Supporting Australian Mathematics Project

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

A guide for teachers – Years 11 and 12

Calculus: Module 10

Introduction to differential calculus

Education Services Australia

AMSI
AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Introduction to differential calculus – 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

Assumed knowledge 4

Motivation 4

How fast are you *really* going? 4

The more things change 6

Content 7

The gradient of secants and tangents to a graph 7

Calculating the gradient of $y = x^2$ 9

Definition of the derivative 11

Notation for the derivative 13

Some derivatives 14

Properties of the derivative 16

The product, quotient and chain rules 19

Summary of differentiation rules 26

The tangent line to a graph 26

The second derivative 27

Differentiation of inverses 28

Implicit differentiation 30

Links forward 32

Higher derivatives 32

Approximations of functions 32

History and applications 34

The discoverers of calculus 34

The Newton–Leibniz controversy 34

Mathematically rigorous calculus 35

Appendix 36

Functions differentiable and not 36

Answers to exercises 39

References 44

Introduction to differential calculus

Assumed knowledge

The content of the modules:

- *Coordinate geometry*
- *The binomial theorem*
- *Functions I*
- *Functions II*
- *Limits and continuity*.

Motivation

How fast are you *really* going?

You ride your bicycle down a straight bike path. As you proceed, you keep track of your exact position — so that, at every instant, you know exactly where you are. *How fast were you going one second after you started?*

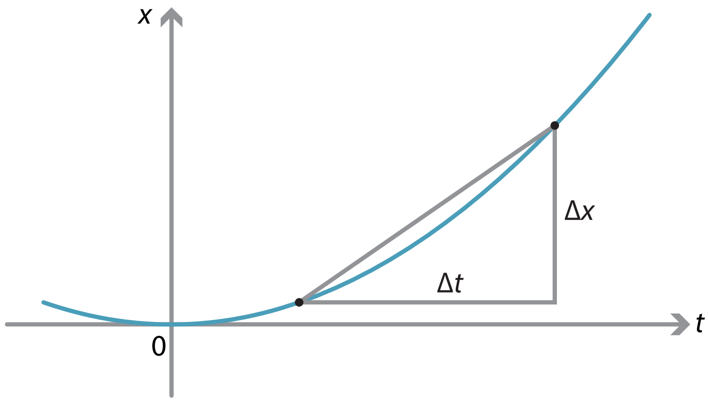
This is not a very realistic situation — it's hard to measure your position precisely at every instant! Nonetheless, let us suspend disbelief and imagine you have this information. (Perhaps you have an extremely accurate GPS or a high-speed camera.) By considering this question, we are led to some important mathematical ideas.

First, recall the formula for **average velocity**:

$$v = \frac{\Delta x}{\Delta t},$$

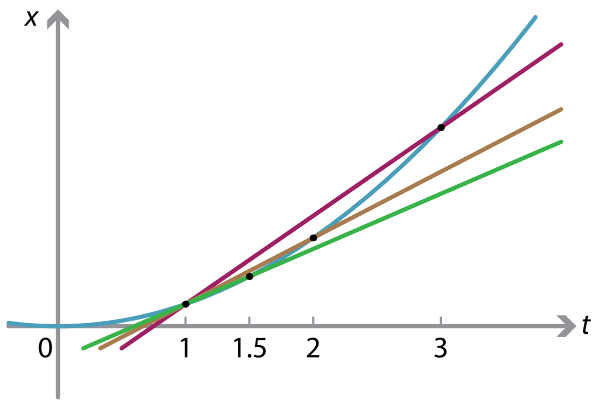
where Δx is the change in your position and Δt is the time taken. Thus, average velocity is the rate of change of position with respect to time.

Draw a graph of your position $x(t)$ at time t seconds. Connect two points on the graph, representing your position at two different times. The *gradient* of this line is your average velocity over that time period.



Average velocity $v = \frac{\Delta x}{\Delta t}$.

Trying to discover your velocity at the one-second mark ($t = 1$), you calculate your *average velocity* over the period from $t = 1$ to a slightly later time $t = 1 + \Delta t$. Trying to be more accurate, you look at shorter time intervals, with Δt smaller and smaller. If you really knew your position at every single instant of time, then you could work out your average velocity over any time interval, no matter how short. The three lines in the following diagram correspond to $\Delta t = 2$, $\Delta t = 1$ and $\Delta t = 0.5$.



Average velocity over shorter time intervals.

As Δt approaches 0, you obtain better and better estimates of your instantaneous velocity at the instant $t = 1$. These estimates correspond to the gradients of lines connecting closer and closer points on the graph.

In the limit, as $\Delta t \to 0$,

$$\lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}$$

gives the precise value for the instantaneous velocity at $t = 1$. This is also the gradient of the *tangent* to the graph at $t = 1$. **Instantaneous velocity** is the instantaneous rate of change of position with respect to time.

Although the scenario is unrealistic, these ideas show how you could answer the question of how fast you were going at $t = 1$. Given a function $x(t)$ describing your position at time t, you could calculate your exact velocity at time $t = 1$.

In fact, it is possible to calculate the instantaneous velocity at *any* value of t, obtaining a function which gives your instantaneous velocity at time t. This function is commonly denoted by $x'(t)$ or $\frac{dx}{dt}$, and is known as the *derivative* of $x(t)$ with respect to t.

In this module, we will discuss derivatives.

The more things change . . .

Velocity is an important example of a derivative, but this is just one example. The world is full of quantities which change with respect to each other — and these rates of change can often be expressed as derivatives. It is often important to understand and predict how things will change, and so derivatives are important.

Here are some examples of derivatives, illustrating the range of topics where derivatives are found:

- **Mechanics.** We saw that the derivative of position with respect to time is velocity. Also, the derivative of velocity with respect to time is *acceleration*. And the derivative of momentum with respect to time is the (net) *force* acting on an object.
- **Civil engineering, topography.** Let $h(x)$ be the height of a road, or the altitude of a mountain, as you move along a horizontal distance x. The derivative $h'(x)$ with respect to distance is the *gradient* of the road or mountain.
- **Population growth.** Suppose a population has size $p(t)$ at time t. The derivative $p'(t)$ with respect to time is the *population growth rate*. The growth rates of human, animal and cell populations are important in demography, ecology and biology, respectively.
- **Economics.** In macroeconomics, the rate of change of the gross domestic product (GDP) of an economy with respect to time is known as the *economic growth rate*. It is often used by economists and politicians as a measure of progress.

- **Mechanical engineering.** Suppose that the total amount of energy produced by an engine is $E(t)$ at time t. The derivative $E'(t)$ of energy with respect to time is the *power* of the engine.

All of these examples arise from a more abstract question in mathematics:

- **Mathematics.** Consider the graph of a function $y = f(x)$, which is a curve in the plane. What is the *gradient* of a tangent to this graph at a point? Equivalently, what is the instantaneous rate of change of y with respect to x?

In this module, we discuss purely mathematical questions about derivatives. In the three modules *Applications of differentiation*, *Growth and decay* and *Motion in a straight line*, we discuss some real-world examples.

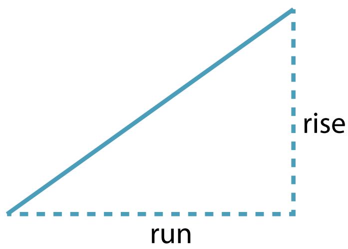
Therefore, although the *Motivation* section has focused on instantaneous velocity, which is an important motivating example, we now concentrate on calculating the gradient of a tangent to a curve.

Content

The gradient of secants and tangents to a graph

Consider a function $f : \mathbb{R} \to \mathbb{R}$ and its graph $y = f(x)$, which is a curve in the plane. We wish to find the *gradient* of this curve at a point. But first we need to define properly what we mean by the gradient of a curve at a point!

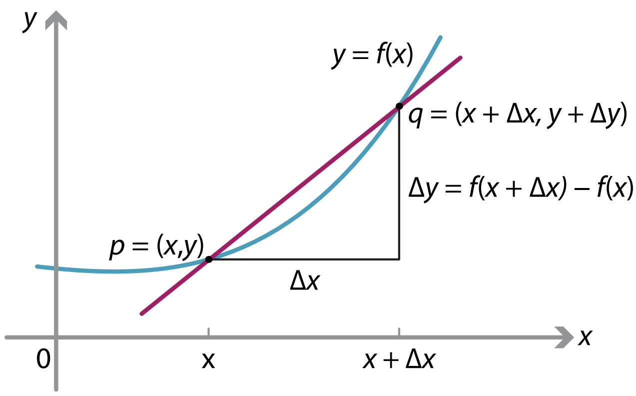
The module *Coordinate geometry* defines the gradient of a line in the plane: Given a non-vertical line and two points on it, the **gradient** is defined as $\dfrac{\text{rise}}{\text{run}}$.



Gradient of a line.

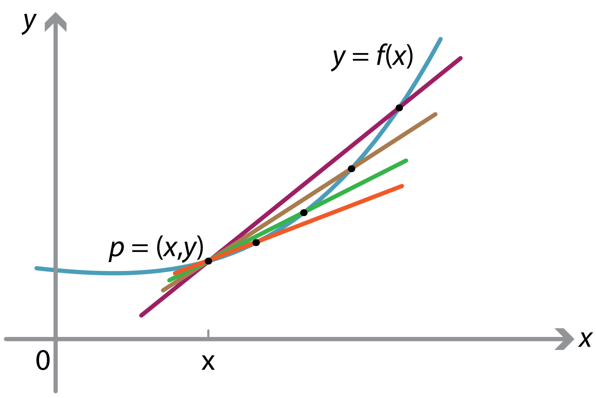
Now, given a *curve* defined by $y = f(x)$, and a point p on the curve, consider another point q on the curve near p, and draw the line pq connecting p and q. This line is called a **secant line**.

We write the coordinates of p as (x, y), and the coordinates of q as $(x + \Delta x, y + \Delta y)$. Here Δx represents a small change in x, and Δy represents the corresponding small change in y.



Secant connecting points on the graph $y = f(x)$ at x and $x + \Delta x$.

As Δx becomes smaller and smaller, the point q approaches p, and the secant line pq approaches a line called the **tangent** to the curve at p. We define the **gradient of the curve** at p to be the gradient of this tangent line.



Secants on $y = f(x)$ approaching the tangent line at x.

Note that, in this definition, the approximation of a tangent line by secant lines is just like the approximation of instantaneous velocity by average velocities, as discussed in the *Motivation* section.

With this definition, we now consider how to compute the gradient of the curve $y = f(x)$ at the point $p = (x, y)$.

Taking $q = (x + \Delta x, y + \Delta y)$ as above, the secant line pq has gradient

$$\frac{\text{rise}}{\text{run}} = \frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Note that the symbol Δ on its own has no meaning: Δx and Δy refer to change in x and y, respectively. You cannot cancel the Δ's!

As $\Delta x \to 0$, the gradient of the tangent line is given by

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

We also denote this limit by $\frac{dy}{dx}$:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx}.$$

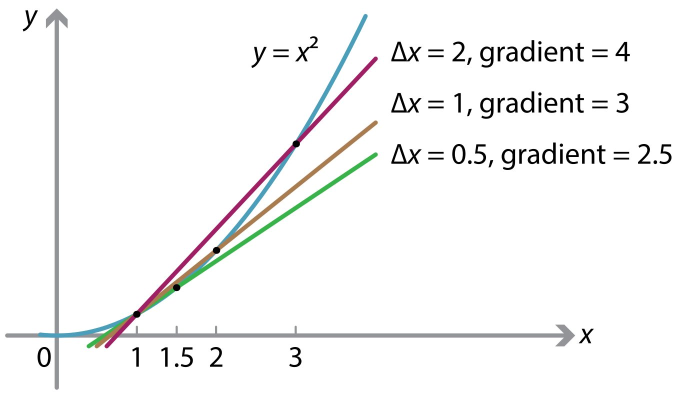
The notation $\frac{dy}{dx}$ indicates the instantaneous rate of change of y with respect to x, and is not a fraction. For our purposes, the expressions dx and dy have no meaning on their own, and the d's do not cancel!

The gradient of a secant is analogous to average velocity, and the gradient of a tangent is analogous to instantaneous velocity. Velocity is the instantaneous rate of change of position with respect to time, and the gradient of a tangent to the graph $y = f(x)$ is the instantaneous rate of change of y with respect to x.

Calculating the gradient of $y = x^2$

Let us consider a specific function $f(x) = x^2$ and its graph $y = f(x)$, which is the standard parabola. To illustrate the ideas in the previous section, we will calculate the gradient of this curve at $x = 1$.

We first construct secant lines between the points on the graph at $x = 1$ and $x = 1 + \Delta x$, and calculate their gradients.



Gradients of secants from $x = 1$ to $x = 1 + \Delta x$.

For instance, taking $\Delta x = 2$, we consider the secant connecting the points at $x = 1$ and $x = 3$. Between these two points, $f(x)$ increases from $f(1) = 1$ to $f(3) = 9$, giving $\Delta y = 8$, and hence

$$\frac{\Delta y}{\Delta x} = \frac{8}{2} = 4.$$

We compute gradients of secants for various values of Δx in the following table.

Secants of the parabola $f(x) = x^2$

Secant between points	Δx	$\Delta y = f(x + \Delta x) - f(x)$	Gradient of secant $\frac{\Delta y}{\Delta x}$
$x = 1, \quad x = 3$	2	8	4
$x = 1, \quad x = 2$	1	3	3
$x = 1, \quad x = 1.5$	0.5	1.25	2.5
$x = 1, \quad x = 1.1$	0.1	0.21	2.1
$x = 1, \quad x = 1.001$	0.001	0.002001	2.001

As Δx approaches 0, the gradients of the secants approach 2. It turns out that indeed the gradient of the tangent at $x = 1$ is 2. To see why, consider the interval of length Δx, from $x = 1$ to $x = 1 + \Delta x$. We have

$$\begin{aligned} \Delta y &= f(1 + \Delta x) - f(1) \\ &= (1 + \Delta x)^2 - 1^2 \\ &= 2\,\Delta x + (\Delta x)^2, \end{aligned}$$

so that

$$\begin{aligned} \frac{\Delta y}{\Delta x} &= \frac{2\,\Delta x + (\Delta x)^2}{\Delta x} \\ &= 2 + \Delta x. \end{aligned}$$

In the limit, as $\Delta x \to 0$, we obtain the instantaneous rate of change

$$\begin{aligned} \frac{dy}{dx} &= \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \\ &= \lim_{\Delta x \to 0} \big(2 + \Delta x\big) = 2. \end{aligned}$$

(So, if you were riding your bike and your position was $f(x) = x^2$ metres after x seconds, then your instantaneous velocity after 1 second would be 2 metres per second.)

There's nothing special about the point $x = 1$ or the function $f(x) = x^2$, as the following example illustrates.

Example

Let $f(x) = x^3$. What is the gradient of the tangent line to the graph $y = f(x)$ at the point $(2, 8)$?

Solution

The gradient at $x = 2$ is given by the limit

$$\begin{aligned}\lim_{\Delta x \to 0} \frac{f(2+\Delta x) - f(2)}{\Delta x} &= \lim_{\Delta x \to 0} \frac{(2+\Delta x)^3 - 2^3}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{8 + 12(\Delta x) + 6(\Delta x)^2 + (\Delta x)^3 - 8}{\Delta x} \\ &= \lim_{\Delta x \to 0} \left(12 + 6(\Delta x) + (\Delta x)^2\right) = 12.\end{aligned}$$

Thus the gradient of the tangent line to $y = f(x)$ at $(2, 8)$ is 12.

Exercise 1

Using a similar method, find the gradient of the tangent line to $y = x^4$ at $(-1, 1)$, and find the equation of this line.

In the examples so far, we have been given a curve, and we have found the gradient of the curve at one particular point on the curve. But we can also find the gradient at all points simultaneously, as the next section illustrates.

Definition of the derivative

The method we used in the previous section to find the gradient of a tangent to a graph at a point can actually be used to work out the gradient everywhere, simultaneously.

Example

Let $f(x) = x^2$. What is the gradient of the tangent line to the graph $y = f(x)$ at a general point $(x, f(x))$ on this graph?

Solution

We calculate the same limit as in previous examples, using the variable x in place of a number:

$$\begin{aligned}\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x} &= \lim_{\Delta x\to 0}\frac{(x+\Delta x)^2-x^2}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{x^2+2x(\Delta x)+(\Delta x)^2-x^2}{\Delta x}\\ &= \lim_{\Delta x\to 0}\bigl(2x+\Delta x\bigr)=2x.\end{aligned}$$

Thus, the gradient of the tangent line at the point $(x, f(x))$ is $2x$.

The **derivative** of a function $f(x)$ is the function $f'(x)$ which gives the gradient of the tangent to the graph $y = f(x)$ at each value of x. It is often also denoted $\frac{dy}{dx}$. Thus

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

The previous example shows that the derivative of $f(x) = x^2$ is $f'(x) = 2x$.

Exercise 2

Show that the derivative of $f(x) = x^3$ is $f'(x) = 3x^2$.

Functions which have derivatives are called **differentiable**. Not all functions are differentiable; in particular, to be differentiable, a function must be continuous. Almost all functions we meet in secondary school mathematics are differentiable. In particular, all polynomials, rational functions, exponentials, logarithms and trigonometric functions (such as sin, cos and tan) are differentiable.

Derivatives of exponential, logarithmic and trigonometric functions are discussed in the two modules *Exponential and logarithmic functions* and *The calculus of trigonometric functions*.

We shall say a little more about which functions are differentiable and which are not in the *Appendix* to this module.

In the following example and exercises, we differentiate constant and linear functions.

Exercise 3

Show that the derivative of a constant function $f(x) = c$, with c a real constant, is given by $f'(x) = 0$.

Example

What is the derivative of the linear function $f(x) = 3x + 7$?

Solution

We calculate

$$\begin{aligned} f'(x) &= \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{3(x + \Delta x) + 7 - 3x - 7}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{3\Delta x}{\Delta x} = 3. \end{aligned}$$

Thus $f'(x) = 3$. (This answer makes sense, as the graph of $y = f(x)$ is the line $y = 3x + 7$, which has gradient 3.)

Exercise 4

Show that the derivative of a linear function $f(x) = ax + b$, with a, b real constants, is $f'(x) = a$.

Notation for the derivative

We have introduced two different notations for the derivative. Both are standard, and it is necessary to be proficient with both.

The first notation is to write $f'(x)$ for the derivative of the function $f(x)$. This functional notation was introduced by Lagrange, based on Isaac Newton's ideas. The dash in $f'(x)$ denotes that $f'(x)$ is *derived* from $f(x)$.

The other notation is to write $\frac{dy}{dx}$. This notation refers to the instantaneous rate of change of y with respect to x, and was introduced by Gottfried Wilhelm Leibniz, one of the discoverers of calculus. (The other discoverer was Isaac Newton. In fact, the question of who discovered calculus first was historically a point of great controversy. For more details, see the *History and applications* section.)

If we have a function $f(x)$ and its graph $y = f(x)$, then the derivative $f'(x)$ is the gradient of the tangent of $y = f(x)$, which is also the instantaneous rate of change $\frac{dy}{dx}$. Thus the notations are equivalent:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = f'(x).$$

For example, we calculated earlier that the derivative of x^2 was $2x$. Thus $y = x^2$ implies $\frac{dy}{dx} = 2x$. Alternatively, $f(x) = x^2$ implies $f'(x) = 2x$. The two notations express the same result.

Another common usage of Leibniz notation is to consider $\frac{d}{dx}$ as an *operator*, meaning 'differentiate with respect to x'. So

$$\frac{dy}{dx} = \frac{d}{dx}(y)$$

means: take y, and differentiate with respect to x. An alternative way to denote differentiation of x^2 using Leibniz notation would be

$$\frac{d}{dx}(x^2) = 2x.$$

At secondary school level, a 'd' or 'dx' or 'dy' has no meaning in itself; it only makes sense as part of a $\frac{d}{dx}$ or $\frac{dy}{dx}$ or similar.[1] The d's do not cancel.

The two types of notation each have their advantages and disadvantages. It is also common to mix notation, and we will do so in this module. For example, we can write

$$\frac{d}{dx} f(x) = f'(x).$$

Some derivatives

So far in the examples and exercises we have found the following derivatives.

	$f(x)$	$f'(x)$
constant	c	0
linear	$ax+b$	a
	x^2	$2x$
	x^3	$3x^2$

[1] However, in more advanced pure mathematics, d is regarded as a more complicated object called a *differential operator*, and objects like dx and dy are studied as examples of *differential forms*.

Note that the example $f(x) = ax + b$ includes the case $f(x) = x$, which has derivative $f'(x) = 1$. Since we have seen that

$$\frac{d}{dx}(x) = 1, \qquad \frac{d}{dx}(x^2) = 2x \qquad \text{and} \qquad \frac{d}{dx}(x^3) = 3x^2,$$

it is natural to conjecture that the derivative of x^n is nx^{n-1}.

We will now compute the derivative of $f(x) = x^n$, for any positive integer n. To do so, we need the binomial expansion

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-2}a^2b^{n-2} + \binom{n}{n-1}ab^{n-1} + b^n.$$

See the module *The binomial theorem* for details.

We begin by using the binomial theorem to expand

$$\begin{aligned} f(x+\Delta x) &= (x+\Delta x)^n \\ &= x^n + \binom{n}{1}x^{n-1}(\Delta x) + \binom{n}{2}x^{n-2}(\Delta x)^2 + \cdots + \binom{n}{n-1}x(\Delta x)^{n-1} + (\Delta x)^n. \end{aligned}$$

We can then compute

$$\begin{aligned} \frac{f(x+\Delta x) - f(x)}{\Delta x} &= \frac{(x+\Delta x)^n - x^n}{\Delta x} \\ &= \binom{n}{1}x^{n-1} + \binom{n}{2}x^{n-2}(\Delta x) + \cdots + \binom{n}{n-1}x(\Delta x)^{n-2} + (\Delta x)^{n-1}. \end{aligned}$$

The last line has a Δx in every term except the first. Since $\binom{n}{1} = n$, we have

$$\begin{aligned} f'(x) &= \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(\Delta x)}{\Delta x} \\ &= nx^{n-1}. \end{aligned}$$

The fact that the derivative of x^n is nx^{n-1} holds more generally than just when n is a positive integer. The next exercise shows it is also true when $n = -1$. Exercise 14, later in this module, shows how to find the derivative of negative integer powers of x in general.

Exercise 5

Prove that the derivative of $f(x) = \dfrac{1}{x}$ is $f'(x) = -\dfrac{1}{x^2}$. That is, prove that

$$\frac{d}{dx}(x^{-1}) = -x^{-2}.$$

In fact, it's also true that for any non-zero *rational number* (i.e., fraction) n, the derivative of $f(x) = x^n$ is $f'(x) = nx^{n-1}$. See exercises 20 and 21 later in this module.

Even more generally, for any *real number* a, including irrational a, the derivative of

$$f(x) = x^a \qquad \text{is} \qquad f'(x) = ax^{a-1}.$$

It is not obvious how to even *define* what it means to raise a number to the power of an irrational number. For instance, 2^3 just means $2 \times 2 \times 2$, and $2^{\frac{7}{5}}$ just means $\sqrt[5]{2^7}$, but what does $2^{\sqrt{3}}$ mean? In the module *Exponential and logarithmic functions*, we explore these issues, show how to define x^a precisely for any real number a, and show that the derivative of x^a is ax^{a-1}.

In summary, the following theorem is true.

Theorem

For any real number a, the derivative of $f(x) = x^a$ is $f'(x) = ax^{a-1}$, wherever $f(x)$ is defined.

Properties of the derivative

We now consider various properties of differentiation. As we proceed, we will be able to differentiate wider and wider classes of functions.

Throughout this section and the next, we will be manipulating limits as we compute derivatives. We therefore recall some basic rules for limits; see the module *Limits and continuity* for details. The following hold provided the limits $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist.

- The limit of a sum (or difference) is the sum (or difference) of the limits:

$$\lim_{x\to a}\big(f(x) \pm g(x)\big) = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x).$$

- The limit of a product is the product of the limits:

$$\lim_{x\to a} f(x)\,g(x) = \big(\lim_{x\to a} f(x)\big)\big(\lim_{x\to a} g(x)\big).$$

  This includes the case of multiplication by a constant c:

$$\lim_{x\to a} c\,f(x) = c\,\lim_{x\to a} f(x).$$

The derivative of a constant multiple

Suppose we want to differentiate $4x^7$. Rather than returning to the definition of a derivative, we can use the following theorem.

Theorem

Let f be a differentiable function and let c be a constant. Then the derivative of $cf(x)$ is $cf'(x)$. That is,

$$\frac{d}{dx}[cf(x)] = cf'(x).$$

This fact can also be written in Leibniz notation as

$$\frac{d}{dx}(cf) = c\frac{df}{dx}.$$

Proof

The derivative of $cf(x)$ is given by

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

We may factor out the c, since it is just a constant. □

Example

What is the derivative of $4x^7$?

Solution

The theorem tells us that the derivative of $4x^7$ is 4 times the derivative of x^7. Hence,

$$\begin{aligned}\frac{d}{dx}(4x^7) &= 4\frac{d}{dx}(x^7)\\ &= 4 \cdot 7x^6\\ &= 28x^6.\end{aligned}$$

The derivative of a sum

A function like $f(x) = x^3 + 5x^2$ can be differentiated from first principles; alternatively, we can use the following theorem.

Theorem

Suppose both f and g are differentiable functions. Then the derivative of $f(x)+g(x)$ is $f'(x)+g'(x)$ and, similarly, the derivative of $f(x)-g(x)$ is $f'(x)-g'(x)$. That is,

$$\frac{d}{dx}\big[f(x)+g(x)\big] = f'(x)+g'(x), \qquad \frac{d}{dx}\big[f(x)-g(x)\big] = f'(x)-g'(x).$$

In Leibniz notation, these statements can be written as

$$\frac{d}{dx}(f+g) = \frac{df}{dx}+\frac{dg}{dx}, \qquad \frac{d}{dx}(f-g) = \frac{df}{dx}-\frac{dg}{dx}.$$

Proof

We prove the first statement. The derivative of $f(x)+g(x)$ is given by

$$\begin{aligned}
&\lim_{\Delta x \to 0} \frac{\big(f(x+\Delta x)+g(x+\Delta x)\big)-\big(f(x)+g(x)\big)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x} + \lim_{\Delta x \to 0} \frac{g(x+\Delta x)-g(x)}{\Delta x} \\
&= f'(x)+g'(x).
\end{aligned}$$

Note that, as both the limits in the second line exist, we are retrospectively justified in splitting the first limit into two pieces. □

Exercise 6

Adapt the above proof to prove the second statement of the theorem: the derivative of the difference of two functions is the difference of the derivatives.

Example

Find the derivative of $f(x) = x^3 + 5x^2$.

Solution

We have, using the above theorems,

$$\begin{aligned}
f'(x) &= \frac{d}{dx}\big(x^3+5x^2\big) \\
&= \frac{d}{dx}\big(x^3\big) + \frac{d}{dx}\big(5x^2\big) && \text{(derivative of a sum)} \\
&= \frac{d}{dx}\big(x^3\big) + 5\frac{d}{dx}\big(x^2\big) && \text{(derivative of a constant multiple)} \\
&= 3x^2+10x && \text{(derivative of } x^n\text{)}.
\end{aligned}$$

Note. The solution to the previous example shows every individual step explicitly and states which theorems are used. Once proficient with these properties of the derivative, however, there is no need to justify each step in this way. It is common, for instance, to go straight from

$$f(x) = x^3 + 5x^2 \qquad \text{to} \qquad f'(x) = 3x^2 + 10x.$$

Exercise 7

Let $f(x) = \dfrac{x^5 + \sqrt{x}}{x^2}$. Find $f'(x)$.

Linearity of the derivative

The two previous theorems (for the derivative of a sum and the derivative of a constant multiple) can be summarised as follows. If f, g are differentiable functions and a, b are constants, then

$$\frac{d}{dx}\big[a\,f(x) + b\,g(x)\big] = a\,f'(x) + b\,g'(x).$$

The same fact can be written in Leibniz notation as

$$\frac{d}{dx}\big(af + bg\big) = a\frac{df}{dx} + b\frac{dg}{dx}.$$

This property is sometimes expressed by saying that 'differentiation is linear'.

The product, quotient and chain rules

We now move to some more involved properties of differentiation. To summarise, so far we have found that:

- the derivative of a constant multiple is the constant multiple of the derivative
- the derivative of a sum is the sum of the derivatives
- the derivative of a difference is the difference of the derivatives.

However, it turns out that:

- the derivative of a product $f(x)\,g(x)$ is *not* the product of the derivatives
- the derivative of a quotient $\dfrac{f(x)}{g(x)}$ is *not* the quotient of the derivatives
- the derivative of the composition $f(g(x))$ is *not* the composition of the derivatives.

The product, quotient and chain rules tell us how to differentiate in these three situations. We consider the three rules in turn.

The product rule

Theorem (Product rule)

Let f, g be differentiable functions. Then the derivative of their product is given by

$$\frac{d}{dx}\big[f(x)\,g(x)\big] = f(x)\,g'(x) + g(x)\,f'(x).$$

The product rule is also often written as

$$\frac{d}{dx}(fg) = f\,\frac{dg}{dx} + g\,\frac{df}{dx}.$$

Proof

As before, we evaluate the limit which gives the derivative:

$$\frac{d}{dx}\big[f(x)\,g(x)\big] = \lim_{\Delta x\to 0} \frac{f(x+\Delta x)\,g(x+\Delta x) - f(x)\,g(x)}{\Delta x}.$$

The trick is to add and subtract an extra term in the numerator, so that we can factorise and obtain some familiar-looking expressions:

$$\begin{aligned}
&f(x+\Delta x)\,g(x+\Delta x) - f(x)\,g(x)\\
&= f(x+\Delta x)\,g(x+\Delta x) - f(x)\,g(x+\Delta x) + f(x)\,g(x+\Delta x) - f(x)\,g(x)\\
&= \big[f(x+\Delta x) - f(x)\big]g(x+\Delta x) + f(x)\big[g(x+\Delta x) - g(x)\big].
\end{aligned}$$

We can then rewrite the limit as

$$\lim_{\Delta x\to 0}\Big[\frac{f(x+\Delta x) - f(x)}{\Delta x}\,g(x+\Delta x) + f(x)\,\frac{g(x+\Delta x) - g(x)}{\Delta x}\Big].$$

Now, since the limit of a sum is the sum of the limits, and the limit of a product is the product of the limits, we obtain

$$\Big(\lim_{\Delta x\to 0}\frac{f(x+\Delta x) - f(x)}{\Delta x}\Big)\Big(\lim_{\Delta x\to 0} g(x+\Delta x)\Big) + f(x)\Big(\lim_{\Delta x\to 0}\frac{g(x+\Delta x) - g(x)}{\Delta x}\Big).$$

We also used the fact that $f(x)$ does not depend on Δx. Recognising $f'(x)$ and $g'(x)$, and substituting $\Delta x = 0$ into $g(x+\Delta x)$, we obtain

$$\frac{d}{dx}\big[f(x)\,g(x)\big] = f'(x)\,g(x) + f(x)\,g'(x),$$

which is equivalent to the desired formula. □

Exercise 8

Starting from the fact that the derivative of x is 1, use the product rule to prove by induction on n, that for all positive integers n,

$$\frac{d}{dx}\big(x^n\big) = nx^{n-1}.$$

Example

Let $f(x) = (x^3+2)(x^2+1)$. Find $f'(x)$.

Solution

We could expand out $f(x)$ and differentiate term-by-term. Alternatively, with the product rule, we obtain

$$\begin{aligned} f'(x) &= (x^3+2)\,\frac{d}{dx}\big(x^2+1\big) + (x^2+1)\,\frac{d}{dx}\big(x^3+2\big) \\ &= (x^3+2)\cdot 2x + (x^2+1)\cdot 3x^2 \\ &= 5x^4 + 3x^2 + 4x. \end{aligned}$$

Exercise 9

Using the product rule, prove that in general, for a differentiable function $f\colon \mathbb{R} \to \mathbb{R}$, the derivative of $(f(x))^2$ with respect to x is $2\,f(x)\,f'(x)$.

Exercise 10

By using the product rule, prove the following 'extended product rule':

$$\frac{d}{dx}\big[f(x)\,g(x)\,h(x)\big] = f'(x)\,g(x)\,h(x) + f(x)\,g'(x)\,h(x) + f(x)\,g(x)\,h'(x).$$

Generalise to the product of any number of functions.

The chain rule

The chain rule allows us to differentiate the *composition* of two functions. Recall from the module *Functions II* that the **composition** of two functions g and f is

$$(f \circ g)(x) = f(g(x)).$$

We start with x, apply g, then apply f. The chain rule tells us how to differentiate such a function.

Theorem (Chain rule)

Let f, g be differentiable functions. Then the derivative of their composition is

$$\frac{d}{dx}\left[f(g(x))\right] = f'(g(x))\, g'(x).$$

In Leibniz notation, we may write $u = g(x)$ and $y = f(u) = f(g(x))$; diagrammatically,

$$x \xrightarrow{g} u \xrightarrow{f} y.$$

Then the chain rule says that 'differentials cancel' in the sense that

$$\frac{dy}{dx} = \frac{dy}{du}\,\frac{du}{dx}.$$

Proof

To calculate the derivative, we must evaluate the limit

$$\frac{d}{dx}\left[f(g(x))\right] = \lim_{\Delta x \to 0} \frac{f(g(x+\Delta x)) - f(g(x))}{\Delta x}.$$

The trick is to multiply and divide by an extra term in the expression above, as shown, so that we obtain two expressions which both express rates of change:

$$\frac{f(g(x+\Delta x)) - f(g(x))}{\Delta x} = \frac{f(g(x+\Delta x)) - f(g(x))}{g(x+\Delta x) - g(x)}\,\frac{g(x+\Delta x) - g(x)}{\Delta x}.$$

We can then rewrite the desired limit as

$$\left(\lim_{\Delta x \to 0} \frac{f(g(x+\Delta x)) - f(g(x))}{g(x+\Delta x) - g(x)}\right)\left(\lim_{\Delta x \to 0} \frac{g(x+\Delta x) - g(x)}{\Delta x}\right).$$

The ratio in the first limit expresses the change in the function f, from its value at $g(x)$ to its value at $g(x+\Delta x)$, relative to the difference between $g(x+\Delta x)$ and $g(x)$. So as $\Delta x \to 0$, this first term approaches the derivative of f at the point $g(x)$, namely $f'(g(x))$. The second limit is clearly $g'(x)$. We conclude that

$$\frac{d}{dx}\left[f(g(x))\right] = f'(g(x))\, g'(x),$$

as required. □

The proof above is not entirely rigorous: for instance, if there are values of Δx close to zero such that $g(x+\Delta x) - g(x) = 0$, then we have division by zero in the first limit. However, a fully rigorous proof is beyond the secondary school level.

The next two examples illustrate 'functional' and 'Leibniz' methods of attacking the same problem using the chain rule.

Example

Let $f(x) = (x^7 - x^2)^{42}$. Find $f'(x)$.

Solution

The function $f(x)$ is the composition of the functions $g(x) = x^7 - x^2$ and $h(x) = x^{42}$, that is, $f(x) = h(g(x))$. We compute

$$g'(x) = 7x^6 - 2x, \qquad h'(x) = 42x^{41},$$

and the chain rule gives

$$\begin{aligned} f'(x) &= h'(g(x))\, g'(x) \\ &= 42(x^7 - x^2)^{41}(7x^6 - 2x). \end{aligned}$$

Example

Let $y = (x^7 - x^2)^{42}$. Find $\dfrac{dy}{dx}$.

Solution

Let $u = x^7 - x^2$, so that $y = u^{42}$. We then have

$$\frac{dy}{du} = 42u^{41}, \qquad \frac{du}{dx} = 7x^6 - 2x,$$

and the chain rule gives

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = 42u^{41}(7x^6 - 2x).$$

Rewriting u in terms of x gives

$$\frac{dy}{dx} = 42(x^7 - x^2)^{41}(7x^6 - 2x).$$

Exercise 11

Find the derivative of $(x^2 + 7)^{100}$ with respect to x.

Exercise 12

In exercise 9, we proved that the derivative of $(f(x))^2$ with respect to x is $2f(x)\,f'(x)$. Re-prove this fact using the chain rule.

Exercise 13

Prove the following 'extended chain rule':

$$\frac{d}{dx}\big[f(g(h(x)))\big] = f'(g(h(x)))\,g'(h(x))\,h'(x).$$

Generalise to the composition of any number of functions.

The following exercise shows how, if you know the derivative of x^n for a positive number n, you can find the derivative of x^{-n}.

Exercise 14

Let $g(x) = x^n$, where n is positive. Using the facts $g'(x) = nx^{n-1}$ and $\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}$ and the chain rule, calculate $\frac{d}{dx}\left(\frac{1}{x^n}\right)$.

The quotient rule

Theorem (Quotient rule)

Let f, g be differentiable functions. Then the derivative of their quotient is

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{g(x)\,f'(x) - f(x)\,g'(x)}{(g(x))^2}.$$

Alternatively, we can write

$$\frac{d}{dx}\left(\frac{f}{g}\right) = \frac{g\,\frac{df}{dx} - f\,\frac{dg}{dx}}{g^2}.$$

Example

Let $f(x) = \dfrac{x^2+1}{x^2-1}$. What is $f'(x)$?

Solution

Using the quotient rule, we have

$$\begin{aligned} f'(x) &= \frac{(x^2-1)\,\frac{d}{dx}(x^2+1) - (x^2+1)\,\frac{d}{dx}(x^2-1)}{(x^2-1)^2} \\ &= \frac{(x^2-1)\cdot 2x - (x^2+1)\cdot 2x}{(x^2-1)^2} = \frac{-4x}{(x^2-1)^2}. \end{aligned}$$

Example

Let $f(x) = \dfrac{x}{\sqrt{x^2+1}}$. Find $f'(x)$.

Solution

We first apply the quotient rule:

$$f'(x) = \frac{\sqrt{x^2+1}\,\frac{d}{dx}(x) - x\,\frac{d}{dx}\sqrt{x^2+1}}{x^2+1}.$$

To differentiate $\sqrt{x^2+1}$, we use the chain rule:

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

Now returning to $f'(x)$, we obtain

$$\begin{aligned}f'(x) &= \frac{\sqrt{x^2+1}\cdot 1 - x\cdot x(x^2+1)^{-\frac{1}{2}}}{x^2+1}\\ &= \frac{(x^2+1)-x^2}{(x^2+1)^{\frac{3}{2}}}\\ &= (x^2+1)^{-\frac{3}{2}}.\end{aligned}$$

The quotient rule can be proved using the product and chain rules, as the next two exercises show.

Exercise 15

Let g be a differentiable function. Using the chain rule, show that

$$\frac{d}{dx}\left(\frac{1}{g(x)}\right) = -\frac{g'(x)}{(g(x))^2}.$$

(This is a generalisation of exercise 14.)

Exercise 16

Using the previous exercise and the product rule, find the derivative of $f(x)\cdot\dfrac{1}{g(x)}$, and hence prove the quotient rule.

Summary of differentiation rules

We can summarise the differentiation rules we have found as follows. They can be expressed in both functional and Leibniz notation. First, the linearity of differentiation.

Linearity of differentiation

	Functional notation	Leibniz notation
Constant multiple	$\dfrac{d}{dx}\big[c f(x)\big] = c f'(x)$	$\dfrac{d}{dx}(cf) = c\dfrac{df}{dx}$
Sum	$\dfrac{d}{dx}\big[f(x)+g(x)\big] = f'(x)+g'(x)$	$\dfrac{d}{dx}(f+g) = \dfrac{df}{dx}+\dfrac{dg}{dx}$
Difference	$\dfrac{d}{dx}\big[f(x)-g(x)\big] = f'(x)-g'(x)$	$\dfrac{d}{dx}(f-g) = \dfrac{df}{dx}-\dfrac{dg}{dx}$

We also have the product, quotient and chain rules. In Leibniz notation, these rules are often written with u, v rather than f, g.

Product, quotient and chain rules

	Functional notation	Leibniz notation
Product	$\dfrac{d}{dx}\big[f(x)\,g(x)\big] = f(x)\,g'(x)+g(x)\,f'(x)$	$\dfrac{d}{dx}(uv) = u\dfrac{dv}{dx}+v\dfrac{du}{dx}$
Quotient	$\dfrac{d}{dx}\left[\dfrac{f(x)}{g(x)}\right] = \dfrac{g(x)\,f'(x)-f(x)\,g'(x)}{(g(x))^2}$	$\dfrac{d}{dx}\left(\dfrac{u}{v}\right) = \dfrac{v\,\frac{du}{dx}-u\,\frac{dv}{dx}}{v^2}$
Chain	$\dfrac{d}{dx}\big[f(g(x))\big] = f'(g(x))\,g'(x)$	$\dfrac{dy}{dx} = \dfrac{dy}{du}\,\dfrac{du}{dx}$

The tangent line to a graph

Given a graph $y = f(x)$, we have seen how to calculate the gradient of a tangent line to this graph. We can go further and find the *equation* of a tangent line.

Consider the tangent line to the graph $y = f(x)$ at $x = a$. This line has gradient $f'(a)$ and passes through the point $(a, f(a))$. Once we know a point on the line and its gradient, we can write down its equation:

$$y - f(a) = f'(a)\,(x-a).$$

(See the module *Coordinate geometry*.)

Example

Find the equation of the tangent line to the graph $y = \frac{1}{2}x^2$ at $x = 3$.

Solution

Letting $f(x) = \frac{1}{2}x^2$, we have $f'(x) = x$, so $f(3) = \frac{9}{2}$ and $f'(3) = 3$. Thus the tangent line has gradient 3 and passes through $(3, \frac{9}{2})$, and is given by

$$y - \frac{9}{2} = 3(x-3)$$

or, equivalently,

$$y = 3x - \frac{9}{2}.$$

Exercise 17

What is the equation of the tangent line to the graph of $y = \sqrt{9-x^2}$ at $x = \frac{3\sqrt{2}}{2}$?

The second derivative

Given a function $f(x)$, we can differentiate it to obtain $f'(x)$. It can be useful for many purposes to differentiate again and consider the **second derivative** of a function.

In functional notation, the second derivative is denoted by $f''(x)$. In Leibniz notation, letting $y = f(x)$, the second derivative is denoted by $\frac{d^2y}{dx^2}$.

The placement of the 2's in the notation $\frac{d^2y}{dx^2}$ may appear unusual. We consider that we have applied the differentiation operator $\frac{d}{dx}$ twice to y:

$$\left(\frac{d}{dx}\right)^2 y = \frac{d^2}{dx^2}\, y = \frac{d^2y}{dx^2};$$

or that we have applied the differentiation operator $\frac{d}{dx}$ to $\frac{dy}{dx}$:

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right).$$

As we will see in the module *Applications of differentiation*, the second derivative can be very useful in curve-sketching. The second derivative determines the *convexity* of the graph $y = f(x)$ and can, for example, be used to distinguish maxima from minima.

The second derivative can also have a physical meaning. For example, if $x(t)$ gives position at time t, then $x'(t)$ is the velocity and the second derivative $x''(t)$ is the *acceleration* at time t. This is discussed in the module *Motion in a straight line*.

Example

Find the second derivative of $f(x) = x^2$.

Solution

We have $f'(x) = 2x$, and so $f''(x) = 2$.

This example implies that, if you ride your bike and your position is x^2 metres after x seconds, then your acceleration is a constant $2\ \mathrm{m/s^2}$.

Example

Let $y = x^7 + 3x^5 + x^{\frac{3}{2}}$. Find $\dfrac{d^2y}{dx^2}$.

Solution

The first derivative is

$$\frac{dy}{dx} = 7x^6 + 15x^4 + \frac{3}{2}x^{\frac{1}{2}},$$

so the second derivative is

$$\frac{d^2y}{dx^2} = 42x^5 + 60x^3 + \frac{3}{4}x^{-\frac{1}{2}}.$$

Exercise 18

Let $f(x) = (x^2+7)^{100}$, as in exercise 11. What is $f''(x)$?

Exercise 19

Suppose the position of an object at time t is given by

$$x(t) = 1 - 7t + (t-5)^4.$$

Show that $x''(t) \geq 0$ for all t, so that acceleration is always non-negative.

Differentiation of inverses

A clever use of the chain rule arises when we have a function f and its *inverse* function f^{-1}. Refer to the module *Functions II* for a discussion of inverse functions.

Letting $y = f(x)$, we can express x as the inverse function of y:

$$y = f(x), \qquad x = f^{-1}(y).$$

The composition of f and its inverse f^{-1}, by definition, is just x. That is, $f^{-1}(f(x)) = x$. We can think of this diagrammatically as

$$x \xrightarrow{f} y \xrightarrow{f^{-1}} x.$$

Using the chain rule, we can differentiate this composition of functions to obtain

$$\frac{dx}{dx} = \frac{dx}{dy}\frac{dy}{dx}.$$

The derivative $\frac{dx}{dx}$ of x with respect to x is just 1, so we obtain the important formula

$$1 = \frac{dx}{dy}\frac{dy}{dx},$$

which can also be expressed as

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} \qquad \text{or} \qquad \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}.$$

This formula allows us to differentiate inverse functions, as in the following example.

Example

Let $y = \sqrt[3]{x}$. Find $\dfrac{dy}{dx}$.

Solution

(We assume that we only know the derivative of x^n when n is a positive integer.) The inverse function of the cube-root function is the cube function: if $y = \sqrt[3]{x}$, then $x = y^3$. We know the derivative of the cube function, $\frac{dx}{dy} = 3y^2$. We use this to find $\frac{dy}{dx}$:

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{3y^2}.$$

Substituting $y = \sqrt[3]{x} = x^{\frac{1}{3}}$ gives

$$\frac{dy}{dx} = \frac{1}{3x^{\frac{2}{3}}} = \frac{1}{3}x^{-\frac{2}{3}},$$

as expected.

The following exercises generalise the above example to find the derivative of $\sqrt[n]{x} = x^{\frac{1}{n}}$, and then of any rational power of x.

Exercise 20

Let n be a positive integer. Using the fact that the derivative of x^n is nx^{n-1}, prove that the derivative of $x^{\frac{1}{n}}$ is $\frac{1}{n}x^{\frac{1}{n}-1}$.

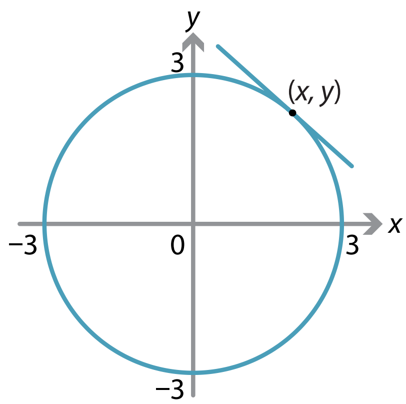
Exercise 21

Using the chain rule and the previous exercise, prove that for any rational number $\frac{p}{q}$ the derivative of $x^{\frac{p}{q}}$ is $\frac{p}{q}x^{\frac{p}{q}-1}$. (Here p, q are integers and $q > 0$.)

Implicit differentiation

Implicit differentiation is a powerful technique to find an instantaneous rate of change $\frac{dy}{dx}$ when there is an equation relating x and y. It applies even when y is not a function of x. All that is required is that there is an equation relating x and y.

For example, consider the curve in the plane described by the equation $x^2 + y^2 = 9$. This is a circle centred at the origin, of radius 3.



The circle $x^2 + y^2 = 9$.

Note that y is not a function of x! The circle fails the vertical-line test; we have a relation, not a function. (See the module *Functions I* for a discussion of functions and relations.) For a given value of x, there may be two distinct values of y:

$$x^2 + y^2 = 9 \quad \Longleftrightarrow \quad y = \pm\sqrt{9 - x^2}.$$

Although y is not a function of x, we can say that the equation $x^2 + y^2 = 9$ expresses y as an **implicit function** of x.

Suppose you want to find the gradient of the tangent to this circle at a point (x, y). One approach would be to consider either $f_1(x) = \sqrt{9 - x^2}$ or $f_2(x) = -\sqrt{9 - x^2}$, depending on whether y is positive or negative, and then differentiate. (We saw the derivative of $\sqrt{9 - x^2}$ in exercise 17.)

The much better approach of **implicit differentiation** is to *differentiate both sides* of the equation $x^2 + y^2 = 9$ with respect to x. Differentiating y^2 requires the chain rule, since y^2 is a function of y and y is a function of x:

$$\frac{d}{dx}(y^2) = \frac{d}{dy}(y^2)\,\frac{dy}{dx} = 2y\,\frac{dy}{dx}.$$

We can now differentiate each term of $x^2 + y^2 = 9$, and we obtain

$$2x + 2y\,\frac{dy}{dx} = 0.$$

Note the power of implicit differentiation: this equation involving $\frac{dy}{dx}$ is valid for *all* points on the circle except $y = 0$. When $y \neq 0$, we may solve for $\frac{dy}{dx}$ and obtain

$$\frac{dy}{dx} = -\frac{x}{y}.$$

Without implicit differentiation, we would not have found such a neat formula!

From this formula, we can see that if x, y have the same sign, then the gradient of the tangent is negative; while if x, y are of opposite sign, then the gradient is positive.

Exercise 22

Give a geometric argument why the tangent at the point (x, y) to a circle centred at the origin has gradient $-\dfrac{x}{y}$.

Example

What is the gradient of the curve $y^2 = x^3 - 7x$ at the points where $x = 4$?

Solution

When $x = 4$, we have $y^2 = 64 - 28 = 36$, so $y = \pm 6$. Implicit differentiation of $y^2 = x^3 - 7x$ gives

$$2y\,\frac{dy}{dx} = 3x^2 - 7.$$

So when $y \neq 0$, we obtain $\dfrac{dy}{dx} = \dfrac{3x^2 - 7}{2y}$. Hence:

- At $(x, y) = (4, 6)$, we have $\dfrac{dy}{dx} = \dfrac{3 \cdot 4^2 - 7}{2 \cdot 6} = \dfrac{41}{12}$.
- At $(x, y) = (4, -6)$, we have $\dfrac{dy}{dx} = \dfrac{3 \cdot 4^2 - 7}{2 \cdot (-6)} = -\dfrac{41}{12}$.

The curve in the previous example is called an *elliptic curve*.

Exercise 23

Consider the hyperbola $x^2 - y^2 = 5$.

a By writing y in terms of x, find the gradient of the tangent to the hyperbola at $(3, -2)$.

b Find the same gradient by implicit differentiation.

Links forward

Higher derivatives

So far in this module, we have looked at first and second derivatives. But there's no need to stop differentiating after doing it twice!

Given a function $f(x)$, if it can be differentiated k times, then we write the kth derivative as $f^{(k)}(x)$. In Leibniz notation, the kth derivative of y with respect to x is denoted by $\dfrac{d^k y}{dx^k}$.

Higher derivatives are useful for finding polynomial approximations to functions, as we see in the next section.

Exercise 24

Let n be a positive integer. What is the nth derivative of the function $f(x) = x^n$?

Approximations of functions

We have seen that the equation of the tangent line to the graph $y = f(x)$ at the point $x = a$ is given by

$$y = f(a) + f'(a)(x - a).$$

Therefore, if we define a linear function $T_1(x)$ as

$$T_1(x) = f(a) + f'(a)(x - a),$$

then $T_1(x)$ approximates $f(x)$ near $x = a$. In fact, since $T_1'(x) = f'(a)$, we have

$$T_1(a) = f(a) \qquad \text{and} \qquad T_1'(a) = f'(a).$$

So the functions $T_1(x)$ and $f(x)$ agree at $x = a$, and their derivatives also agree at $x = a$. In this sense, $T_1(x)$ is the 'best linear approximation' to $f(x)$ at $x = a$.

However, there's no reason to stop at *linear* approximations: we might ask for the best *quadratic* approximation to $f(x)$ at $x = a$.

Consider the function

$$T_2(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2.$$

This is a quadratic function. Its derivatives are

$$T_2'(x) = f'(a) + f''(a)(x-a) \qquad \text{and} \qquad T_2''(x) = f''(a),$$

so

$$T_2(a) = f(a), \qquad T_2'(a) = f'(a) \qquad \text{and} \qquad T_2''(a) = f''(a).$$

That is, $T_2(x)$ and $f(x)$ agree at $x = a$, as do their first *and second* derivatives. In this sense, $T_2(x)$ is the best quadratic approximation to $f(x)$ at $x = a$.

In general, one can check that the function

$$T_k(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k$$

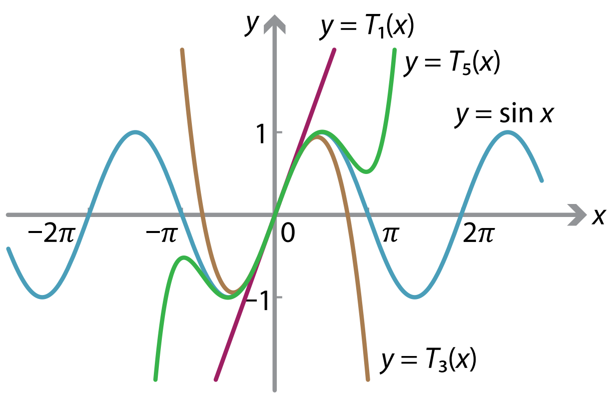
is a polynomial of degree k which agrees with $f(x)$ at $x = a$, along with all its first k derivatives. That is,

$$T_k(a) = f(a), \quad T_k'(a) = f'(a), \quad T_k''(a) = f''(a), \quad \dots, \quad T^{(k)}(a) = f^{(k)}(a).$$

In this sense, $T_k(x)$ is the best polynomial approximation of degree k to $f(x)$ at $x = a$.

The polynomials $T_k(x)$, called **Taylor polynomials**, are very important throughout mathematics and science for approximating functions. Taylor polynomials and their generalisations are studied extensively in university-level mathematics courses.

The following diagram shows some of the first few Taylor polynomials approximating the function $f(x) = \sin x$ near $x = 0$. Note how they provide successively better approximations to $f(x)$. (We will see how to differentiate $\sin x$ in the module *The calculus of trigonometric functions*.)



Taylor polynomials approximating $\sin x$ at $x = 0$.

History and applications

The discoverers of calculus

Today it is generally believed that calculus was discovered independently in the late 17th century by two great mathematicians: Isaac Newton and Gottfried Leibniz. However, the dispute over who first discovered calculus became a major scandal around the turn of the 18th century.

Like most scientific discoveries, the discovery of calculus did not arise out of a vacuum. In fact, many mathematicians and philosophers going back to ancient times made discoveries relating to calculus.

The ancient Greeks made many discoveries that we would today think of as part of calculus — however, mostly *integral* calculus, which will be discussed in the module *Integration*. Indian mathematicians in Kerala had developed Taylor polynomials for functions like $\sin x$ and $\cos x$ before 1500. (See the article *Was calculus invented in India?* listed in the *References* section.)

In the early 17th century, Fermat developed a method called *adequality* for finding where the derivative of a function is zero, that is, for solving $f'(x) = 0$. But it was not until Newton and Leibniz that gradients of tangents to curves could be calculated in general.

The Newton–Leibniz controversy

Newton described his version of differential calculus as 'the method of fluxions'. He wrote a paper on fluxions in 1666, but like many of his works, it was not published until decades later. His magnum opus *Philosophiae naturalis principia mathematica* (Mathematical principles of natural philosophy) was published in 1687. This work includes his theories of motion and gravitation, but does not include much calculus explicitly — although there is some explanation of calculus at the beginning, and Newton certainly used calculus to formulate his theories. Nonetheless, Newton's 'method of fluxions' did not explicitly appear in print until 1693.

Leibniz, on the other hand, published his first paper on calculus in 1684 — and claimed to have discovered calculus in the 1670s. From the published record, at least, Leibniz seemed to have discovered calculus first.

While Newton and Leibniz initially had a cordial relationship, Leibniz and his followers did not take kindly to a statement made by the English mathematician John Wallis. With a rather xenophobic and quarrelsome character, Wallis fought priority disputes on behalf of English scientists throughout his life. In 1695, perhaps inadvertently, Wallis intimated that Leibniz learned about calculus from Newton — a claim now known to be false.

Then, offended by a statement of Leibniz that certain mathematical problems could only be solved by Leibniz's own version of the calculus, a mathematician named Fatio de Duiller in 1699 accused Leibniz of plagiarism. Things only went downhill from there. It did not help matters that Newton and Leibniz also disagreed on philosophical questions.

In 1712 the Royal Society in England wrote a report purporting to settle the matter — except, the whole investigation was effectively directed by Newton himself. The report found that Leibniz had concealed his knowledge of Newton's work — based on facts now known to be false. In response, Leibniz accused Newton and his followers of stealing Leibniz's own calculus and making errors in their applications of it. The dispute went on well after Leibniz's death in 1716, full of accusations and counter-accusations.

Both Newton and Leibniz were capable of incredible mathematical discoveries, but their dispute demonstrated they were also capable of some rather less impressive behaviour.

Mathematically rigorous calculus

Both Newton's and Leibniz's versions of the calculus fell far short of the standards of rigour demanded by mathematicians today. Leibniz's infinitesimals like dx and Newton's fluxions were concepts that many argued were poorly defined or incoherent.

The most scathing criticism perhaps came from Bishop Berkeley (*The Analyst*, 1734), who ridiculed fluxions and infinitesimals:

> And what are these Fluxions? The Velocities of evanescent Increments? And what are these same evanescent Increments? They are neither finite Quantities nor Quantities infinitely small, nor yet nothing. May we not call them the Ghosts of departed Quantities?

It was not until over a century later that ideas like *limits* were formally introduced, and put on a firm mathematical footing, so that today we present the derivative as

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

As with many branches of mathematics, the way that calculus is taught and learned bears little relation to its historical development.

Appendix

Functions differentiable and not

We mentioned earlier that not all functions have derivatives. We now say a little more, although a full discussion of these issues is the subject matter of higher level university courses on *analysis*.

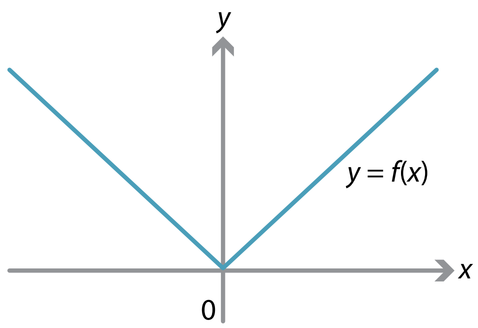
For a function to have a derivative, it must first of all be *continuous* — you must be able to draw the graph without taking your pen off the page! More rigorously, continuity at a point $x = a$ means that we must have $\lim_{x \to a} f(x) = f(a)$.

Roughly speaking, a function will only have a derivative where its graph is *smooth*, and can be drawn on the page without sharp corners.

For instance, consider the *absolute value function*

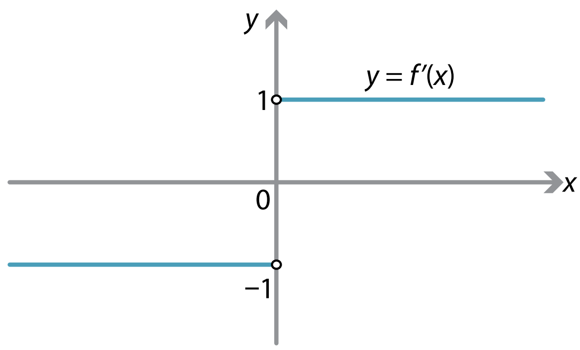
$$f(x) = |x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

The following diagram shows the graph of $y = f(x)$.



Graph of $f(x) = |x|$.

For $x > 0$, we have $f(x) = x$ and so $f'(x) = 1$. For $x < 0$, we have $f(x) = -x$ and so $f'(x) = -1$. However, at $x = 0$ we have a problem.



Graph of $f'(x)$.

By definition, if $f'(0)$ were to exist,

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

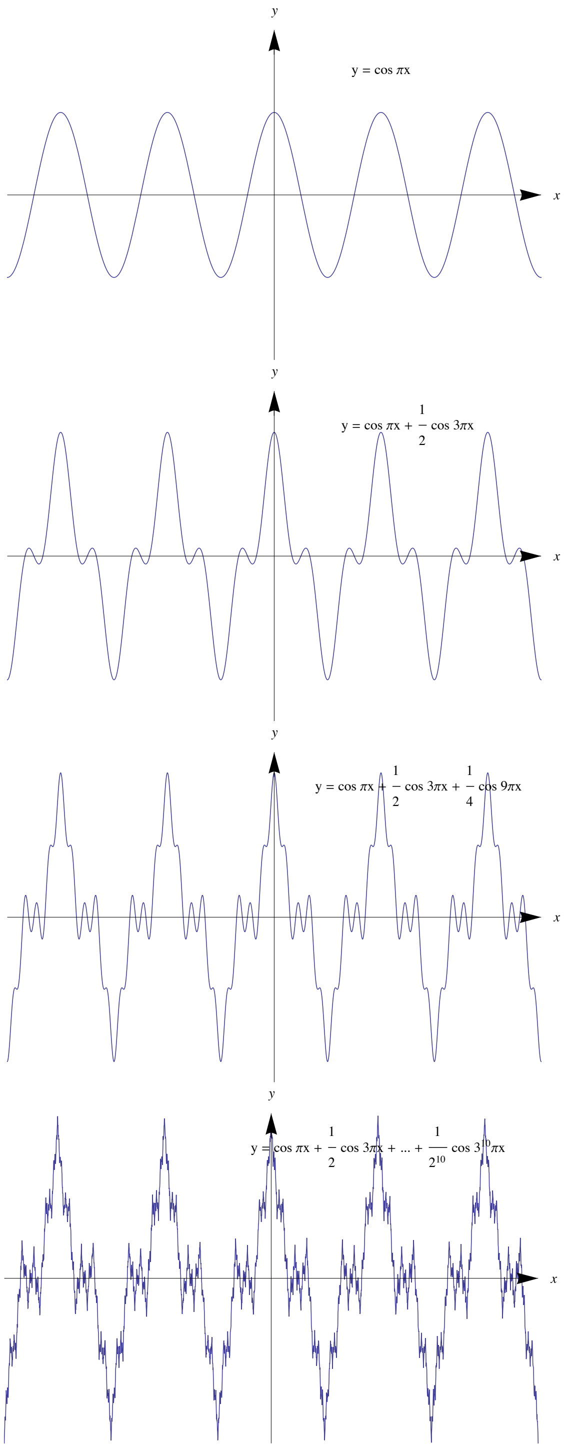
Unfortunately, for any $\Delta x > 0$ we have $\frac{|\Delta x|}{\Delta x} = 1$, while for any $\Delta x < 0$ we have $\frac{|\Delta x|}{\Delta x} = -1$. So if Δx approaches 0 from above, the limit is 1, but if Δx approaches 0 from below, the limit is -1. Therefore, the limit does not exist, and $f'(0)$ is not defined.

The function $f(x) = |x|$ is therefore differentiable at all $x \neq 0$, but is not differentiable at $x = 0$. It is differentiable almost everywhere, but not everywhere.

The absolute value function, however, is rather tame in comparison to examples like the following. Consider the infinite series

$$f(x) = \cos(\pi x) + \frac{1}{2}\cos(3\pi x) + \frac{1}{2^2}\cos(3^2 \pi x) + \cdots = \sum_{n=0}^{\infty} \frac{1}{2^n}\cos(3^n \pi x).$$

The following graphs show a few partial sums of this series.



Graphs of approximations to a nowhere-differentiable function.

Note how the functions oscillate more rapidly and the graphs become ‘bumpier’ as each term is added. In the limit, the function oscillates so wildly that, although it remains continuous, it is too bumpy for the derivative to exist. It can be shown that this function is continuous everywhere, but is *differentiable nowhere*!

Answers to exercises

Exercise 1

Let $f(x)=x^4$. We first compute

$$f(-1+\Delta x)=(-1+\Delta x)^4=1-4(\Delta x)+6(\Delta x)^2-4(\Delta x)^3+(\Delta x)^4$$

and $f(-1)=1$. The gradient at $x=-1$ is then given by

$$\begin{aligned}\lim_{\Delta x\to 0}\frac{f(-1+\Delta x)-f(-1)}{\Delta x} &= \lim_{\Delta x\to 0}\frac{-4(\Delta x)+6(\Delta x)^2-4(\Delta x)^3+(\Delta x)^4}{\Delta x}\\ &= \lim_{\Delta x\to 0}\left(-4+6(\Delta x)-4(\Delta x)^2+(\Delta x)^3\right)=-4.\end{aligned}$$

The tangent line at $(-1,1)$ has gradient -4, and hence has equation $y-1=-4(x+1)$ or, equivalently, $y=-4x-3$.

Exercise 2

Let $f(x)=x^3$. We compute

$$\begin{aligned}f'(x) &= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}=\lim_{\Delta x\to 0}\frac{(x+\Delta x)^3-x^3}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{3x^2(\Delta x)+3x(\Delta x)^2+(\Delta x)^3}{\Delta x}=\lim_{\Delta x\to 0}\left(3x^2+3x(\Delta x)+(\Delta x)^2\right)=3x^2.\end{aligned}$$

Exercise 3

Let $f(x)=c$. Then, for any x and Δx, we have $f(x+\Delta x)-f(x)=c-c=0$. Hence, the derivative is

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

Exercise 4

Let $f(x)=ax+b$. Then

$$\begin{aligned}f'(x) &= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}=\lim_{\Delta x\to 0}\frac{a(x+\Delta x)+b-ax-b}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{a(\Delta x)}{\Delta x}=\lim_{\Delta x\to 0}a=a.\end{aligned}$$

Exercise 5

Let $f(x) = \frac{1}{x}$. We first compute the quotient

$$\frac{f(x+\Delta x)-f(x)}{\Delta x} = \frac{1}{\Delta x}\left(\frac{1}{x+\Delta x}-\frac{1}{x}\right) = \frac{1}{\Delta x}\cdot\frac{x-(x+\Delta x)}{x(x+\Delta x)}$$
$$= \frac{-\Delta x}{(\Delta x)x(x+\Delta x)} = -\frac{1}{x(x+\Delta x)}.$$

Hence, the derivative is

$$f'(x) = \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x} = \lim_{\Delta x\to 0} -\frac{1}{x(x+\Delta x)} = -\frac{1}{x^2}.$$

Thus the derivative of $f(x) = \frac{1}{x}$ is $f'(x) = -\frac{1}{x^2}$.

Exercise 6

The derivative of $f(x) - g(x)$ is given by

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

Hence, the derivative of $f(x) - g(x)$ is $f'(x) - g'(x)$.

Exercise 7

Rewriting $f(x)$ as $x^3 + x^{-\frac{3}{2}}$, we obtain $f'(x) = 3x^2 - \frac{3}{2}x^{-\frac{5}{2}}$.

Exercise 8

We assume that the derivative of x is 1. We now prove that, if the derivative of x^n is nx^{n-1}, then the derivative of x^{n+1} is $(n+1)x^n$. To do this we use the product rule:

$$\frac{d}{dx}\left(x^{n+1}\right) = \frac{d}{dx}\left(x^n\cdot x\right) = x^n\frac{d}{dx}(x) + x\frac{d}{dx}(x^n)$$
$$= x^n\cdot 1 + x\cdot nx^{n-1} = (n+1)x^n.$$

It follows by induction that $\frac{d}{dx}(x^n) = nx^{n-1}$, for all positive integers n.

Exercise 9

Using the product rule, we have

$$\frac{d}{dx}f(x)^2 = \frac{d}{dx}\big(f(x)\cdot f(x)\big) = f(x)\cdot\frac{d}{dx}\big(f(x)\big) + f(x)\cdot\frac{d}{dx}\big(f(x)\big)$$

$$= f(x)\,f'(x) + f(x)\,f'(x) = 2\,f(x)\,f'(x).$$

Exercise 10

We first use the product rule on the product of $f(x)\,g(x)$ and $h(x)$:

$$\frac{d}{dx}\big[f(x)\,g(x)\,h(x)\big] = \frac{d}{dx}\big[f(x)\,g(x)\big]\,h(x) + f(x)\,g(x)\,\frac{d}{dx}\big[h(x)\big].$$

Then we use the product rule on $f(x)\,g(x)$:

$$\frac{d}{dx}\big[f(x)\,g(x)\,h(x)\big] = \big(f'(x)\,g(x) + f(x)\,g'(x)\big)\,h(x) + f(x)\,g(x)\,h'(x)$$

$$= f'(x)\,g(x)\,h(x) + f(x)\,g'(x)\,h(x) + f(x)\,g(x)\,h'(x).$$

For a general product $f_1(x)\,f_2(x)\cdots f_n(x)$, the derivative is a sum of n terms, with $f_i'(x)$ occurring in the ith term:

$$\frac{d}{dx}\big[f_1(x)\,f_2(x)\cdots f_n(x)\big]$$

$$= f_1'(x)\,f_2(x)\cdots f_n(x) + f_1(x)\,f_2'(x)\,f_3(x)\cdots f_n(x) + \cdots + f_1(x)\cdots f_{n-1}(x)\,f_n'(x).$$

Exercise 11

Let $f(x) = (x^2+7)^{100} = g(h(x))$, where $g(x) = x^{100}$ and $h(x) = x^2+7$. Then $g'(x) = 100x^{99}$ and $h'(x) = 2x$, so by the chain rule

$$f'(x) = g'(h(x))\,h'(x) = 100(x^2+7)^{99}\cdot 2x = 200x(x^2+7)^{99}.$$

Exercise 12

We can write $f(x)^2$ as $h(f(x))$ where $h(x) = x^2$. The chain rule then gives

$$\frac{d}{dx}h(f(x)) = h'(f(x))\,f'(x) = 2\,f(x)\,f'(x).$$

Exercise 13

We first think of $f(g(h(x)))$ as the composition of $f(x)$ and $g(h(x))$, so the chain rule gives

$$\frac{d}{dx}\big[f(g(h(x)))\big] = f'(g(h(x)))\,\frac{d}{dx}g(h(x)).$$

Then using the chain rule again gives

$$\frac{d}{dx}\big[f(g(h(x)))\big] = f'(g(h(x)))\,g'(h(x))\,h'(x).$$

In general, for the composition of n functions $f_1 \circ f_2 \circ \cdots \circ f_n$, the derivative is a product of n factors, and the ith factor is $f_i'(f_{i+1}(\cdots(f_n(x))\cdots))$.

Exercise 14

Let $g(x) = x^n$ and $h(x) = \frac{1}{x}$, so that $\frac{1}{x^n} = h(g(x))$. Then $g'(x) = nx^{n-1}$ and $h'(x) = -\frac{1}{x^2}$. By the chain rule,

$$\frac{d}{dx}\left(\frac{1}{x^n}\right) = \frac{d}{dx}h(g(x)) = h'(g(x))\,g'(x)$$

$$= -\frac{1}{(g(x))^2}\,g'(x) = -\frac{1}{x^{2n}}\,nx^{n-1} = -nx^{-n-1}.$$

Thus, the derivative of x^{-n} is $-nx^{-n-1}$.

Exercise 15

Let $h(x) = \frac{1}{x}$. Then $h'(x) = -\frac{1}{x^2}$ and $\frac{1}{g(x)} = h(g(x))$. By the chain rule,

$$\frac{d}{dx}\left(\frac{1}{g(x)}\right) = \frac{d}{dx}h(g(x)) = h'(g(x))\,g'(x) = -\frac{1}{(g(x))^2}\,g'(x) = -\frac{g'(x)}{(g(x))^2}.$$

Exercise 16

Since the derivative of $\frac{1}{g(x)}$ is $-\frac{g'(x)}{(g(x))^2}$, the product rule gives

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{1}{g(x)}\,\frac{d}{dx}\big(f(x)\big) + f(x)\,\frac{d}{dx}\left(\frac{1}{g(x)}\right)$$

$$= \frac{f'(x)}{g(x)} + f(x)\left(-\frac{g'(x)}{(g(x))^2}\right) = \frac{g(x)\,f'(x) - f(x)\,g'(x)}{(g(x))^2}.$$

This is the quotient rule.

Exercise 17

Let $y = \sqrt{9 - x^2}$. We differentiate:

$$\frac{dy}{dx} = \frac{1}{2}(9 - x^2)^{-\frac{1}{2}} \cdot (-2x) = \frac{-x}{\sqrt{9 - x^2}}.$$

So, at $x = \frac{3\sqrt{2}}{2}$, we have $y = \frac{3\sqrt{2}}{2}$ and $\frac{dy}{dx} = \frac{-3\sqrt{2}}{2}\,\frac{2}{3\sqrt{2}} = -1$. The tangent line has gradient -1 and passes through the point $(\frac{3\sqrt{2}}{2}, \frac{3\sqrt{2}}{2})$, and hence has equation $y = -x + 3\sqrt{2}$.

Exercise 18

From $f(x) = (x^2+7)^{100}$, we have $f'(x) = 200x(x^2+7)^{99}$. Using the product and chain rules, we obtain

$$\begin{aligned} f''(x) &= 200x\,\frac{d}{dx}\big[(x^2+7)^{99}\big] + (x^2+7)^{99}\,\frac{d}{dx}\big[200x\big] \\ &= 200x\cdot 99(x^2+7)^{98}\cdot 2x + (x^2+7)^{99}\cdot 200 \\ &= 200(x^2+7)^{98}(199x^2+7). \end{aligned}$$

Exercise 19

We compute the derivatives of $x(t) = 1 - 7t + (t-5)^4$ with respect to t:

$$\begin{aligned} x'(t) &= -7 + 4(t-5)^3 \\ x''(t) &= 12(t-5)^2. \end{aligned}$$

Since squares are non-negative, we have $x''(t) \geq 0$ for all t. That is, the acceleration is always non-negative.

Exercise 20

Let $y = x^{\frac{1}{n}}$, where n is a positive integer. We wish to find $\frac{dy}{dx}$. We have $x = y^n$, and so $\frac{dx}{dy} = ny^{n-1}$. Thus

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{ny^{n-1}} = \frac{1}{n}y^{1-n},$$

and substituting $y = x^{\frac{1}{n}}$ gives

$$\frac{dy}{dx} = \frac{1}{n}x^{\frac{1-n}{n}} = \frac{1}{n}x^{\frac{1}{n}-1},$$

as expected.

Exercise 21

Let $y = x^{\frac{p}{q}}$, where p, q are integers with $q > 0$. We wish to find $\frac{dy}{dx}$. We let $u = x^{\frac{1}{q}}$. Then $y = u^p$ and, by the previous exercise, $\frac{du}{dx} = \frac{1}{q}x^{\frac{1}{q}-1}$. The chain rule then gives

$$\frac{dy}{dx} = \frac{dy}{du}\,\frac{du}{dx} = pu^{p-1}\cdot\frac{1}{q}x^{\frac{1}{q}-1} = \frac{p}{q}u^{p-1}x^{\frac{1}{q}-1}.$$

Substituting $u = x^{\frac{1}{q}}$ gives

$$\frac{dy}{dx} = \frac{p}{q}x^{\frac{p-1}{q}}x^{\frac{1}{q}-1} = \frac{p}{q}x^{\frac{p}{q}-1}.$$

Exercise 22

If the circle is centred at the origin, then the radius of the circle from $(0,0)$ to (x,y) has gradient $\frac{y}{x}$. The tangent to the circle is perpendicular to the radius, and hence its gradient is the negative reciprocal of $\frac{y}{x}$, that is, the gradient is $-\frac{x}{y}$.

Exercise 23

a From $y^2 = x^2 - 5$, we have $y = \pm\sqrt{x^2-5}$. As we want to include the point $(3,-2)$, we take the negative square root and consider $y = -\sqrt{x^2-5}$. Then

$$\frac{dy}{dx} = -\frac{1}{2}(x^2-5)^{-\frac{1}{2}} \cdot 2x = \frac{-x}{\sqrt{x^2-5}}.$$

At $x = 3$, we have $\frac{dy}{dx} = \frac{-3}{\sqrt{4}} = -\frac{3}{2}$.

b Implicit differentiation of $x^2 - y^2 = 5$ gives $2x - 2y\frac{dy}{dx} = 0$, and so $\frac{dy}{dx} = \frac{x}{y}$. Hence, at the point $(3,-2)$, we have $\frac{dy}{dx} = -\frac{3}{2}$.

Exercise 24

The kth derivative of x^n is $n(n-1)\cdots(n-k+1)x^{n-k}$, and hence the nth derivative is the constant $n!$.

References

- David Bressoud, ‘Was calculus invented in India?’, *College Mathematics Journal* 33 (2002), 2–13.
- Alfred Rupert Hall, *Philosophers at War: The Quarrel Between Newton and Leibniz*, Cambridge University Press, 1980.

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