example

Estimate the area under the graph of $y = x^3$ between 0 and 2, using the right-endpoint estimate with four subintervals. Is this an overestimate or an underestimate?

Solution

We split the interval $[0,2]$ into $n = 4$ equal subintervals $[0,\frac{1}{2}]$, $[\frac{1}{2},1]$, $[1,\frac{3}{2}]$, $[\frac{3}{2},2]$ and consider rectangles of width $\Delta x = \frac{1}{2}$. The right endpoints are $\frac{1}{2}$, 1, $\frac{3}{2}$ and 2, so letting $f(x) = x^3$, the estimate for the area is

$$\begin{aligned}\sum_{j=1}^{4} f(x_j)\,\Delta x &= \big[f(x_1) + f(x_2) + f(x_3) + f(x_4)\big]\,\Delta x \\ &= \Big[f\Big(\frac{1}{2}\Big) + f(1) + f\Big(\frac{3}{2}\Big) + f(2)\Big]\,\frac{1}{2} \\ &= \Big[\frac{1}{8} + 1 + \frac{27}{8} + 8\Big]\,\frac{1}{2} \\ &= \frac{25}{2}\cdot\frac{1}{2} = \frac{25}{4}.\end{aligned}$$

In each interval $[x_{j-1}, x_j]$, the function f takes its maximum value at x_j. So this is an overestimate.

Midpoint estimate

A third way to estimate the area under a graph is to set the height of each rectangle equal to the value of f at the *midpoint* of each subinterval, obtaining the **midpoint estimate**.

Specifically, above the interval $[x_0, x_1]$ there is again a rectangle of width Δx, but its height is the value of f at the midpoint of the interval. The midpoint is $\frac{1}{2}(x_0 + x_1)$, so the height is $f\big(\frac{1}{2}(x_0 + x_1)\big)$ and the area of the rectangle is $f\big(\frac{1}{2}(x_0 + x_1)\big)\,\Delta x$. Similarly, above each interval $[x_{j-1}, x_j]$, which has midpoint $\frac{1}{2}(x_{j-1} + x_j)$, there is a rectangle of width Δx and height $f\big(\frac{1}{2}(x_{j-1} + x_j)\big)$, and hence of area $f\big(\frac{1}{2}(x_{j-1} + x_j)\big)\,\Delta x$. Therefore the total area estimate is

$$\left[f\left(\frac{x_0 + x_1}{2}\right) + f\left(\frac{x_1 + x_2}{2}\right) + \cdots + f\left(\frac{x_{n-1} + x_n}{2}\right)\right]\Delta x = \sum_{j=1}^{n} f\left(\frac{x_{j-1} + x_j}{2}\right)\Delta x.$$

Exercise 2

Estimate the area under the graph of $y = x^2 + x$ between $x = 0$ and $x = 4$, using the midpoint estimate with two subintervals.

Exercise 3

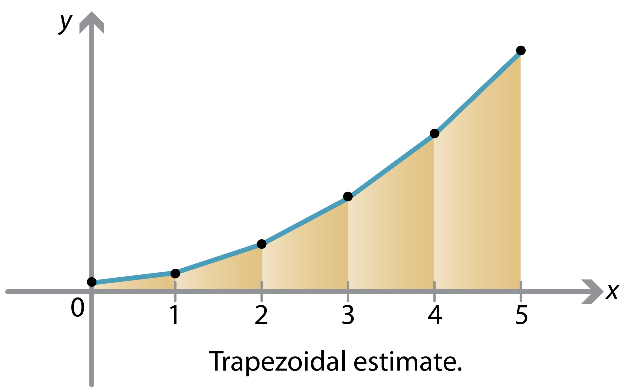
Assume as above that the graph of $y = f(x)$ is above the x-axis, and assume also that f is an **increasing function** (that is, $u \leq v$ implies $f(u) \leq f(v)$). Explain why, under these conditions, we have

left-endpoint estimate $\leq$ midpoint estimate $\leq$ right-endpoint estimate.

What happens if f is a decreasing function (that is, $u \leq v$ implies $f(u) \geq f(v)$)?

Trapezoidal estimate

We need not restrict ourselves to rectangles. Instead, we could plot the points on the graph $y = f(x)$ for each x_j, and join the dots. In this way, the function $f(x)$ is approximated by a **piecewise linear function**, that is, a sequence of straight line segments. The area under $y = f(x)$ is then approximated by *trapezia*. This estimate is known as the **trapezoidal estimate**.



Trapezoidal estimate.

Above the interval $[x_0, x_1]$ we have a trapezium of width $\frac{1}{n}(b-a) = \Delta x$. Its left side has height $f(x_0)$, and its right side has height $f(x_1)$. So the area of the trapezium is

$$\left(\frac{f(x_0)+f(x_1)}{2}\right)\Delta x.$$

Similarly, above the interval $[x_{j-1}, x_j]$, we have a trapezium of width Δx with two sides of height $f(x_{j-1})$ and $f(x_j)$, hence the area of the trapezium is $\frac{1}{2}\big(f(x_{j-1})+f(x_j)\big)\,\Delta x$. The total area estimate is

$$\left[\frac{f(x_0)+f(x_1)}{2}+\frac{f(x_1)+f(x_2)}{2}+\frac{f(x_2)+f(x_3)}{2}+\cdots+\frac{f(x_{n-1})+f(x_n)}{2}\right]\Delta x,$$

which simplifies to

$$\left[\frac{1}{2}f(x_0)+f(x_1)+f(x_2)+\cdots+f(x_{n-1})+\frac{1}{2}f(x_n)\right]\Delta x.$$

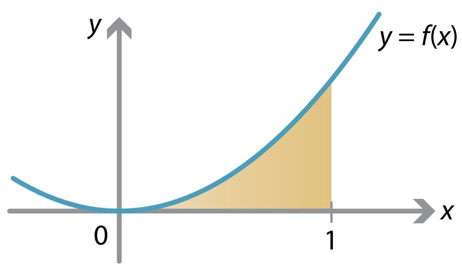
Exercise 4

Show that the trapezoidal estimate is the average of the left-endpoint estimate and the right-endpoint estimate.

Limits of estimates

In some simple cases, it is possible to calculate an area estimate with n rectangles (or trapezia) and explicitly take the limit as $n \to \infty$ to obtain the exact area. In the rest of this module, and in practice, we use other techniques to calculate areas, but it is worth seeing that sometimes a direct approach is possible.

For example, consider the function $f(x) = x^2$ and the area under the graph $y = f(x)$ between $x = 0$ and $x = 1$. We'll compute the right-endpoint estimate with n rectangles, and then take the limit as $n \to \infty$.



Dividing the interval $[0,1]$ into n subintervals $[x_0, x_1], \dots, [x_{n-1}, x_n]$, we have

$$\Delta x = \frac{1}{n} \qquad \text{and} \qquad x_j = \frac{j}{n}.$$

So the right-endpoint estimate is

$$\begin{aligned}\sum_{j=1}^{n} f(x_j)\,\Delta x &= \sum_{j=1}^{n} f\left(\frac{j}{n}\right)\frac{1}{n}\\ &= \sum_{j=1}^{n}\left(\frac{j}{n}\right)^2\frac{1}{n} = \frac{1}{n^3}\sum_{j=1}^{n} j^2.\end{aligned}$$

As it turns out, there is a formula for the sum of the first n squares:

$$\sum_{j=1}^{n} j^2 = 1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

The right-endpoint estimate with n rectangles then becomes

$$\begin{aligned} \frac{n(n+1)(2n+1)}{6n^3} &= \frac{2n^3+3n^2+n}{6n^3} \\ &= \frac{1}{6}\left(2+\frac{3}{n}+\frac{1}{n^2}\right). \end{aligned}$$

Taking the limit as $n \to \infty$, since $\frac{3}{n} \to 0$ and $\frac{1}{n^2} \to 0$, the area under the curve is

$$\lim_{n\to\infty} \frac{1}{6}\left(2+\frac{3}{n}+\frac{1}{n^2}\right) = \frac{1}{3}.$$

Note that the method used above relies crucially on the formula for the sum of squares. For more complicated functions, no such formula may be available, and this approach becomes impractical or impossible.

Exercise 5

Using the formula

$$\sum_{j=1}^{n} j = 1 + 2 + \cdots + n = \frac{n(n+1)}{2},$$

calculate the right-endpoint estimate for the area under the graph $y = x$ between $x = 0$ and $x = 1$. Take the limit as $n \to \infty$ to find the exact area. Confirm your answer using elementary geometry.

Definition of the integral

One can show that, as long as the function $f(x)$ is *continuous* (see below), then using any of these estimates, as we take more and more thinner and thinner rectangles (or trapezia) — i.e., as $n \to \infty$ or $\Delta x \to 0$ in the limit — the area estimates computed will converge to an answer for the exact area under the graph. Moreover, no matter which estimate we use, we will end up with the *same answer*.

For our purposes, **continuous** means that you can draw the graph $y = f(x)$ without taking your pen off the paper. Actually, it's sufficient that the graph is **piecewise continuous**, i.e., constructed out of several continuous functions.

For a function $f(x)$, considered between $x = a$ and $x = b$, the limit we obtain for the exact area is denoted by

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

and is called the **definite integral**, or simply the **integral**, of $f(x)$ with respect to x from $x = a$ to $x = b$. The symbol $\int$ is called an **integral sign**, and the numbers a, b are called the **terminals** or **endpoints** of the integral. The function $f(x)$ is called the **integrand**.

We can think of the integral as being the limit of a sum. Roughly speaking, the area estimates are given by expressions like $\sum f(x_j)\,\Delta x$ and in the limit these become $\int f(x)\,dx$.

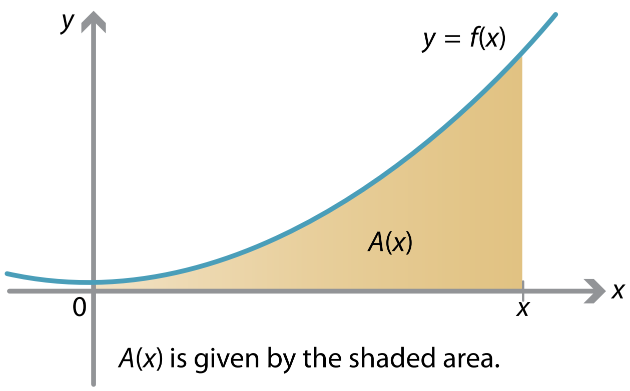
Area and the antiderivative

Let's return to our function $f(x) = x^2 + 1$, and our quest to find the area under the graph $y = f(x)$ between $x = 0$ and $x = 5$, i.e.,

$$\int_0^5 (x^2 + 1)\,dx.$$

As we've seen, we can estimate the area by many narrow rectangles, each over a subinterval of $[0,5]$, with width Δx and height given by a value of $f(x)$. As we take the limit of more and more narrower and narrower rectangles, we get a better approximation to the area.

To calculate the exact area, we introduce an **area function** A. For any fixed number $c > 0$, let $A(c)$ be the area under the graph $y = f(x)$ between $x = 0$ and $x = c$. So we can define the function $A(x)$ to be the area between 0 and x. Clearly $A(0) = 0$. We seek $A(5)$.



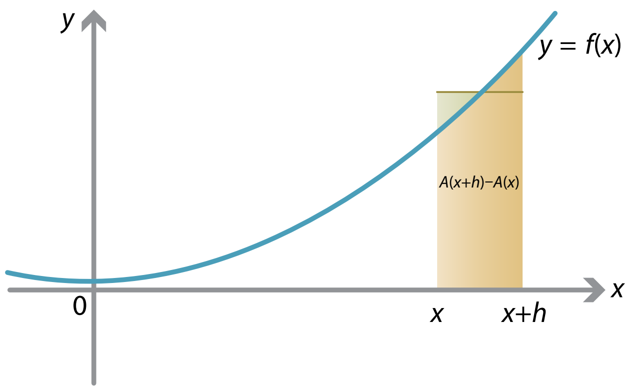
A(x) is given by the shaded area.

The fundamental idea is to consider the *derivative* of the area function A. What is $A'(x)$?

In fact, $A'(x) = f(x)$. We can see this geometrically. To differentiate $A(x)$ from first principles, we consider

$$\frac{A(x+h) - A(x)}{h}.$$

Now $A(x+h)$ is the area under the graph up to $x+h$, and $A(x)$ is the area under the graph up to x. Thus $A(x+h) - A(x)$ is the area under the graph between x and $x+h$.



If you were to 'level off' the area under the graph between x and $x+h$, you'd obtain a *rectangle* (as shown in the graph above) with width h and height equal to the average value of f; dividing the area by h gives the average value:

$$\begin{aligned}\frac{A(x+h) - A(x)}{h} &= \frac{\text{area of rectangle}}{\text{width of rectangle}} \\ &= \text{height of rectangle} \\ &= \text{average value of } f \text{ over the interval } [x, x+h].\end{aligned}$$

If h becomes very small, then the interval $[x, x+h]$ approaches the single point x. And the average value of f over this interval must approach $f(x)$. So we have

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

In other words,

$$A'(x) = f(x),$$

and $A(x)$ is an *antiderivative* of $f(x)$.

Returning again to our example $f(x) = x^2 + 1$, what then is the area function $A(x)$? Its derivative must be $f(x) = x^2 + 1$. We're looking for a function A such that

$$A'(x) = x^2 + 1 \qquad \text{and} \qquad A(0) = 0.$$

Thinking back to derivatives, we can see that the antiderivative

$$A(x) = \frac{1}{3}x^3 + x$$

satisfies these criteria. Hence the total area is

$$A(5) = \frac{125}{3} + 5 = \frac{140}{3} = 46\frac{2}{3},$$

confirming our earlier estimates. We can write this as

$$\int_0^5 f(x)\,dx = A(5) - A(0) = \frac{140}{3}.$$

Antiderivatives and indefinite integrals

In order to calculate integrals, we now see that it's important to be able to find antiderivatives of functions. An **antiderivative** of a function $f(x)$ is a function whose derivative is equal to $f(x)$. That is, if $F'(x) = f(x)$, then $F(x)$ is an antiderivative of $f(x)$.

Importantly, *antiderivatives are not unique.* A given function can have many antiderivatives. For instance, the following functions are all antiderivatives of x^2:

$$\frac{x^3}{3}, \qquad \frac{x^3}{3} + 1, \qquad \frac{x^3}{3} - 42, \qquad \frac{x^3}{3} + \pi.$$

However, any two antiderivatives of a given function *differ by a constant.* This allows us to write a general formula for any antiderivative of x^2:

$$\frac{x^3}{3} + c, \qquad \text{where } c \text{ is a constant.}$$

Since, as we saw in the previous section, antiderivatives are intimately connected with areas, we write this antiderivative as

$$\int x^2\,dx = \frac{x^3}{3} + c$$

and call this an indefinite integral.

An **indefinite integral** is an integral written without terminals; it simply asks us to find a general antiderivative of the integrand. It is not one function but a *family of functions*, differing by constants; and so the answer must have a '+ constant' term to indicate all antiderivatives.

A definite integral, on the other hand, is an integral with terminals. It is a number which measures an area under a graph between the terminals.

Exercise 6

Find the derivatives of the following functions:

a for n, c any real constants with $n \neq -1$,

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

b for a, b, c, n any real constants with $a \neq 0$ and $n \neq -1$,

$$f(x) = \frac{1}{a(n+1)}(ax+b)^{n+1} + c.$$

From our knowledge of differentiation, and in particular the above exercise, we have the following formulas for antiderivatives.

Basic antiderivatives

Function	General antiderivative	Comment
x^n	$\frac{1}{n+1}x^{n+1} + c$	for n, c any real constants with $n \neq -1$
$(ax+b)^n$	$\frac{1}{a(n+1)}(ax+b)^{n+1} + c$	for a, b, c, n any real constants with $a \neq 0$, $n \neq -1$

Example

Find

1 $\int 1\, dx$ **2** $\int (3x-1)^4\, dx$.

Solution

1 We can observe directly that x is an antiderivative of 1, or we can use the above rule for the antiderivative of x^n when $n = 0$. Either way, the general antiderivative is $x + c$ and so $\int 1\, dx = x + c$, where c is a constant.

2 Using the above rule for the antiderivative of $(ax+b)^n$ with $a = 3$, $b = -1$ and $n = 4$, we obtain

$$\int (3x-1)^4\, dx = \frac{1}{15}(3x-1)^5 + c, \qquad \text{where } c \text{ is a constant.}$$

Exercise 7

Find the indefinite integrals

a $\displaystyle\int \frac{1}{\sqrt{3-2x}}\,dx$ **b** $\displaystyle\int \frac{1}{(3-2x)^2}\,dx.$

The following theorem gives some useful rules for computing integrals.

Theorem (Linearity of integration)

a *If f and g are continuous functions, then*

$$\int \big(f(x) \pm g(x)\big)\,dx = \int f(x)\,dx \pm \int g(x)\,dx.$$

b *If f is a continuous function and k is a real constant, then*

$$\int kf(x)\,dx = k\int f(x)\,dx.$$

This theorem allows us to compute an antiderivative by treating a function term-by-term and factoring out constants, as the next example illustrates.

Example

Find

$$\int (3x^3 - 4x^2 + 2)\,dx.$$

Solution

Using the above rules, we can rewrite the integral as follows:

$$\begin{aligned}\int (3x^3 - 4x^2 + 2)\,dx &= \int 3x^3\,dx - \int 4x^2\,dx + \int 2\,dx\\ &= 3\int x^3\,dx - 4\int x^2\,dx + 2\int 1\,dx\\ &= \frac{3x^4}{4} - \frac{4x^3}{3} + 2x + c.\end{aligned}$$

Note. You might think that, since there are several indefinite integrals on the second line of these equations, there should be several different constants in the answer. But since the sum of several constants is still a constant, we can just write a single constant.

Exercise 8

Find

$$\int (3x^2 + \sqrt[3]{x})\,dx.$$

Exercise 9

Prove the theorem above (linearity of integration) using similar rules for differentiation.

Calculating areas with antiderivatives

A general method to find the area under a graph $y = f(x)$ between $x = a$ and $x = b$ is given by the following important theorem.

Theorem (Fundamental theorem of calculus)

Let $f(x)$ be a continuous real-valued function on the interval $[a, b]$. Then

$$\int_a^b f(x)\,dx = \Big[F(x)\Big]_a^b = F(b) - F(a),$$

where $F(x)$ is any antiderivative of $f(x)$.

The notation $[F(x)]_a^b$ is just a shorthand to substitute $x = a$ and $x = b$ into $F(x)$ and subtract; it is synonymous with $F(b) - F(a)$.

Note that *any* antiderivative of $f(x)$ will work in the above theorem. Indeed, if we have two different antiderivatives $F(x)$ and $G(x)$ of $f(x)$, then they must differ by a constant, so $G(x) = F(x) + c$ for some constant c. Then we have to get the same answer whether we use F or G, since

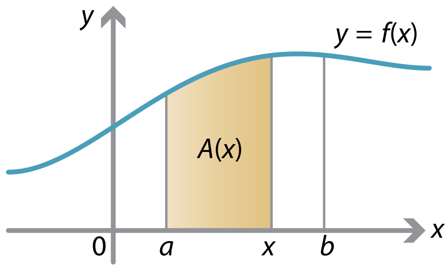
$$\begin{aligned} G(b) - G(a) &= \big(F(b) + c\big) - \big(F(a) + c\big) \\ &= F(b) - F(a). \end{aligned}$$

We give a proof of the theorem, assuming our previous statements:

- the derivative of the area function is $f(x)$
- any two antiderivatives of $f(x)$ differ by a constant.

Proof

As before, let $A(c)$ be the area under the graph $y = f(x)$ between $x = a$ and $x = c$. So we have an area function $A(x)$ which measures the area between a and x. We must find $A(b)$, the area from a to b. The derivative of the area function $A(x)$ is $f(x)$, so $A(x)$ is an antiderivative of $f(x)$. As both $A(x)$ and $F(x)$ are antiderivatives of $f(x)$, they must differ by a constant, i.e., $A(x) = F(x) + K$ for some constant K. Since $A(a) = 0$, this implies $F(a) + K = 0$. Hence $K = -F(a)$, and it follows that $A(x) = F(x) - F(a)$. Therefore the desired integral is equal to $A(b) = F(b) - F(a)$. □



We can return to our original problem and solve it using the fundamental theorem.

Example

Find

$$\int_0^5 (x^2 + 1)\,dx.$$

Solution

An antiderivative of $x^2 + 1$ is $\frac{1}{3}x^3 + x$, so we have

$$\begin{aligned}\int_0^5 (x^2+1)\,dx &= \left[\frac{1}{3}x^3 + x\right]_0^5 \\ &= \left(\frac{1}{3}\cdot 5^3 + 5\right) - \left(\frac{1}{3}\cdot 0^3 + 0\right) \\ &= \frac{125}{3} + 5 = \frac{140}{3} = 46\frac{2}{3}.\end{aligned}$$

Exercise 10

Find

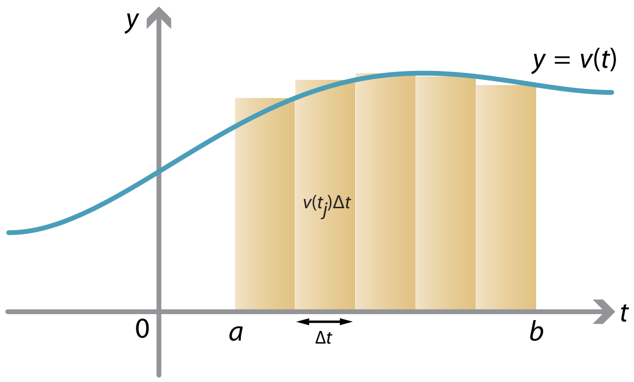
$$\int_0^8 (3x^2 + \sqrt[3]{x})\,dx.$$

Velocity–time graphs

Let us now return to our original motivating question with the bike and speedometer. Let $v(t)$ denote your velocity on your bike at time t, and we now calculate your total change in position between time $t = a$ and $t = b$. As mentioned earlier, if $v(t)$ is a constant function, and so you are travelling at constant velocity, then your total change in position, or displacement, is $\Delta s = v\,\Delta t$.

Our idea of estimating your distance travelled by breaking your trip up into 1-minute intervals is nothing more than an area estimate. We may divide the time interval $[a, b]$ into n subintervals $[t_0, t_1], [t_1, t_2], \ldots, [t_{n-1}, t_n]$, each of width $\Delta t = \frac{1}{n}(b-a)$. We can estimate your velocity over the interval $[t_{j-1}, t_j]$ by the constant $v(t_j)$ (i.e., using the right-endpoint estimate); over this interval of time, you travelled a distance of approximately $v(t_j)\,\Delta t$. The total distance travelled, then, is approximately

$$\sum_{j=1}^{n} v(t_j)\,\Delta t.$$



The right-endpoint estimate for area under $y = v(t)$ estimates displacement.

In the limit, as we take more and more shorter and shorter time intervals, we obtain the exact change in position (displacement) as

$$\int_a^b v(t)\,dt.$$

Therefore, the total displacement is the *area under the velocity–time graph* $y = v(t)$.

Note that this integrand is expressed as a function of t, and dt indicates that we integrate over t. If we used $v(x)$ and dx instead, the meaning would be identical:

$$\int_a^b v(t)\,dt = \int_a^b v(x)\,dx.$$

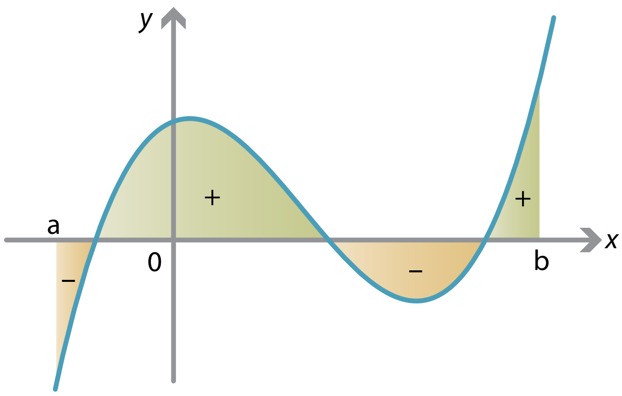
The variable t (or x) just indexes the values of the integrand; it is only used for the integral's own bookkeeping purposes. It is known as a *dummy variable*, analogous to the dummy variable in summation notation, like the j in $\sum_{j=1}^{5} j^2$.

Areas above and below the axis

So far we have only considered functions $f(x)$ whose graphs are *above* the x-axis, i.e., $f(x) > 0$. But, of course, a graph can go below the x-axis. We now consider this situation.

When we computed the right-endpoint estimate, we approximated the area under the graph $y = f(x)$ by rectangles of width $\Delta x = \frac{1}{n}(b-a)$ and height $f(x_j)$. The important point now is that, if $f(x_j)$ is negative, then $f(x_j)\,\Delta x$ is negative and we obtain the *negative* of the area of the rectangle. So when the graph of $y = f(x)$ is below the x-axis, our method approximates the *negative* of the area between the graph and the x-axis.

In regions where the graph of $y = f(x)$ is above the x-axis (i.e., $f(x) > 0$), a definite integral calculates the area between the graph and the x-axis. In regions where the graph is below the x-axis (i.e., $f(x) < 0$), the integral calculates the *negative* of the area between the graph and the x-axis. A general function $f(x)$ is sometimes positive and sometimes negative, so the integral calculates the **signed area**, that is, the total area *above* the x-axis *minus* the total area *below* the x-axis.



The integral counts areas above/below the x-axis as positive/negative.

When we have a velocity–time graph $y = v(t)$, the signed area has a natural interpretation. The regions where the graph is below the t-axis occur when $v(t) < 0$, i.e., the velocity is *negative* and you are going *backwards*. When you are going backwards, your change in position is negative. Your total displacement is given by your total forwards movements minus your total backwards movements; these correspond to the regions above and below the t-axis. Thus *signed area* corresponds to displacement. The *total area* enclosed between the graph and the axis, on the other hand, is the total distance travelled, forwards and backwards.

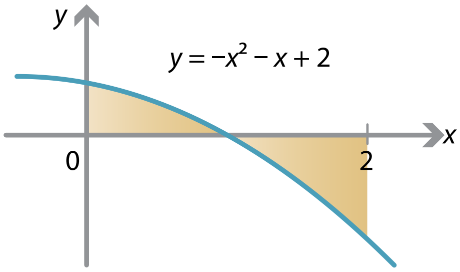
In general, to find the total area enclosed between a graph and the axis, we find where the graph crosses the axis and compute the areas separately. See the following example.

Example

1 Find $\displaystyle\int_0^2 (-x^2 - x + 2)\,dx$.

2 Find the area of the shaded region.

3 If your velocity at time t is given by $v(t) = -t^2 - t + 2$, then what is your displacement and your distance travelled between time $t = 0$ and $t = 2$?



Solution

1
$$\int_0^2 (-x^2 - x + 2)\,dx = \left[-\frac{1}{3}x^3 - \frac{1}{2}x^2 + 2x\right]_0^2$$
$$= \left(-\frac{8}{3} - 2 + 4\right) - (0) = -\frac{2}{3}.$$

2 The area is partly above and partly below the x-axis. Factorising

$$-x^2 - x + 2 = -(x-1)(x+2),$$

we find the x-intercept is at $x = 1$. Therefore the total area shown is

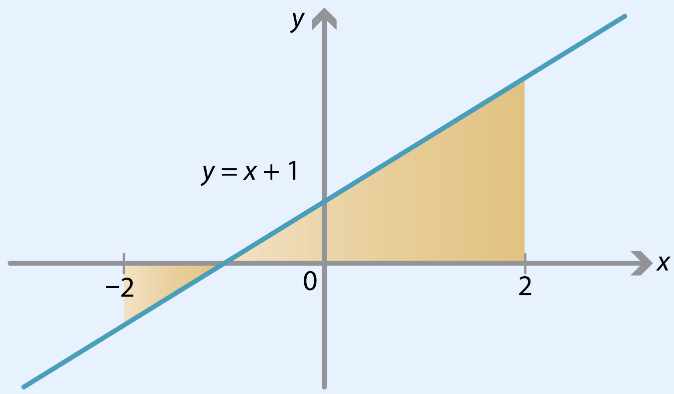
$$\int_0^1 (-x^2 - x + 2)\,dx - \int_1^2 (-x^2 - x + 2)\,dx$$
$$= \left[-\frac{1}{3}x^3 - \frac{1}{2}x^2 + 2x\right]_0^1 - \left[-\frac{1}{3}x^3 - \frac{1}{2}x^2 + 2x\right]_1^2$$
$$= \left(\left(-\frac{1}{3} - \frac{1}{2} + 2\right) - (0)\right) - \left(\left(-\frac{8}{3} - 2 + 4\right) - \left(-\frac{1}{3} - \frac{1}{2} + 2\right)\right)$$
$$= \left(\frac{7}{6} - 0\right) - \left(-\frac{2}{3} - \frac{7}{6}\right)$$
$$= \frac{7}{6} - \left(-\frac{11}{6}\right) = 3.$$

3 Your displacement is $\int_0^2 v(t)\,dt$, which was computed in part 1, and is $-\frac{2}{3}$ (i.e., you end up $\frac{2}{3}$ backwards of where you started). Your total distance travelled is given by the area calculated in part 2, which is 3.

Beware! Computing integrals often involves many fractions, subtractions and negative signs, as in the above example! It is good practice to carefully bracket all terms, as shown.

Exercise 11

Find the area shown, both by integration and directly.



Exercise 12

a Compute $\displaystyle\int_{-2}^{2} (x^3 - x)\,dx$.

b What is the total area enclosed between the graph of $y = x^3 - x$ and the x-axis, from $x = -2$ to $x = 2$?

Properties of the definite integral

Definite integrals obey rules similar to those for indefinite integrals. The following theorem is analogous to one for indefinite integrals.

Theorem (Linearity of integration)

a *If f and g are continuous functions on the interval $[a, b]$, then*

$$\int_a^b \big(f(x) \pm g(x)\big)\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx.$$

b *If f is a continuous function on $[a, b]$, and k is a real constant, then*

$$\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx.$$

The endpoints on a definite integral obey the following theorem.

Theorem (Additivity of integration)

Let $f\colon [a, b] \to \mathbb{R}$ be continuous, where a, b are real numbers. Let c be a real number between a and b. Then

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

As with many mathematical statements, it's useful to understand these two theorems both algebraically (in terms of antiderivatives) and geometrically.

For instance, we can think of the second theorem (additivity of integration) as saying geometrically that, if we consider the signed area between $y = f(x)$ and the x-axis from $x = a$ to $x = b$, then this signed area is equal to the sum of the signed area from $x = a$ to $x = c$ and that from $x = c$ to $x = b$. Alternatively, if we let $F(x)$ be an antiderivative of $f(x)$, we can regard the theorem as just expressing that

$$F(b) - F(a) = \big(F(c) - F(a)\big) + \big(F(b) - F(c)\big).$$

This piece of algebra and the fundamental theorem of calculus together give a rigorous proof of the theorem.

Exercise 13

Find a geometric interpretation of part (b) of the first theorem of this section (linearity of integration). You may assume the graph of $y = f(x)$ lies above the x-axis. Also find an interpretation in terms of antiderivatives.

We have stated the second theorem (additivity of integration) so that $a < c < b$. But in fact, this theorem works when a, b, c are in *any order*, as long as f, g are defined and continuous over the appropriate intervals. We just have to make sense of integrals which have their terminals 'in the wrong order'. When $a > b$, we define

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx.$$

We can think of this as saying that, when x goes *backwards* from a to b, we count areas as *negative*. In our area estimates, $\Delta x = \frac{1}{n}(b - a)$ is the negative of the rectangle widths. This fits with our previous definitions, as you can show in the following exercise.

Exercise 14

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, and let $F(x)$ be an antiderivative of $f(x)$. Using the fundamental theorem of calculus, show that for any real numbers a, b (even when $a > b$),

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Exercise 15

Find $\displaystyle\int_5^0 (-2x - 3)\,dx$.

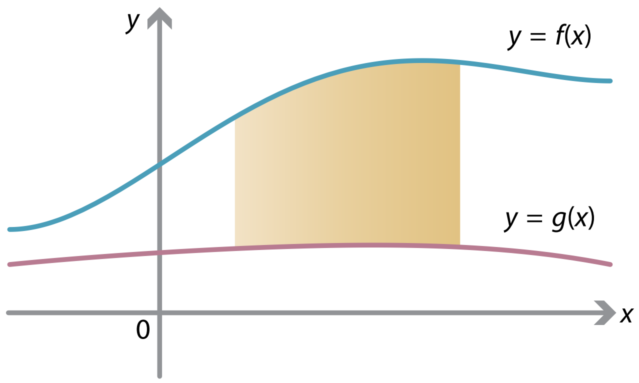
(Careful with signs! Should your answer be positive or negative?)

Area between two curves

So far we have only found areas between the graph $y = f(x)$ and the x-axis. In general we can find the area enclosed between *two* graphs $y = f(x)$ and $y = g(x)$. If $f(x) > g(x)$, then the desired area is that which is below $y = f(x)$ but above $y = g(x)$, which is

$$\int_a^b \big(f(x) - g(x)\big)\, dx.$$

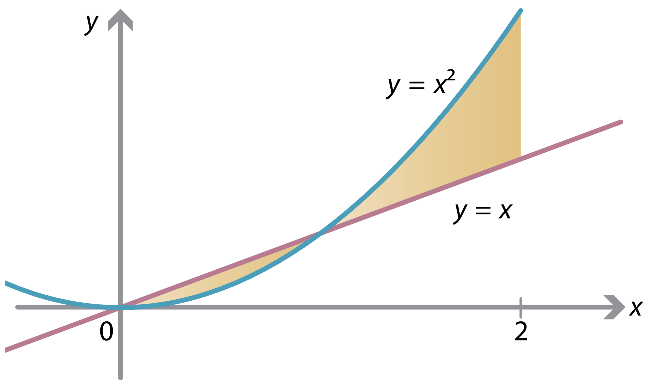
This formula works regardless of whether the graphs are above or below the x-axis, as long as the graph of $f(x)$ is above the graph of $g(x)$. (Can you see why?)



In general, two graphs $y = f(x)$ and $y = g(x)$ may cross. Sometimes $f(x)$ may be larger and sometimes $g(x)$ may be larger. In order to find the area enclosed by the curves, we can find where $f(x)$ is larger, and where $g(x)$ is larger, and then take the appropriate integrals of $f(x) - g(x)$ or $g(x) - f(x)$ respectively.

Example

Find the area enclosed between the graphs $y = x^2$ and $y = x$, between $x = 0$ and $x = 2$.



Solution

Solving $x^2 = x$, we see that the two graphs intersect at $(0,0)$ and $(1,1)$. In the interval $[0,1]$, we have $x \geq x^2$, and in the interval $[1,2]$, we have $x^2 \geq x$. Therefore the desired area is

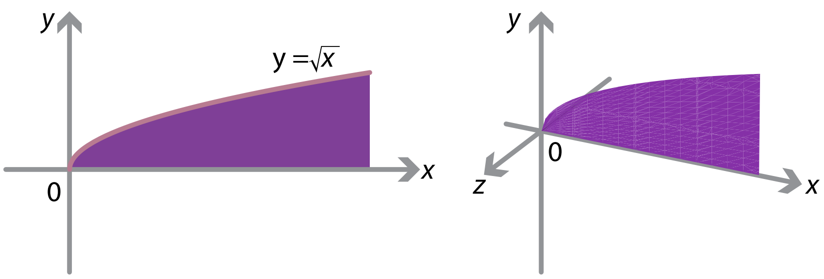
$$\begin{aligned}\int_0^1 (x - x^2)\,dx + \int_1^2 (x^2 - x)\,dx &= \left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^1 + \left[\frac{1}{3}x^3 - \frac{1}{2}x^2\right]_1^2 \\ &= \left(\left(\frac{1}{2} - \frac{1}{3}\right) - (0)\right) + \left(\left(\frac{8}{3} - 2\right) - \left(\frac{1}{3} - \frac{1}{2}\right)\right) \\ &= \left(\frac{1}{6} - 0\right) + \left(\frac{2}{3} - \left(-\frac{1}{6}\right)\right) \\ &= \frac{1}{6} + \frac{5}{6} = 1.\end{aligned}$$

Links forward

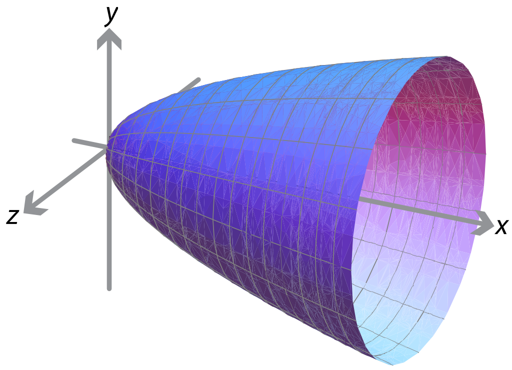
Integration and three dimensions

So far we have used integration to calculate areas, which are 2-dimensional. Integration can also be used to obtain 3-dimensional volumes.

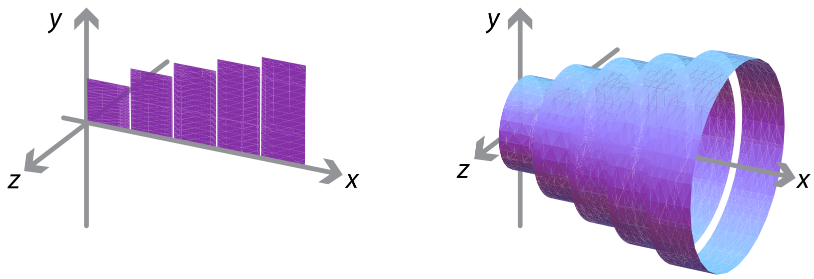
For instance, take a graph $y = f(x)$ in the x–y plane, and consider adding a third dimension: points in 3-dimensional space have coordinates (x, y, z), and there is now also a z-axis. The figures below show the graph of $y = \sqrt{x}$, for $0 \leq x \leq 5$, in the plane and then adding a z-axis.



We can now consider rotating the curve around the x-axis. As we do, the curve sweeps out a *surface*, and the area under the curve sweeps out a 3-dimensional *solid*. The solid has rotational symmetry about the x-axis and is called a **solid of revolution**.



We can estimate the area under the graph $y = f(x)$ in the plane by subdividing the x-interval and forming rectangles. Rotating these rectangles about the x-axis, we estimate the *volume* of the solid of revolution by *cylinders*.



Use our usual notation: we consider the graph of $y = f(x)$ between $x = a$ and $x = b$, and we divide the interval $[a,b]$ into n subintervals $[x_0, x_1]$, ..., $[x_{n-1}, x_n]$ of width Δx. As we have seen, for the right-endpoint estimate, the rectangles have width Δx and height $f(x_j)$, and hence area $f(x_j)\,\Delta x$. Similarly, the cylinders have radius $f(x_j)$ and width Δx, and hence volume $\pi f(x_j)^2\,\Delta x$. We can estimate the volume of the solid of revolution as $\sum_{j=1}^{n} \pi f(x_j)^2\,\Delta x$. Taking the limit as $n \to \infty$, the volume of the solid of revolution is

$$\int_a^b \pi f(x)^2\,dx.$$

More generally, if we have a solid bounded by the two parallel planes $x = a$ and $x = b$, and at each value of x in $[a,b]$ the cross-sectional area of the solid is $A(x)$, then the volume of the solid is

$$\int_a^b A(x)\,dx.$$

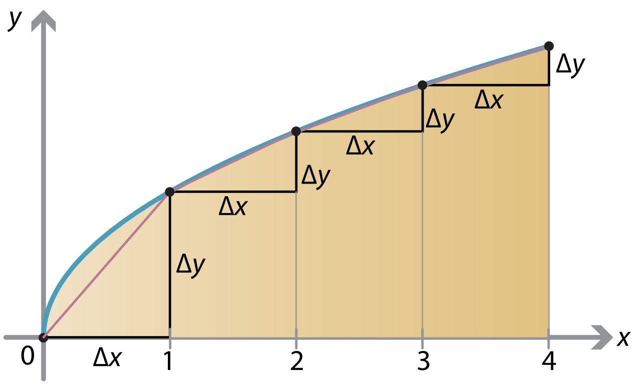
Using this idea, it's possible to derive the formulas for the volumes of pyramids, prisms, and other solids. In higher mathematics courses one studies 3-dimensional (and higher-dimensional) volumes using *multiple integrals*.

Arc length

How long is a piece of string? If the piece of string is given as a graph $y = f(x)$ then its length can be calculated using integration.

As we have now seen many times, we consider the curve between $x = a$ and $x = b$ and split the x-interval into n subintervals $[x_0, x_1]$, ..., $[x_{n-1}, x_n]$ of length Δx. Over each subinterval there is an arc segment that travels Δx to the right and Δy upwards, where $\Delta y = f(x_j) - f(x_{j-1})$. Approximating the arc by straight line segments, the length of each line segment is $\sqrt{(\Delta x)^2 + (\Delta y)^2}$. This gives an estimate for the length of the curve as

$$\sum_{j=1}^{n} \sqrt{(\Delta x)^2 + (\Delta y)^2} = \sum_{j=1}^{n} \sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}\, \Delta x.$$



Taking the limit as $n \to \infty$, the length of the curve is given by

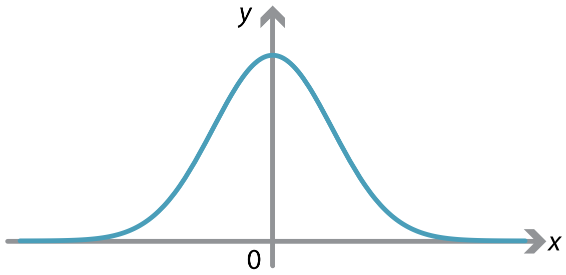
$$\int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_a^b \sqrt{1 + f'(x)^2}\, dx.$$

Difficulties with antiderivatives

We saw in the modules on differentiation that many functions can be differentiated. Indeed, knowing how to differentiate elementary functions, and using the product, chain and quotient rules, we can find the derivative of most functions we come across.

So far in this module, we have been able to antidifferentiate every function we encountered. But as it turns out, there are some relatively simple functions for which the antiderivative has no simple formula. Although there are many techniques and tricks for finding antiderivatives, there are no rules as simple as the product, chain and quotient rules for differentiation.

For an interesting and important example, consider the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = e^{-x^2}$.



Graph of $f(x) = e^{-x^2}$.

It is easy to differentiate $f(x)$ using the chain rule: $f'(x) = -2xe^{-x^2}$. However, the antiderivative of $f(x)$ has no simple formula. Yet the integral of $f(x)$ is very important, because it essentially describes the *normal distribution*, which is fundamental to probability and statistics.

It is possible to show that the integral of $f(x)$, over the whole real line, is

$$\int_{-\infty}^{+\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

This may seem to be a surprising result; it is not clear at all where the π came from.[2] Some clever techniques and higher mathematics, which are beyond our scope here, such as *area integrals* and *polar coordinates*, can be used to obtain this answer.

History and applications

The differential and integral calculus, as we have discussed it in these modules, using antiderivatives and the fundamental theorem of calculus, was developed in the 17th century by such mathematicians as Pierre de Fermat, James Gregory, Isaac Barrow, Isaac Newton and Gottfried Wilhelm Leibniz — but principally by Newton and Leibniz. However, some precursors to integration were known in ancient times.

Several ancient Greek philosophers and mathematicians considered the problem of calculating the area bounded by a curve. In the 5th and 4th centuries BCE, Antiphon and

[2] Lord Kelvin, the mathematical physicist and engineer, once joked that a mathematician is a person for whom this equation is as obvious as the equation $2+2=4$ is to the average person!

Eudoxus developed the idea of finding the area bounded by a curve by inscribing polygons in it, now known as the **method of exhaustion**. The idea is to inscribe infinitely many polygons within the curve, which do not overlap and which cover the entire area, so that the sum of the areas of polygons approaches the area bounded by the curve.

The method of exhaustion was perhaps taken furthest by Archimedes, who used it in the 3rd century BCE to perform calculations of areas and volumes bounded by circles, ellipses, spheres and cylinders. By calculating the area of the polygons inside and outside a circle, he was able to show that $\frac{223}{71} < \pi < \frac{22}{7}$. He was also able to compute the area bounded by a chord in a parabola.

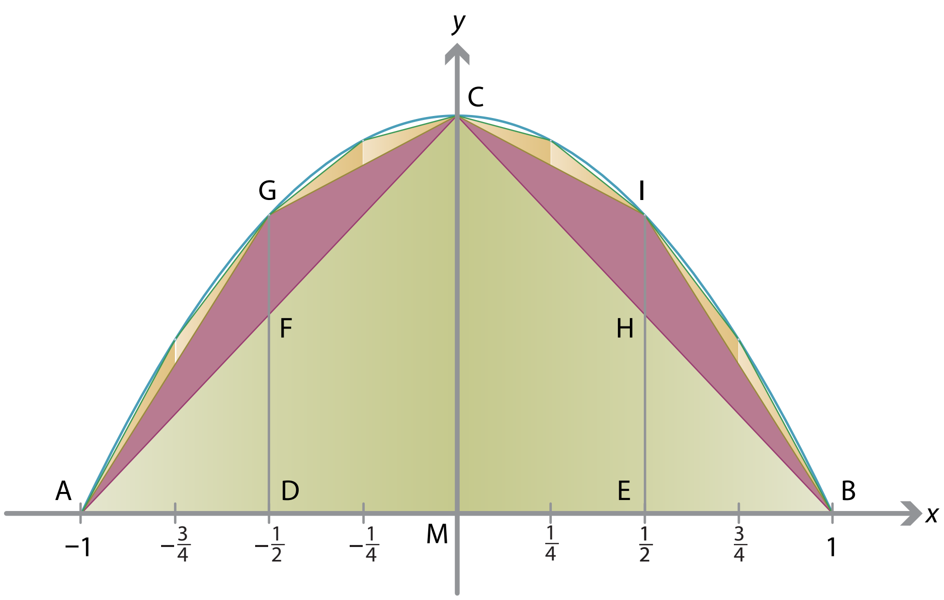
Though our notation is very different, Archimedes would have been able to calculate an integral like

$$\int_{-1}^{1} (1 - x^2)\, dx.$$

It's interesting to compare our 'modern' (17th century CE) technique to Archimedes'. We would compute this area as

$$\int_{-1}^{1} (1 - x^2)\, dx = \left[x - \frac{x^3}{3}\right]_{-1}^{1} = \left(1 - \frac{1}{3}\right) - \left((-1) - \left(-\frac{1}{3}\right)\right) = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}.$$

Although Archimedes did not have coordinate geometry, or even symbolic algebra, to work with, we can give a rough outline of his ideas in his work *Quadrature of the parabola*. (For further details, see the article *Archimedes' quadrature of the parabola revisited* listed in the *References* section.)



Using Archimedes' technique to find the area under a parabola.

We begin by inscribing a triangle in the parabola. Let $A=(-1,0)$ and $B=(1,0)$ be the intersection points of the chord (i.e., the x-axis) with the parabola, as shown in the diagram. Let $M=(0,0)$ be the midpoint of AB, and let $C=(0,1)$. The area of $\triangle ABC$ is 1, as it has width 2 and height 1.

We next inscribe two more triangles in the parabola, built on $\triangle ABC$. Let $D=(-\frac{1}{2},0)$ be the midpoint of AM and let $E=(\frac{1}{2},0)$ be the midpoint of BM. We draw a line from D parallel to the axis of the parabola (i.e., the y-axis), intersecting AC at F and intersecting the parabola at $G=(-\frac{1}{2},\frac{3}{4})$. Similarly, we draw a line from E parallel to the parabola's axis, intersecting BC at H and intersecting the parabola at $I=(\frac{1}{2},\frac{3}{4})$. We consider the two triangles $\triangle AGC$ and $\triangle BIC$.

To compute the area of $\triangle AGC$, we consider the segment FG. The area is given by half the 'height' FG times the width of $\triangle AGC$ in the x-direction. The width (in the x-direction) is half that of $\triangle ABC$, i.e., 1. To compute FG, note that $\triangle ADF$ is similar to $\triangle AMC$, so $\frac{DF}{AD}=\frac{MC}{AM}$ and hence $DF=AD\cdot\frac{MC}{AM}=\frac{1}{2}$. Thus $FG=DG-DF=\frac{3}{4}-\frac{1}{2}=\frac{1}{4}$; the 'height' of $\triangle AGC$ is a quarter that of $\triangle ABC$. As $\triangle AGC$ has half the width and a quarter the height of $\triangle ABC$, it has an eighth the area. Similarly, we can compute that $\triangle BIC$ has an eighth the area of $\triangle ABC$.

As we proceed, we inscribe more and more triangles inside the parabola. In the next step we inscribe four triangles, after that we inscribe eight triangles. At the nth stage, we inscribe 2^n triangles in the gaps between previous triangles. Archimedes was able to show that, at each stage, the triangles all have half the width and a quarter the height of the triangles from the previous stage, and hence an eighth the area.

The total area of all the triangles is given by a geometric series:

$$1+\frac{2}{8}+\frac{2^2}{8^2}+\frac{2^3}{8^3}+\cdots=1+\frac{1}{4}+\frac{1}{4^2}+\frac{1}{4^3}+\cdots.$$

Using the formula

$$1+r+r^2+\cdots=\frac{1}{1-r}$$

for the limiting sum of a geometric series, the total area is

$$\frac{1}{1-\frac{1}{4}}=\frac{4}{3}.$$

In this way, Archimedes succeeded in computing the area under the parabola.

Note. Archimedes actually did better than this. He used his technique to compute the area bounded by *any* chord in a parabola; the chord considered here is a special one (perpendicular to the axis of the parabola).

Appendix

A comparison of area estimates, and Simpson's rule

We have mentioned several different ways of estimating the area under a curve: left-endpoint, right-endpoint, midpoint and trapezoidal estimates. It's interesting to compare them.

Recall that we consider the function $f(x)$ over the interval $[a,b]$ and divide the interval into n subintervals $[x_0,x_1],[x_1,x_2],\dots,[x_{n-1},x_n]$ each of width $\Delta x = \frac{1}{n}(b-a)$.

All of the estimates discussed give the approximate area under $y = f(x)$ in the form of a sum of products of values of f with the width Δx. In fact, the left-endpoint, right-endpoint and trapezoidal estimates are all of the form

$$\left[c_0 f(x_0) + c_1 f(x_1) + c_2 f(x_2) + \cdots + c_{n-1} f(x_{n-1}) + c_n f(x_n)\right] \Delta x.$$

(The midpoint rule is slightly different, since it evaluates the function f at the midpoints of subintervals.) The coefficients c_j are compared in the following table.

The coefficients in the formulas for area estimates

Estimate	c_0	c_1	c_2	c_3	c_4	$\dots$	c_{n-2}	c_{n-1}	c_n
left endpoint	1	1	1	1	1	$\dots$	1	1	0
right endpoint	0	1	1	1	1	$\dots$	1	1	1
trapezoidal	$\frac{1}{2}$	1	1	1	1	$\dots$	1	1	$\frac{1}{2}$
Simpson	$\frac{1}{3}$	$\frac{4}{3}$	$\frac{2}{3}$	$\frac{4}{3}$	$\frac{2}{3}$	$\dots$	$\frac{2}{3}$	$\frac{4}{3}$	$\frac{1}{3}$

There is another estimate, called **Simpson's rule**, which takes a slightly different set of coefficients, as shown in the table. It requires that the number of subintervals n is *even*. The coefficients start at $\frac{1}{3}$, then alternate between $\frac{4}{3}$ and $\frac{2}{3}$, before ending again at $\frac{1}{3}$. (It is as if every second coefficient in the list for the trapezoidal estimate donated $\frac{1}{6}$ to its neighbours.)

Explicitly, the estimate from Simpson's rule for the area under $y = f(x)$ between $x = a$ and $x = b$ is

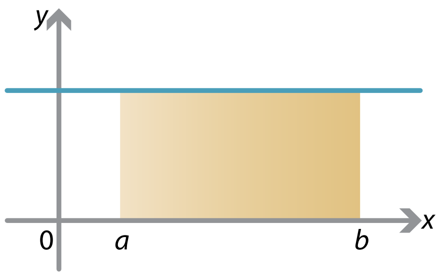
$$\left[\frac{1}{3}f(x_0)+\frac{4}{3}f(x_1)+\frac{2}{3}f(x_2)+\frac{4}{3}f(x_3)+\frac{2}{3}f(x_4)+\cdots+\frac{2}{3}f(x_{n-2})+\frac{4}{3}f(x_{n-1})+\frac{1}{3}f(x_n)\right]\Delta x.$$

Although the coefficients may appear somewhat bizarre, Simpson's rule almost always gives a much more accurate answer than any of the other estimates mentioned. We'll see why in the next section.

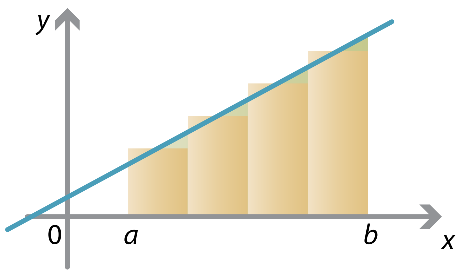
Exactness of area estimates

Consider some simple functions $f(x)$ — constant, linear, quadratic — and see how the various estimates fare in approximating the integral $\int_a^b f(x)\,dx$.

First suppose $f(x)$ is a constant function. In this case, we obtain an exactly correct answer for the area under the graph from the left-endpoint, right-endpoint, midpoint or trapezoidal estimate. The area under $y = f(x)$ is itself a rectangle; breaking it up into rectangles or trapezia, we still have the exactly correct area. This is true regardless of the number of subintervals n: we could split $[a,b]$ into just one subinterval, or a million; in all cases we obtain the exactly correct answer.



Area under a constant function.



Area under a linear function, correctly estimated by midpoint or trapezoidal estimate.

Next suppose $f(x)$ is a (non-constant) linear function. In this case the left- and right-endpoint estimates will definitely *not* give you the correct answer. However, the midpoint and trapezoidal estimates will give you the exactly correct answer. The midpoint estimate builds rectangles which have exactly the same amount of area above and below the line. And breaking the area under $y = f(x)$ into trapezia gives the exactly correct area. Again this is true whatever number of subintervals we take.

However, when $f(x)$ is a quadratic function, in general none of the left-endpoint, right-endpoint, midpoint or trapezoidal estimates will give the correct area. Let's consider the function $f(x) = x^2$, and see how the midpoint and trapezoidal estimates fare. We'll just take one subinterval, $n = 1$, so $\Delta x = b - a$.

First, the midpoint estimate M is just

$$M = f\Big(\frac{a+b}{2}\Big)\,\Delta x = \Big(\frac{1}{4}a^2 + \frac{1}{2}ab + \frac{1}{4}b^2\Big)(b-a).$$

Next, the trapezoidal estimate T (again with only one subinterval) is

$$T = \Big(\frac{1}{2}f(a) + \frac{1}{2}f(b)\Big)\,\Delta x = \Big(\frac{1}{2}a^2 + \frac{1}{2}b^2\Big)(b-a).$$

The correct answer, on the other hand, is

$$\int_a^b x^2\,dx = \left[\frac{x^3}{3}\right]_a^b = \frac{1}{3}b^3 - \frac{1}{3}a^3 = \frac{1}{3}\left(a^2+ab+b^2\right)(b-a).$$

Here we used the factorisation $b^3 - a^3 = (b-a)(a^2+ab+b^2)$.

Note that M and T contain very similar terms to the correct answer. Indeed, *we can combine the midpoint and trapezoidal estimates to get the exact answer*! Observe that

$$\frac{1}{3}(2M+T) = \frac{1}{3}(b-a)\left(a^2+ab+b^2\right) = \int_a^b x^2\,dx.$$

This combination of the midpoint and trapezoidal rules will in fact give us an exactly correct answer for the integral of *any* quadratic function.

This new area estimate is nothing other than Simpson's rule. Writing out $\frac{1}{3}(2M+T)$ for a general function $f(x)$, from the definitions of the midpoint and trapezoidal rules, we have

$$\begin{aligned}\frac{2}{3}M + \frac{1}{3}T &= (b-a)\left(\frac{1}{6}f(a) + \frac{2}{3}f\left(\frac{a+b}{2}\right) + \frac{1}{6}f(b)\right)\\ &= \frac{b-a}{2}\left(\frac{1}{3}f(a) + \frac{4}{3}f\left(\frac{a+b}{2}\right) + \frac{1}{3}f(b)\right).\end{aligned}$$

This pattern of coefficients $\frac{1}{3}$, $\frac{4}{3}$, $\frac{1}{3}$ is exactly Simpson's rule with two subintervals; if we take $\frac{1}{3}(2M+T)$ for the midpoint and trapezoidal estimates with n subintervals, we obtain Simpson's rule with $2n$ subintervals.

Simpson's rule is very accurate because it gets quadratic integrals exactly right. It does this by taking the midpoint and trapezoidal estimates — the previous best estimates, giving exactly correct answers for linear integrals — and combining them. We can also think of Simpson's rule as approximating a function by parabolic segments and computing the integrals of the parabolic segments. It turns out that Simpson's rule actually gets cubic integrals exactly right as well!

Functions integrable and not

The theory of integration that we have discussed was developed rigorously by Bernhard Riemann, and this type of integral is known as the **Riemann integral**.

As we have seen, the Riemann integral calculates the area under the curve $y = f(x)$ over the interval $[a,b]$ by approximating it with n rectangles (or trapezia) of equal width, and in the limit, as $n \to \infty$, the estimate approaches the true area.

Riemann integration works for any function $f : [a, b] \to \mathbb{R}$ that is *continuous* — roughly, when the graph $y = f(x)$ can be drawn without taking your pen off the paper. It also works for any *piecewise continuous* function — where you may take your pen off the paper, but only a finite number of times.

A function $f(x)$ where the area estimates (based on n rectangles or trapezia) approach the true integral as $n \to \infty$ is called **Riemann integrable**. The previous paragraph, then, says that any piecewise continuous function is integrable.[3]

As an example of a non-integrable function, consider the following rather strange function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

This function is discontinuous. In fact, near *any* real number x, there is a rational number arbitrarily close, and there is an irrational number arbitrarily close. So $f(x)$ is discontinuous at *every* real number x.

This example of a non-integrable function is quite pathological, and unlikely to arise in real-world situations. There is, however, a theory of integration that can integrate even functions like these. It is called *Lebesgue integration* and it is developed in university mathematics courses.

Answers to exercises

Exercise 1

Since $x_0, x_1, \dots, x_j$ are all separated by distance $\frac{1}{n}(b-a) = \Delta x$, we have

$$\begin{aligned} x_j &= (x_j - x_{j-1}) + (x_{j-1} - x_{j-2}) + \cdots + (x_1 - x_0) + x_0 \\ &= j\,\Delta x + x_0 \\ &= a + j\,\Delta x. \end{aligned}$$

In the last equality we used the fact that $x_0 = a$, by definition.

[3] It turns out that a function may have infinitely many discontinuities and still be Riemann integrable. The *Riemann–Lebesgue theorem* says that a function $f : [a, b] \to \mathbb{R}$ is integrable if and only if it is bounded and its set of discontinuities has *measure zero*. The concept of measure zero is, however, well beyond the scope of these notes.

Exercise 2

The two subintervals are $[0,2]$ and $[2,4]$, each of width $\Delta x = 2$, with respective midpoints 1 and 3. So the area estimate is

$$\big[f(1) + f(3)\big]\,\Delta x = (2+12)\cdot 2 = 28.$$

Exercise 3

Over each subinterval $[x_{j-1}, x_j]$, the three estimates give rectangles of the same width but different heights. For the left-endpoint, midpoint and right-endpoint estimates, respectively, the heights are $f(x_{j-1})$, $f\big(\frac{1}{2}(x_{j-1}+x_j)\big)$ and $f(x_j)$. Now by definition $x_{j-1} < x_j$ and obviously the midpoint $\frac{1}{2}(x_{j-1}+x_j)$ lies between them, so

$$x_{j-1} < \frac{x_{j-1}+x_j}{2} < x_j.$$

As f is an increasing function, we have

$$f(x_{j-1}) \le f\left(\frac{x_{j-1}+x_j}{2}\right) \le f(x_j).$$

It follows that the rectangles for the left-endpoint estimate are shorter than the rectangles for the midpoint estimate, which are shorter again than those for the right-endpoint estimate. This gives the desired inequalities.

When f is decreasing, we have the opposite inequalities

$$f(x_{j-1}) \ge f\left(\frac{x_{j-1}+x_j}{2}\right) \ge f(x_j),$$

which imply

$$\text{left-endpoint estimate} \;\ge\; \text{midpoint estimate} \;\ge\; \text{right-endpoint estimate}.$$

Exercise 4

With the interval $[a,b]$ divided into n subintervals $[x_{j-1}, x_j]$, for $j = 1, \dots, n$, and $\Delta x = \frac{1}{n}(b-a)$, as usual, we have

$$\begin{aligned}
\text{left-endpoint estimate} &= \big[f(x_0) + f(x_1) + \cdots + f(x_{n-1})\big]\,\Delta x,\\
\text{right-endpoint estimate} &= \big[f(x_1) + f(x_2) + \cdots + f(x_n)\big]\,\Delta x,\\
\text{average} &= \frac{1}{2}\Big(\big[f(x_0) + \cdots + f(x_{n-1})\big]\,\Delta x + \big[f(x_1) + \cdots + f(x_n)\big]\,\Delta x\Big)\\
&= \frac{1}{2}\Big(f(x_0) + 2f(x_1) + \cdots + 2f(x_{n-1}) + f(x_n)\Big)\,\Delta x\\
&= \Big[\frac{1}{2}f(x_0) + f(x_1) + \cdots + f(x_{n-1}) + \frac{1}{2}f(x_n)\Big]\,\Delta x\\
&= \text{trapezoidal estimate.}
\end{aligned}$$

Exercise 5

Divide $[0,1]$ into n subintervals $[x_0,x_1],\dots,[x_{n-1},x_n]$, so $x_j = \frac{j}{n}$ and $\Delta x = \frac{1}{n}$. The right-endpoint estimate for the area is then

$$\sum_{j=1}^{n} f(x_j)\,\Delta x = \sum_{j=1}^{n} f\Big(\frac{j}{n}\Big)\,\frac{1}{n} = \sum_{j=1}^{n} \Big(\frac{j}{n}\Big)\,\frac{1}{n} = \frac{1}{n^2}\sum_{j=1}^{n} j.$$

Using the formula $\sum_{j=1}^{n} j = \frac{1}{2}n(n+1)$ gives the right-endpoint estimate as

$$\frac{n(n+1)}{2n^2} = \frac{n^2+n}{2n^2} = \frac{1}{2}\Big(1+\frac{1}{n}\Big).$$

Taking the limit as $n \to \infty$, the exact area under the curve is

$$\lim_{n\to\infty} \frac{1}{2}\Big(1+\frac{1}{n}\Big) = \frac{1}{2}.$$

The area under the curve is just a right-angled triangle with height 1 and base 1, so we can confirm that its area is $\frac{1}{2}\cdot 1 \cdot 1 = \frac{1}{2}$.

Exercise 6

a $f'(x) = x^n$

b $f'(x) = (ax+b)^n$.

Exercise 7

a We can rewrite this as $\int (-2x+3)^{-\frac{1}{2}}\,dx$, so the integrand is of the form $(ax+b)^n$. We obtain $-(-2x+3)^{\frac{1}{2}} + c = -\sqrt{3-2x} + c$.

b Rewriting as $\int (-2x+3)^{-2}\,dx$, the integrand is again of the form $(ax+b)^n$, and we obtain

$$\frac{1}{2}(-2x+3)^{-1} + c = \frac{1}{2(3-2x)} + c.$$

Exercise 8

$$\int \big(3x^2 + x^{\frac{1}{3}}\big)\,dx = x^3 + \frac{3}{4}x^{\frac{4}{3}} + c$$

Exercise 9

a Let $F(x)$ be an antiderivative of $f(x)$ and let $G(x)$ be an antiderivative of $g(x)$, so $\int f(x)\,dx = F(x) + c_1$ and $\int g(x)\,dx = G(x) + c_2$, where c_1, c_2 are constants. Then the derivative of $(F \pm G)(x)$ is $F'(x) \pm G'(x) = f(x) \pm g(x)$, hence $F \pm G$ is an antiderivative of $f \pm g$, and $\int \big(f(x) \pm g(x)\big)\,dx = F(x) \pm G(x) + c$, where c is a constant. Noting that the sum of two constants is a constant gives the desired equality.

b Let $F(x)$ be an antiderivative of $f(x)$, so $\int f(x)\,dx = F(x)+c$, where c is a constant. Then the derivative of $(kF)(x)$ is $kF'(x) = kf(x)$, hence kF is an antiderivative of kf and $\int kf(x)\,dx = kF(x)+C$, where C is a constant. As a constant times k is another constant, we have the desired equality.

Exercise 10

$$\int_0^8 (3x^2+\sqrt[3]{x})\,dx = \left[x^3+\frac{3}{4}x^{\frac{4}{3}}\right]_0^8 = \left(512+\frac{3}{4}\cdot 16\right)-(0+0) = 512+12 = 524$$

Exercise 11

The graph $y=x+1$ crosses the x-axis at $x=-1$, so the desired area is

$$\begin{aligned}\int_{-1}^{2}(x+1)\,dx - \int_{-2}^{-1}(x+1)\,dx &= \left[\frac{1}{2}x^2+x\right]_{-1}^{2} - \left[\frac{1}{2}x^2+x\right]_{-2}^{-1}\\ &= \left((2+2)-\left(\frac{1}{2}-1\right)\right)-\left(\left(\frac{1}{2}-1\right)-(2-2)\right)\\ &= \left(4-\left(-\frac{1}{2}\right)\right)-\left(-\frac{1}{2}-0\right) = \frac{9}{2}+\frac{1}{2} = 5.\end{aligned}$$

Alternatively, we can compute the areas of the triangles directly. The left triangle has height 1 and base 1, and so area $\frac{1}{2}$. The right triangle has height 3 and base 3, and so area $\frac{9}{2}$. The total area is 5.

Exercise 12

a $\displaystyle\int_{-2}^{2}(x^3-x)\,dx = \left[\frac{1}{4}x^4-\frac{1}{2}x^2\right]_{-2}^{2} = (4-2)-(4-2) = 0$

b Factorising $x^3-x=(x+1)x(x-1)$, we see the graph has intercepts at $x=-1,0,1$. It is above the x-axis for $-1<x<0$ and $x>1$, and below the x-axis for $x<-1$ and $0<x<1$. We compute four separate integrals:

$$\int_{-2}^{-1}(x^3-x)\,dx = \left[\frac{1}{4}x^4-\frac{1}{2}x^2\right]_{-2}^{-1} = \left(\frac{1}{4}-\frac{1}{2}\right)-(4-2) = -\frac{1}{4}-2 = -\frac{9}{4},$$

$$\int_{-1}^{0}(x^3-x)\,dx = \left[\frac{1}{4}x^4-\frac{1}{2}x^2\right]_{-1}^{0} = (0)-\left(\frac{1}{4}-\frac{1}{2}\right) = \frac{1}{4},$$

$$\int_{0}^{1}(x^3-x)\,dx = \left[\frac{1}{4}x^4-\frac{1}{2}x^2\right]_{0}^{1} = \left(\frac{1}{4}-\frac{1}{2}\right)-(0) = -\frac{1}{4},$$

$$\int_{1}^{2}(x^3-x)\,dx = \left[\frac{1}{4}x^4-\frac{1}{2}x^2\right]_{1}^{2} = (4-2)-\left(\frac{1}{4}-\frac{1}{2}\right) = \frac{9}{4}.$$

Thus the total area is

$$\frac{9}{4}+\frac{1}{4}+\frac{1}{4}+\frac{9}{4} = 5.$$

Exercise 13

The equation $\int_a^b k\,f(x)\,dx = k\int_a^b f(x)\,dx$ has the geometric interpretation that the signed area under the graph of $y = kf(x)$ between $x = a$ and $x = b$ is k times the area under the graph of $y = f(x)$. Indeed, the graph of $y = kf(x)$ is obtained from the graph of $y = f(x)$ by a dilation of factor k from the x-axis. Algebraically, letting F be an antiderivative of f, we have $\int_a^b kf(x)\,dx = \big[kF(x)\big]_a^b = kF(b) - kF(a) = k\big[F(x)\big]_a^b = k\int_a^b f(x)\,dx$.

Exercise 14

If $a < b$, then the equation follows immediately from the fundamental theorem of calculus. If $a > b$, then $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx = -\big(F(a) - F(b)\big) = F(b) - F(a)$, as desired.

Exercise 15

As a reality check: for x between 0 and 5, the integrand is negative; and x is going backwards from 5 to 0; so the answer should be positive.

$$\int_5^0 (-2x-3)\,dx = \big[-x^2 - 3x\big]_5^0 = (0-0) - (-25-15) = 0 - (-40) = 40.$$

References

- G. Swain and T. Dence, ‘Archimedes’ quadrature of the parabola revisited’, *Mathematics Magazine* 71 (1998), 123–130.

0 1 2 3 4 5 6 7 8 9 10 11 12