

MAT1001 Differential Calculus: Lecture Notes

Calum Ross

Last modified: February 13, 2026

Contents

1	Why Do We Need Calculus	1
1.1	Course Overview	1
1.2	Why Calculus	2
2	Functions	7
2.1	What is a function	7
2.2	Trigonometric functions	10
2.3	Logarithms and Exponentials	14
2.4	Hyperbolic functions	18
2.5	Limits and asymptotics	20
2.6	Continuity and differentiability	24
3	Differentiation	29
3.1	Differentiation from first principles	29
3.2	Differentiation techniques	32
3.3	Multiple derivatives	37
3.4	Applications of differentiation	38
4	Integration	43
4.1	Integration and antiderivatives	43
4.2	Techniques for integration	46
4.3	Definite integrals	50
4.4	Applications of integrals	58
5	Differential Equations	61
6	Numerical Methods	63
6.1	Solving equations numerically	63
6.2	Approximating functions	70
6.3	Numerical integration	74
6.4	Numerical differentiation	80

6.5	Numerical approaches to differential equations	84
7	Calculus in Computer Science	85
7.1	LLMs and AI	85
7.2	Robotics	87
8	Advanced Topics	89
8.1	L'Hôpital's rule for evaluating limits	89
8.2	Trigonometric Derivatives in Degrees	90
8.3	Functions of two variables	91
8.4	Multiple integrals	91
8.5	Optimisation Problems	91
8.6	Polynomial approximation	92
9	Background Mathematics	93
9.1	Background and References	93
9.2	Polynomials and roots	93
9.3	Trigonometry Primer	95
9.4	Radians and Degrees	96
9.5	Rearranging Equations	97
9.6	More on Limits	100
10	Further Reading	103
10.1	Why Calculus Extra Reading	103
10.2	Functions Extra Reading	104
10.3	Differentiation Extra Reading	104
10.4	Integration Extra Reading	104
10.5	Differential Equations Extra Reading	104
10.6	Numerical Methods Extra Reading	104
10.7	Advanced Topics Extra Reading	105
11	Tutorial Sheets	107
11.1	Week 1	107
11.2	Week 2	109
11.3	Week 3	111
12	Extra Proofs and Derivations	113
13	Standard Derivatives and Integrals	115
13.1	Derivatives	115
13.2	Integrals	115

Chapter 1

Why Do We Need Calculus

Calculus: A branch of mathematics using the idea of a limit and generally divided into two parts: integral and differential calculus.

*Penguin Dictionary of
Mathematics*

1.1 Course Overview

These lecture notes cover the same material as the lectures for **MAT1001 Differential Calculus**. They will be updated during the module so please check back regularly for the latest version, the date of the most recent update is shown on the front page. In particular, most of the figures used in the lectures are hand drawn and it takes some time to produce good quality digital versions of these.

It is important to note that **reading these notes is not a substitute for attending the lectures and tutorials!** While I have included everything that I intend to cover in the lectures, and have often given more detail in these notes, the direction that we go in lectures will be guided by student questions and the example problems that you are given to solve. These are not reflected in the notes as I cannot anticipate all of the questions that may be asked. If you want to pass the module then the best way to do this is to attend the lectures and tutorials and to solve all of the practice problems that you are given.

In week one there was overview information about the module including a rough schedule, how you will be assessed, and how to contact me. That will not be duplicated here and can be found on the module's virtual learning environment (VLE) page. If you need have any questions during the module then you can send me an email or come by my office during the office hour times. The module reading list, available on the VLE, contains details of recommended and supplementary reading for the module, there will frequently be suggested reading given to complement the material from a given weeks lecture. In these notes I will occasionally cite one of these books in the lecture notes, there will also be extra information and resources given in chapter 10. My favourite reference for much of this material is [Riley et al., 2006], which covers all the areas of mathematics that we need for this course.

We will mainly be using content from chapters 1, 2, and parts of 27 from [Riley et al., 2006]. Another good resource is the calculus page on the website Paul's Online Notes. There are lots of examples given there, some of which I have included in this lecture notes.

Another book that I have sometimes used is [Jordan and Smith, 2008], it contains lots of background material but does not always have the simplest explanations. For numerical methods [Lissamen and West, 2004] contains a lot of material, and there are several copies in the library.

1.2 Why Calculus

So what does calculus add for me? It provides a way for us to construct relatively simple quantitative models of change, and to deduce their consequences.

MIT Calculus for Beginners

Calculus is an important topic in mathematics and at its most basic is the study of how functions change. It makes an appearance whenever we are interested in solving an optimisation problem, e.g. finding the best solution to a given problem. This could be finding out the largest area that can be enclosed by a fence of a given length, or what the fastest way to travel between two different points is. One example of this is sometimes called the **Ocado** or **vehicle routing problem** and wants to find the optimal route for a delivery van¹.

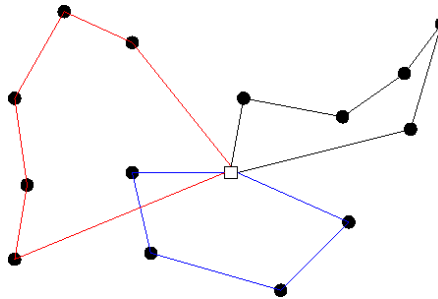


Figure 1.1: A schematic of the vehicle routing problem. This is a programming problem where you want to find the optimal set of routes for a fleet of vehicles to travel when making deliveries to a given set of customers, the dots in the figure. This sort of optimisation problem is important for any delivery service as making their routes more efficient can save them money.

You may be thinking, “that all sounds sensible in mathematics or physics, but as a computer science why should I care about calculus?” If so you are not alone, many people ask the same question and there are many answers given online. For example the Medium post [The uses of Calculus in Computer Science](#) highlights several areas:

- **Algorithms and optimisation:** As highlighted above, whenever you have an optimisation problem calculus invariably rears its head.
- **Numerical analysis/ Computer graphics and computer vision:** Calculus provides a language for solving complex physical and mathematical problems that we are unable to solve by hand. This is important if you want to simulate a physical system, whether in a piece of scientific research or because you are designing a computer game, but also in graphics rendering and computer vision where you need to understand and simulate the path that light rays will follow.
- **AI and machine learning:** Most machine learning algorithms involve calculus in their optimisation and training.

¹More info on this can be found [here](#). Sometimes the Ocado problem refers to routing of the robots that pick up groceries in the warehouse which is another example of a vehicle routing problem.

- **Data science:** Much of data science involves statistical and mathematical modelling to analyse and understand the data. Understanding these techniques relies on understanding calculus.

This all goes to show that calculus is an essential tool in many field beyond pure mathematics and is indispensable for Physicists, Engineers, and computer programmers.

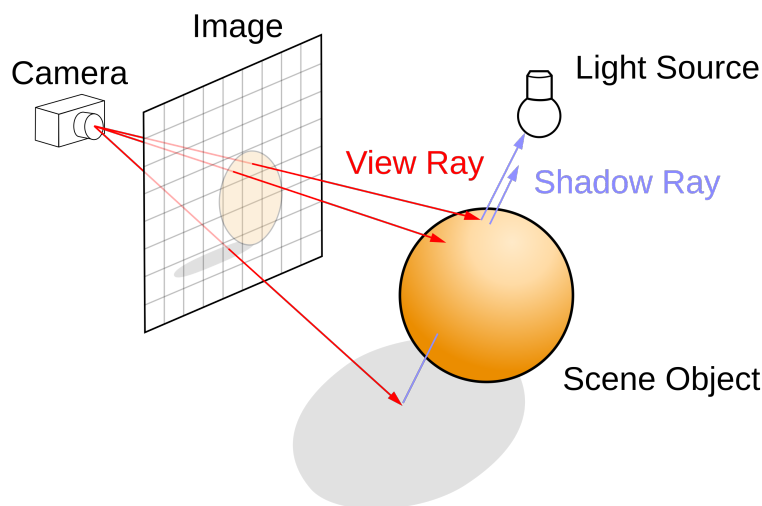


Figure 1.2: A schematic showing how a ray tracing algorithm works to build up an image by extending light rays. This image was made by Henrik for Wikimedia Commons.

Calculus is separated into two main areas: **Differentiation** and **Integration**. Differentiation is the study of how quickly or slowly a function varies. If we have a graph of a function the derivative is the gradient to this graph, this notion will be made precise in a later section. Integration is the “opposite” of differentiation and corresponds to finding the area under the graph of a function.

Both concepts can be extended to much more abstract settings but the intuition gained from thinking about graphs of functions, their gradients, and the area under them, will set you in good stead for everything that comes after this.

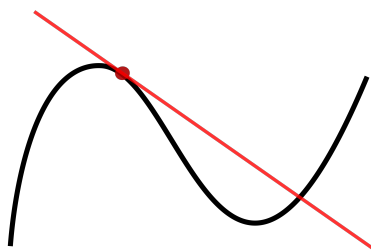


Figure 1.3: The derivative of a function at a point is equivalent to finding the tangent to a curve at this point.

The YouTube channel 3Blue1Brown has an excellent playlist called the essence of calculus available here which goes through the basics of calculus and its implications in a fairly unique way. Grant’s goal in the videos is to give you an intuition for why Calculus works the way it does, rather than just giving you a list of facts and equations to memorise. Hopefully you will find that this module takes a similar approach, the goal is for you to understand the how and why of the mathematics and not just approach it as a set of rules to memorise.

If you find that you enjoy any of the topics in this module, or just want to read about interesting mathematics in a popular science setting I recommend having a look at *Chalkdust Magazine* magazine. As one of the

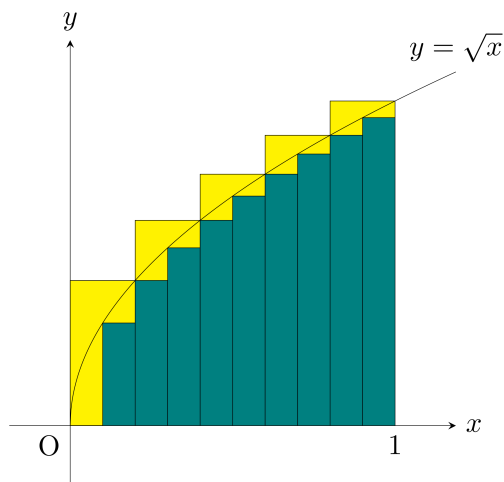


Figure 1.4: The integral, or area under the curve, can be approximated by rectangles. Depending on the size of the rectangles this can be an over estimate or an underestimate, but it will improve the thinner the rectangles become. This figure comes from Wikimedia Commons.

editors of *Chalkdust* I may be biased, but I think that it is a nice way to find out about some interesting maths without having to go into all of the details.

Hopefully, this has given you a feel for why you are doing this module. Now we can get started with the mathematics. As calculus is the study of how functions change we first need to remind ourselves of some of the properties of functions and introduce some notation.

How to use these notes

As stated above, these notes are here to complement the in person lectures rather than to replace them. For some topics there will be more detail given here, while at other points there may be explanations that I give in the lectures, or examples that I use which do not appear in these notes. Also, I will frequently draw pictures in the lectures, as the module goes on I will try to add these figures to the lecture notes. However, that will not always be possible. These notes also contain exercises, some of which will appear in the tutorial sheets, that you should solve to help reinforce the concepts from the module.

To complement the theory presented in these notes you will find a range to examples and exercises to help you build familiarity with the mathematical concepts that we will meet during this module. As this is not a module for mathematicians we will not always completely precise and give all the details, but rather focus on the practical applications of the mathematics. Occasionally I will include a **Mathematical Diversion** where I give some more of these details. If you are interested in the mathematics for its own sake then these will hopefully be interesting to you. If you are only interested in the mathematics that you need to pass this module, then feel free to skip these diversions.

There are also **Real World Application** examples at certain points in these notes. These are there to highlight how the mathematics that you are learning in this module plays a role in what you will see later in your degree, as well as showing you how it is relevant to your everyday life. Initially there are not many of these sections in the notes but more will be added as the notes are expanded.

These lecture notes will continue to develop as the module goes on so make sure to check back frequently to see what has been added. There is also an online html version of these notes that you can check out as well. The html version should be accessible with alttext on the figures. If you have any difficulty viewing either

these notes or the html version let me know.

Chapter 2

Functions

The art of doing mathematics is finding that special case that contains all the germs of generality.

David Hilbert

2.1 What is a function

The concept of a function is essential to understand not just calculus but also computer programming. We can think of a function as being a black box that takes in information, potentially changes it in some way, and then outputs information.

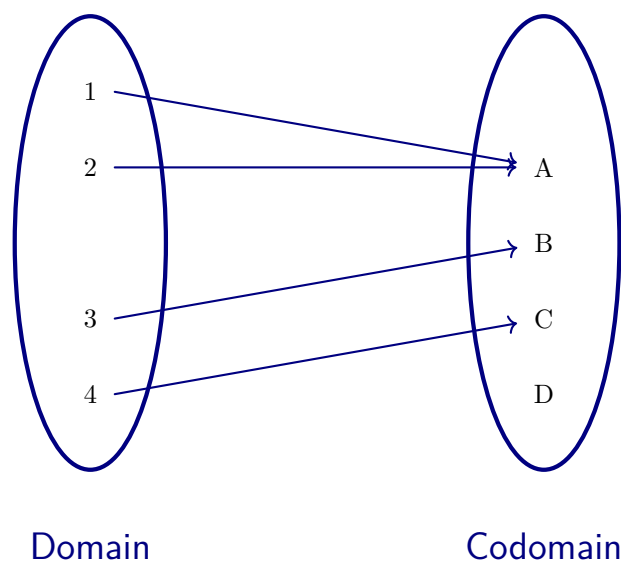


Figure 2.1: A schematic of how a function behaves, it maps objects 1, 2, 3, ... in the Set X to objects A, B, C, D, \dots in the set Y .

Being more precise, a function is a map between two spaces, the domain (X), and the codomain (Y). e.g.

$$f : X \rightarrow Y$$

and it sends a point $x \in X$ to a point $f(x) = y \in Y$. The domain is the space of all allowed inputs, sometimes called the argument of the function. While the codomain is the space that all output of the function have to be in. Many functions cannot reach the whole of the codomain, the subset of outputs that are actually reached by a function are called its image. Functions have the property that they map a point $x \in X$ to a single point $y \in Y$, for example:

$$y = f(x) = x^2 + 1,$$

is a function since every value of x gives a single value of y . If we consider what this looks like for a few examples we get:

$$\begin{aligned} x, &\mapsto f(x) = x^2 + 1 \\ 0, &\mapsto 0 + 1 = 1, \\ \pm 1, &\mapsto 1 + 1 = 2, \\ \pm 2, &\mapsto 4 + 1 = 5 \\ \pm 3, &\mapsto 9 + 1 = 10 \end{aligned}$$

However,

$$y^2 = x + 1,$$

is not a function since every x corresponds to two values of y , e.g. if $x = 3$ then $y^2 = 3 + 1 = 4$ and y can be both 2 and -2 .

Mathematical Diversion 2.1. *Note that we some times refer to a square root function $\sqrt{} : x \rightarrow \sqrt{x}$, even though has more than two outputs for a given input. e.g. $\sqrt{4} = \pm 2$. It is an example of a multi valued function, and they are important in mathematics, but require care.*

Note that when we write $f(x)$ we are not saying f times x , but mean that f is a function of x , e.g. f takes in a value of x and returns a number $y = f(x)$. Sometimes we refer to x as the argument of the function $f(x)$.

Throughout this module we will meet several different types of functions so you will need to get familiar with this notation and understand how to evaluate functions. There will be some questions in the tutorial sheet to help you practice this.

A key property of a function is its **Roots** or zeros, these are the values of x such that $f(x) = 0$. For linear functions finding the roots is just a matter of rearranging the equation, sometimes referred to as changing the subject. In this module I am assuming that you have some familiarity with this, if not take a look at the background material in chapter 9 or ask me to point you towards more resources. There will also be some revision questions on this topic in the first couple of weeks tutorials.

A common type of function where it is, relatively, easy to find the roots is a polynomial. A polynomial is a function given by the sum of powers of a variable x , or several variables. There is more information on polynomials in one variable in the Polynomials and roots section of chapter 9.

Example 2.1. *Find all the roots of $f(x) = 9x^3 - 18x^2 + 6x$.*

Remember the roots are where $f(x) = 0$ so we are solving $9x^3 - 18x^2 + 6x = 0$. First notice that there is a common factor of $3x$ in all of the terms so we can factor that out and get

$$0 = 9x^3 - 18x^2 + 6x = 3x(3x^2 - 6x + 2).$$

This means that either x is zero or the quadratic expression in brackets, $3x^2 - 6x + 2$, is zero. This means that $x = 0$ is a root of the equation, and there will be two more that we find by solving the quadratic. This can be done by using the quadratic formula:

$$\begin{aligned} x &= \frac{6 \pm \sqrt{(-6)^2 - 4(3)(2)}}{2(3)} \\ &= \frac{6 \pm \sqrt{12}}{6} \\ &= 1 \pm \frac{\sqrt{3}}{3} \\ &= 1 \pm \frac{1}{\sqrt{3}}. \end{aligned}$$

So the three roots of $f(x)$ are

$$x = 0, \quad x = 1 + \frac{\sqrt{3}}{3}, \quad x = 1 - \frac{\sqrt{3}}{3}.$$

If you do not remember how to use the quadratic formula then I suggest that you look at the background material in chapter 9.

If this was a course for mathematicians this is where we would spend time talking about the domain and range of a function. How they are defined, and when we need to be careful to avoid dividing by zero. Here, we will not go through this, and I will just remind you that in your algebraic manipulations you should not divide by any quantity that is zero.

A simple, yet useful, class of functions are the rational functions. There are functions who are the ratio of two polynomials, e.g.

$$f(x) = \frac{4x + 10}{x^2 - 2x - 15}.$$

A useful way to understand a function is draw a graph. In a graph, the domain of the function X is drawn horizontally and the codomain, Y , is drawn vertically, using **Cartesian axes** (x, y) . The graph consists of the points (x, y) with $y = f(x)$. In many books x is called the independent variable and y the dependent variable. In the tutorials we will discuss how to produce plots using a computer, either using **MATLAB**, **Python**, or **WolframAlpha**. There are also online graphing programs like **desmos** and **GeoGebra**. In these notes most of the graphs are produced using the Tikz package for L^AT_EX, which is a typesetting and markup language¹.

Example 2.2. Plot the graph of $f(x) = 2x^2 - 1$.

Note that the equation $y^2 = x + 1$ also defines a parabola, see fig. 2.3. However if we plot this function the curve is not the graph of a function. You should think about why this is.

Example 2.3. The function $f(x) = \sin(x)$ has the graph shown in fig. 2.4.

Example 2.4. Consider the two functions $f(x) = 3x - 2$ and $g(x) = x/3 + 2/3$. These satisfy the relationship

$$\begin{aligned} (f \circ g)(x) &= f(g(x)) \\ &= f\left(\frac{x}{3} + \frac{2}{3}\right) \\ &= 3\left(\frac{x}{3} + \frac{2}{3}\right) - 2 \\ &= x + 2 - 2 = x \end{aligned}$$

¹If you want to produce professional looking documents which include mathematical formulas then it is worth your time learning how to use L^AT_EX.

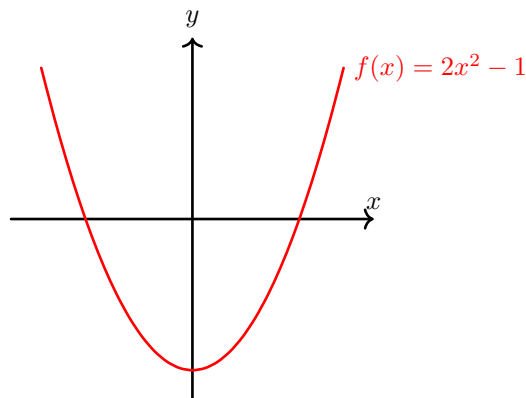


Figure 2.2: The graph of the quadratic function $f(x) = 2x^2 - 1$ is a parabola which intersects the y -axis at $x = 0$ and the x -axis at $x = \pm \frac{1}{\sqrt{2}}$.

and

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

When we have two functions whose composition leaves x unchanged, we say that they are **Inverse** to each other, and g is the *inverse*² of f . As can be seen in Section 1.2 of [Dawkins, 2025b] we can understand this a meaning that if $f : x \mapsto y$ then $g : y \mapsto x$, e.g. $f(-1) = -5$ while $g(-5) = -1$.

The concept of an inverse function makes intuitive sense. However, if we pretend to be mathematicians and treat this carefully we quickly encounter some problems. For example, what happens if we have a function which sends two different values of x to the same value? Then we are unable to know which value we started with, so cannot build an inverse function. For example $f(x) = x^2$ sends both x and $-x$ to x^2 so we cannot find an inverse that works everywhere, this is why when we first meet the square root function you often see it as $\pm\sqrt{}$.

Mathematicians fix this by introducing the notion of a **One-to-one** function, sometimes called an **injective** function. A function is called one-to-one if no two values of x produce the same value of y , so that

$$f(x_1) \neq f(x_2) \quad \text{for } x_1 \neq x_2.$$

The advantage of one-to-one functions is that we can find inverses for them. If we have two one-to-one functions which satisfy

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

then f and g are inverses and we write $g(x) = f^{-1}(x)$.

2.2 Trigonometric functions

Now that we have discussed functions and how to graph them, we can focus on some specific functions which it is very common to encounter.

From your previous maths experience you have probably come across **Trigonometric** (trig) functions in the context of triangles, where they are used to calculate lengths and angles. If you do not remember how this works then I suggest that you have a look at the Trigonometry primer in chapter 9 or check out some revision material available here. Another important thing to remember is that you should always work in radians rather than degrees when using trig functions. This is because radians are a more natural unit for

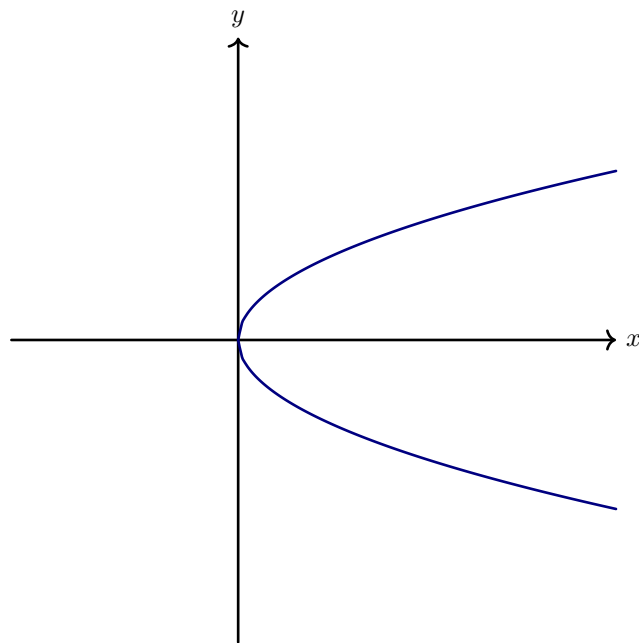


Figure 2.3: A plot of $y^2 = x + 1$, you should think about why this is not the graph of a function.

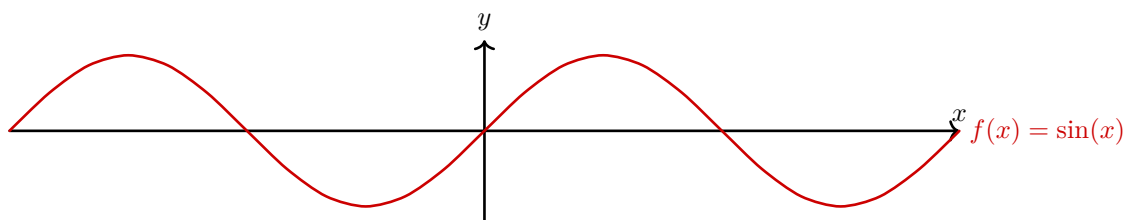


Figure 2.4: The graph of the sine function $f(x) = \sin(x)$.

angles and if we did not use them lots of formulas would need extra factors of π to be added for them to be valid.

In the setting of triangles the angles are restricted to run between 0 and π radians as for angles greater than this we would no longer have a triangle. However, the functions are valid for any real values of the argument, x . Typically we will be interested in angles within the range $[0, 2\pi)$, but need to remember that, as shown in fig. 2.6, these functions are 2π periodic. This means that

$$\begin{aligned}\sin(x + 2\pi) &= \sin(x), \\ \cos(x + 2\pi) &= \cos(x), \\ \tan(x + 2\pi) &= \tan(x).\end{aligned}$$

We can think of these as functions from $[0, 2\pi) \rightarrow [-1, 1]$ which reduce to the familiar trig functions for angles between 0 and π .

The trig functions are clearly not one-to-one, since they are periodic, in fact even restricted to $x \in [0, 2\pi)$ there are repeated values. However, just like with the square root, it is convenient to introduce inverse functions, where we need to be careful which quadrant our angle is in to get a single value out. These inverse functions

²Usually the inverse function is denoted as f^{-1} , though we need to be careful to understand that this does not mean $1/f$.

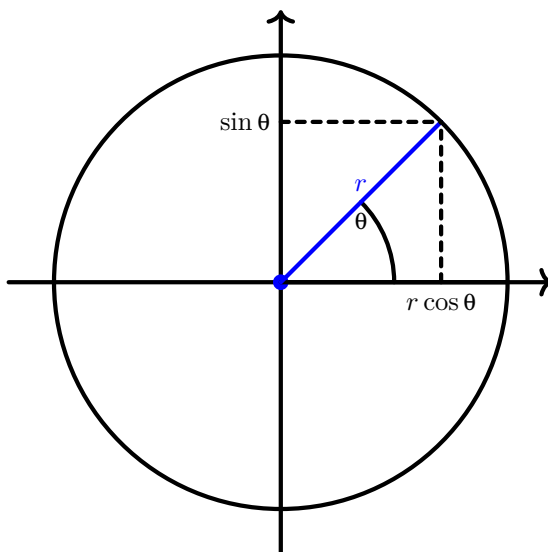


Figure 2.5: Trig functions and their relationship to a unit circle, this is true whatever the angle θ is, though we need to be careful about the signs of x and y

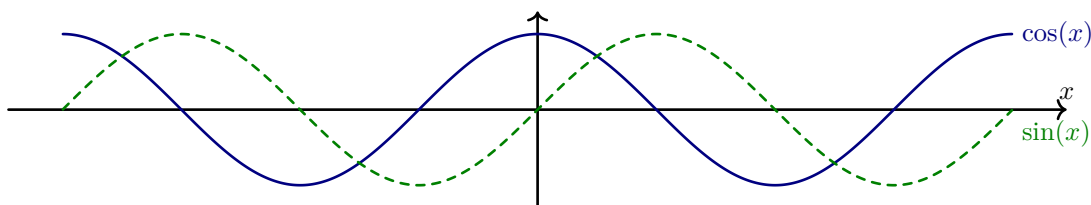


Figure 2.6: Plots of the trig functions $\sin(x)$ and $\cos(x)$ for x between -2π and 2π .

are often denoted

$$\begin{aligned}\sin^{-1}(x) &= \arcsin(x), \\ \cos^{-1}(x) &= \arccos(x), \\ \tan^{-1}(x) &= \arctan(x).\end{aligned}$$

Your calculator will have a button for each trig function, and a way to select the inverse functions, typically calculators return answers in the following ranges:

$$0 \leq \arccos(x) \leq \pi, \quad -\frac{\pi}{2} \leq \arcsin(x) \leq \frac{\pi}{2}, \quad -\frac{\pi}{2} < \arctan(x) < \frac{\pi}{2}.$$

The inverse trig functions are useful if we need to solve equations involving trig functions.

Example 2.5. Find the solutions to $4 \cos(t) = 3$ for $t \in [-8, 10]$.

The first step is to divide both sides by 4 to give us

$$\cos(t) = \frac{3}{4},$$

now we can use

$$t = \arccos\left(\frac{3}{4}\right) = 0.7227.$$

Not that this is just one of an infinite³ number of solutions to the above equation. in the range from 0 to 2π

³“Why infinitely many?” you ask. This is because $\cos(x)$ is 2π periodic so if t is a solution in the range $[0, 2\pi)$ then $t + 2\pi n$ is another solution for $n \in \mathbb{Z}$.

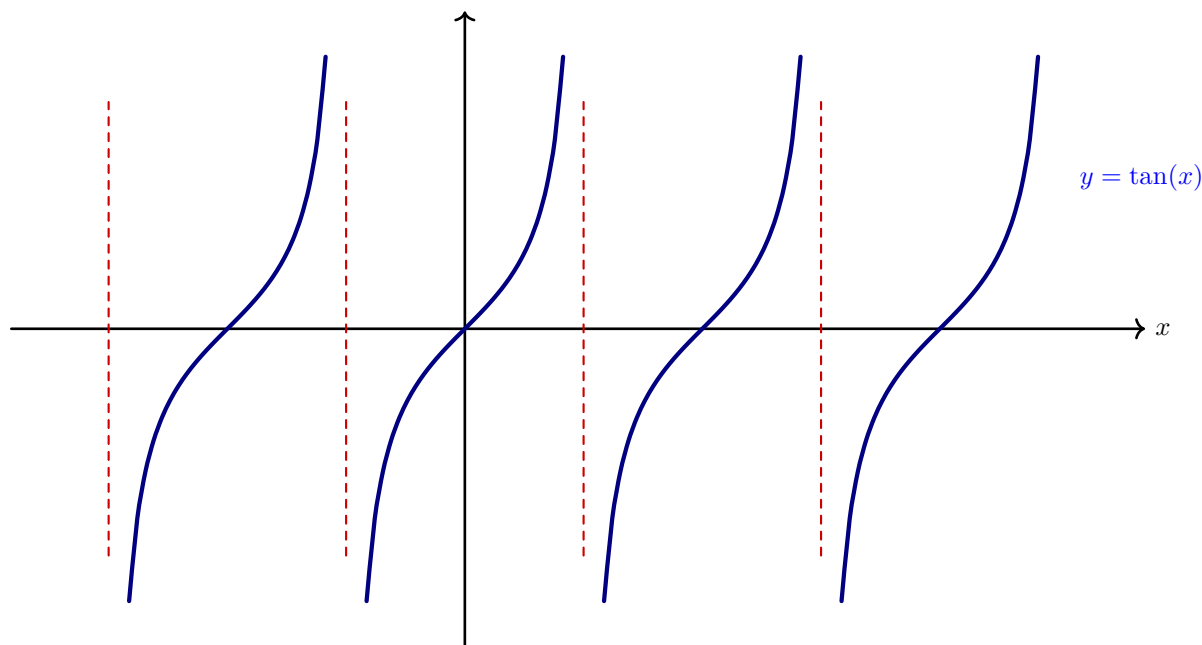


Figure 2.7: A plot of the trig function $\tan(x)$, unlike \sin and \cos above, \tan cannot be drawn without taking your pen off the page.

there are two solutions, $t = 0.7227$ and $t = 2\pi - 0.7227 = 5.5605$. We now add $2\pi n$ to both of our values, testing values of n so that the result stays within the interval :

- $n = -2$ then $t = 0.7227 - 4\pi = -11.8437$ and $t = 5.5605 - 4\pi = -7.0059$,
- $n = -1$ then $t = 0.7227 - 2\pi = -5.5605$ and $t = 5.5605 - 2\pi = -0.7227$,
- $n = 0$ then $t = 0.7227$ and $t = 5.5605$,
- $n = 1$ then $t = 0.7227 + 2\pi = 7.0059$ and $t = 5.5605 + 2\pi = 11.8437$.

Thus there are six solutions in the interval $[-8, 10]$,

$$t = -7.0059, -5.5605, -0.7227, 0.7227, 5.5605, 7.0059.$$

Not that the solutions come in positive and negative pairs.

You will have the opportunity to practice more problems like this either by going through the week one tutorial sheet or by looking at Section 1.5 in [Dawkins, 2025b].

The trig functions satisfy some nice relationships that you should try to learn, and which we quote here without any proof. Some of these are fairly easy to prove and are left as an exercise, while others are harder and can be looked up if you are curious.

The first one is that

$$\tan(x) = \frac{\sin(x)}{\cos(x)}.$$

Squares: The squares of trig functions have the nice property that

$$\sin^2 x + \cos^2 x = 1,$$

which is a consequence of Pythagoras' theorem.

Multiple angles: There are some very useful identities when we consider trig functions for the sum and difference of an angle:

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y,$$

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y.$$

For the special case of $x = y$ this leads to

$$\sin 2x = 2 \sin x \cos x,$$

$$\cos 2x = \cos^2 x - \sin^2 x = 2 \cos^2 x - 1 = 1 - 2 \sin^2 x.$$

You should try to become familiar with these identities, at least for this module, as they will be useful wherever trig functions appear.

Further trig functions: There are three more trig functions that you should be aware of. Again, these are not independent trig functions but are one over the familiar trig functions. These are:

$$\sec x = \frac{1}{\cos x},$$

$$\csc x = \frac{1}{\sin x},$$

$$\cot x = \frac{1}{\tan x}.$$

Plots of these three functions are shown in fig. 2.8.

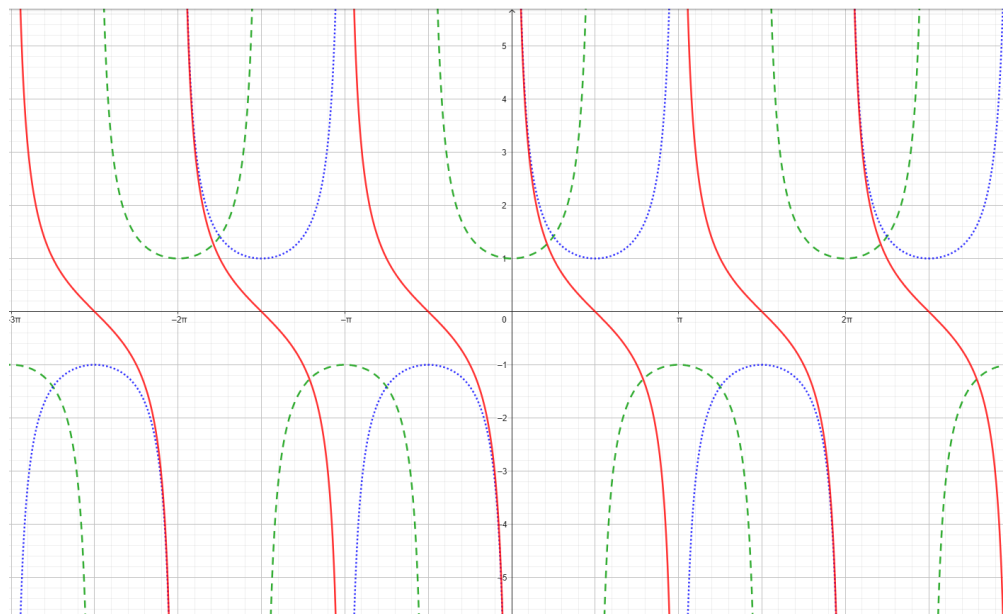


Figure 2.8: A plot of the trig functions produced using GeoGebra: $\sec(x)$ the green dashed line, $\csc(x)$ the blue dotted line, and $\cot(x)$ the solid red line. It is not possible to draw any of these functions without taking the pen off the page, we will see later that this is because these functions have vertical asymptotes where the functions head off to infinity.

2.3 Logarithms and Exponentials

Another pair of very useful functions which we need to be familiar with are the logarithm and the exponential function.

In your previous maths courses you may have come across the fact that we call powers exponents, e.g. in an expression $f(x) = x^b$ then b is called the exponent, or the index. An exponential function is then a function like

$$f(x) = b^x,$$

where a constant⁴ b is raised to the power of the variable x . In this case the number b is called the **Base** of the exponential function.

The exponential function with base 2 is shown in fig. 2.9. It turns out that by appropriate scaling of x different bases are related⁵ so we only need one standard basis.

When working with binary the basis $b = 2$ is chosen, while in scientific notation $b = 10$ is chosen. Nowadays the basis in general use is denoted e or \exp with $\exp(x)$ called the **exponential function**.

Mathematical Diversion 2.2. Warning! When talking to mathematicians there is a difference between e and \exp , the first one is a number, close to $2.71828\dots$, while \exp is a function. Often we will use e^x and $\exp(x)$ interchangeably, but occasionally it is important to know the difference. The main difference is that $\exp(x)$ is a one-to-one function, in particular when $x = 1/2$ we have that $\exp(1/2) \simeq 1.65$ while when taking $\sqrt{e} = e^{1/2}$ we need to pick either the positive or negative square root. Most likely you can just forget this distinction and work with e^x and $\exp(x)$ as if they were the same thing. However, it is worth seeing the distinction at least once.

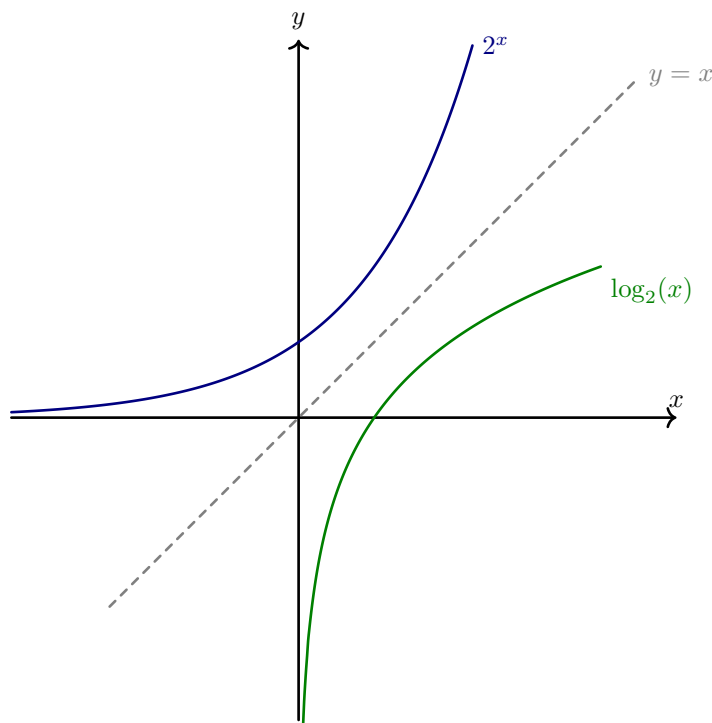


Figure 2.9: Plots of the exponential and logarithm functions, $y = 2^x$ and $y = \log_2(x)$.

The reason for picking e as the basis is that simplifies some of the algebra associated with exponential functions, it also has the advantage that the tangent to the curve has gradient 1 at $x = 0$. This last part boils down to

$$e \simeq (1 + h)^{\frac{1}{h}} \quad (2.1)$$

for h close to zero.

⁴here $b > 0$ and $b \neq 1$.

⁵To get an idea of this consider that $2^x = 4^{x/2}$ so we can always transform from base 2 into base 4. In general we need to use the logarithm function that we will meet shortly.

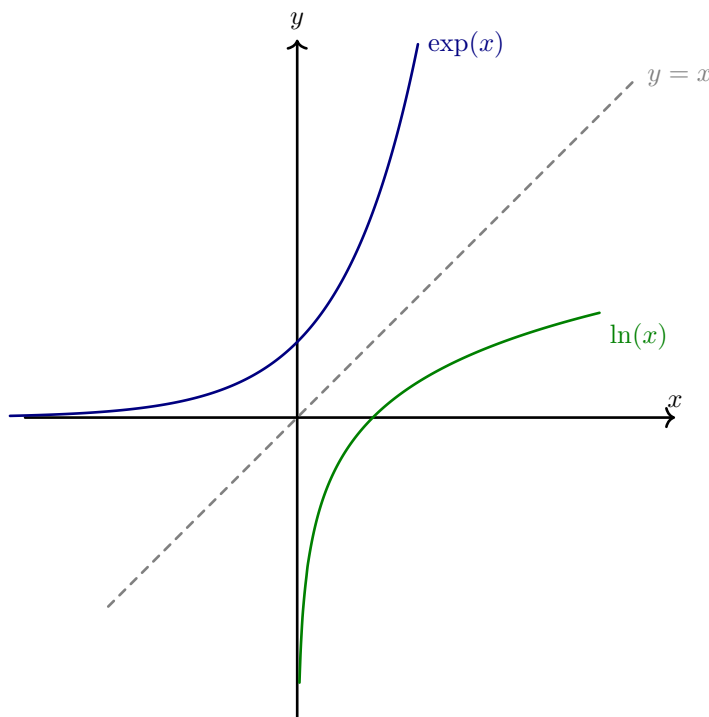


Figure 2.10: Plots of the exponential and logarithm functions, $y = \exp(x)$ and $y = \ln(x)$.

Closely related to \exp is its inverse function, called the **logarithm** function. You will see it written as either $\log(x)$ or $\ln(x)$, or rarely as $\log_e(x)$ where the base is made explicit.

As this is an inverse function it is defined as the solution to the equation $e^y = x$. For example $e^y = 4$ is solved by $y = \ln(4)$.

The graphs of both $\exp(x)$ and $\ln(x)$ are shown in fig. 2.10. Note that we are, currently, not defining \ln when x is negative⁶. You may have guessed that we could take a logarithm in a different basis since it is defined so that

$$\ln(\exp(x)) = x = \exp(\ln(x)).$$

This is why there are several slightly different ways of writing the logarithm. If you learnt about it in school you probably used $\log(x)$ to mean log with base 10. In this module if we work with a basis other than e we will make this explicit by writing $\log_b(x)$ to mean log with base b .

To be clear $\log_b(x)$ is the function defined such that

$$\log_b(b^x) = x = b^{\log_b(x)}.$$

A nice property of exponentials and logarithms is that they map addition to multiplication. What this means is that

$$\begin{aligned}\exp(a)\exp(b) &= \exp(a+b), \\ \ln(ab) &= \ln(a) + \ln(b), \\ \ln\left(\frac{a}{b}\right) &= \ln(a) - \ln(b).\end{aligned}$$

⁶If you study more mathematics you will discover that we can make sense of $\ln(x)$ when x is negative. However, it is no longer real and we need to understand complex numbers to make sense of it.

Some of the other properties of these functions are that:

$$\begin{aligned}\exp(-x) &= \frac{1}{\exp(x)}, \\ \exp(x) &> 0, \\ \exp(x) &\rightarrow \infty \quad \text{as } x \rightarrow \infty, \\ \exp(x) &\rightarrow 0 \quad \text{as } x \rightarrow -\infty, \\ \ln(x) &\rightarrow \infty \quad \text{as } x \rightarrow \infty, \\ \ln(x) &\rightarrow -\infty \quad \text{as } x \rightarrow 0^+.\end{aligned}$$

We will return to discussing limits shortly, but here the statement $\exp(x) \rightarrow \infty$ as $x \rightarrow \infty$ means that the function $\exp(x)$ becomes infinitely large as x becomes infinitely large. The terminology for this is that $\exp(x)$ **diverges** as x tends to infinity.

The final, useful, identity about logarithms is how to change the basis. The change of basis formula is

$$\log_b(x) = \frac{\log_a(x)}{\log_a(b)}. \quad (2.2)$$

As with trig functions in section 2.2, \ln and \exp are useful when solving equations. There are lots of examples in [Dawkins, 2025b] but we will go through a couple of examples here, there will be more on the tutorial sheets.

Example 2.6. Consider the equation

$$xe^{-x} - x = 0.$$

The first step in solving this is to factor out the x common to both terms. Note that we **cannot** divide by x as at this stage we do not know if it will be zero. This gives us

$$\begin{aligned}xe^{-x} - x &= 0 \\ x(e^{-x} - 1) &= 0,\end{aligned}$$

so we get two possibilities⁷: either $x = 0$ or $e^{-x} - 1 = 0$. If we had divided by x we would have missed that $x = 0$ is a solution, and would not have completely solved the problem. Consider the second case,

$$\begin{aligned}e^{-x} &= 1 \\ -x &= \ln(1) = 0\end{aligned}$$

so this other case also reduces to $x = 0$ and we only have one solution in this case.

Example 2.7. Consider the equation

$$\frac{1}{2} \ln(x^2) - \ln(x - 1) = 4.$$

If we make use of the log rules we have that $\ln(x^2)/2 = \ln(\sqrt{x^2}) = \ln x$, where we take the positive square root as currently we do not know how to make sense of logarithms with negative arguments. The equation thus becomes

$$\begin{aligned}4 &= \frac{1}{2} \ln(x^2) - \ln(x - 1) \\ &= \ln(x) - \ln(x - 1) \\ &= \ln\left(\frac{x}{x - 1}\right).\end{aligned}$$

⁷Sometimes we will get more than two possibilities, remember with trig functions their periodicity meant that we had infinitely many solutions.

Now we can exponentiate both sides and rearrange to solve for x :

$$\begin{aligned}\frac{x}{x-1} &= e^4 \\ x &= e^4(x-1) \\ x &= -e^4 + xe^4 \\ x(1-e^4) &= -e^4 \\ x &= -\frac{e^4}{1-e^4} = 1.01866.\end{aligned}$$

When we get an answer like this we need to check that it does not give a negative argument in any of the original logarithms, here we are fine since both x^2 and $x-1$ are positive.

It may seem like example 2.7 took quite a bit of working out, but with practice you will get faster at solving problems like this. In the lectures you will frequently hear me paraphrasing George Pólya and saying that Mathematics is not a spectator sport. This means that while reading these notes and attending the lectures can help you in your learning there is no substitute for actually rolling your sleeves up and solving problems. Remember that most modules *expect* you to be putting in around three hours of self study for every hour of contact time.

2.4 Hyperbolic functions

The next class of functions that it is useful to know about are the **hyperbolic** trig functions. As shown in fig. 2.11, the hyperbolic trig functions are the analogue of the standard trig functions of section 2.2 but adapted to the geometry of the hyperbola, $x^2 - y^2 = 1$ rather than the circle $x^2 + y^2 = 1$. For the standard trig functions, their argument was an angle related to how far round the unit circle we had gone. In the hyperbolic case the *angle* is now twice the shaded area shown in section 2.2.

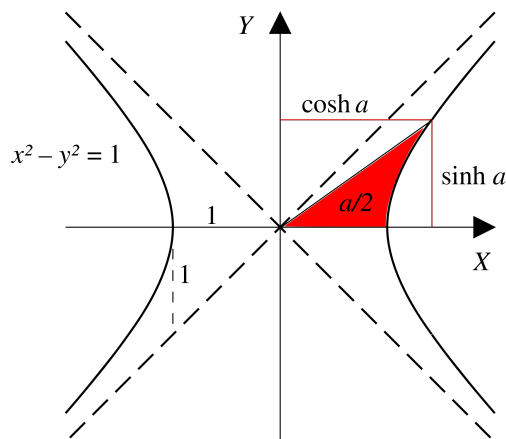


Figure 2.11: Hyperbolic trig functions are defined in terms of the area enclosed by a radial ray and the hyperbola $x^2 - y^2 = 1$. The convention is to assume that a is negative for points below the x -axis. The image is from Wikimedia commons.

There is a hyperbolic equivalent of all of the standard trig functions, and the notation is very similar just with an added h at the end:

- **hyperbolic sine** \sinh ,
- **hyperbolic cosine** \cosh ,
- **hyperbolic tangent** \tanh ,

- **inverse hyperbolic sine** $\operatorname{arcsinh}$,
- **inverse hyperbolic cosine** $\operatorname{arcCosh}$,
- **inverse hyperbolic tangent** arctan .

There are also hyperbolic versions of the other trig functions but we will not be as concerned with them here.

As in the regular trig case the hyperbolic tangent is defined in terms of the other two functions as

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}. \quad (2.3)$$

The hyperbolic trig functions are actually defined in terms of the odd and even parts of the exponential function⁸ as follows:

$$\sinh(x) = \frac{e^x - e^{-x}}{2}, \quad (2.4)$$

$$\cosh(x) = \frac{e^x + e^{-x}}{2}. \quad (2.5)$$

They have the following useful properties:

$$\begin{aligned} \sinh(-x) &= -\sinh(x), \\ \cosh(-x) &= \cosh(x), \\ \cosh^2(x) - \sinh^2(x) &= 1, \\ 1 - \tanh^2(x) &= \operatorname{sech}^2(x). \end{aligned}$$

These are very similar to the identities satisfied by the ordinary trig functions. However, the sign in front of $\sinh(x)$ and $\tanh(x)$ is negative while that in front of $\sin(x)$ and $\tan(x)$ in the equivalent formulae was positive. The rule of thumb when converting from identities for trig functions to identities for hyperbolic trig functions is to send $\cos^2(x) \rightarrow \cosh^2(x)$ and $\sin^2(x) \rightarrow -\sinh^2(x)$. There are also analogues of the addition of angle identity formulas that you can try to derive if you are interested.

Example 2.8. Consider the hyperbolic equation $\cosh(x) - 5\sinh(x) - 5 = 0$ from chapter 3.7 in [Riley et al., 2006]. The easiest way to solve this is to use the definitions of the hyperbolic functions in terms of the exponential function. This transforms the equation into

$$\frac{1}{2}(e^x + e^{-x}) - \frac{5}{2}(e^x - e^{-x}) - 5 = 0,$$

which we rearrange to give

$$0 = e^x \left(\frac{1}{2} - \frac{5}{2} \right) + e^{-x} \left(\frac{1}{2} + \frac{5}{2} \right) - 5 = -2e^x + 3e^{-x} - 5.$$

Then multiply through by $-e^x$, which is allowed since this is never zero, to get

$$2e^{2x} + 5e^x - 3 = 0.$$

Now we could either let $y = e^x$ and use the quadratic formula to solve for y or we can factorise this to get

$$(2e^x - 1)(e^x + 3) = 0,$$

so there are two solutions: $e^x = 1/2$ or $x = -\ln(2)$, and $e^x = -3$ or $x = \ln(-3)$. Remember we have not discussed how to make sense of the logarithm of a negative number so for us there is only one real solution, $x = -\ln(2)$.

⁸If you have come across complex numbers then you may know that this is only true for real x , if the argument is imaginary then relationships like this hold between the exponential function and the standard trig functions. We may see this in the advanced topics section of the module if there is enough time.

Since the hyperbolic trig functions are related to the exponential function you can probably guess that their inverses are related to the logarithm. It is worth thinking about how to show this relationship. I will show you how to do this for $\sinh(x)$ and $\tanh(x)$ but leave deriving the identity for $\cosh(x)$ as an exercise for the interested reader.

Example 2.9. Consider $y = \operatorname{arcsinh}(x)$, we can invert this to give $x = \sinh(y)$. Next if we make use of eqs. (2.4) and (2.5) we have that

$$\begin{aligned} e^y &= \cosh(y) + \sinh(y) \\ &= \sqrt{1 + \sinh^2(y)} + \sinh(y) \\ &= \sqrt{1 + x^2} + x. \end{aligned}$$

Taking \ln of both sides then gives that

$$\operatorname{arcsinh}(x) = y = \ln \left(\sqrt{1 + x^2} + x \right). \quad (2.6)$$

If you do a similar calculation for $y = \operatorname{arcCosh}(x)$ you will find that

$$\operatorname{arcCosh}(x) = \ln \left(x \pm \sqrt{x^2 - 1} \right). \quad (2.7)$$

You should think about why there is a \pm in this formula but there was not one in eq. (2.6).

Example 2.10. Consider $y = \operatorname{arctanh}(x)$, which inverts to $x = \tanh(y)$. Using the definition of $\tanh(y)$ as being $\sinh(y)/\cosh(y)$ and eqs. (2.4) and (2.5) we have that

$$x = \frac{e^y - e^{-y}}{e^y + e^{-y}}$$

which is equivalent to

$$(x + 1)e^{-y} = (1 - x)e^y.$$

which can be further rearranged to give

$$\begin{aligned} e^{2y} &= \frac{1 + x}{1 - x}, \\ \Rightarrow e^y &= \sqrt{\frac{1 + x}{1 - x}}, \\ y &= \ln \left(\sqrt{\frac{1 + x}{1 - x}} \right). \end{aligned}$$

This gives that

$$\operatorname{arctanh}(x) = \frac{1}{2} \ln \left(\frac{1 + x}{1 - x} \right). \quad (2.8)$$

You may be asking why we have spent the time deriving these identities for the hyperbolic functions, and why we discussed their relationship with the exponential function. This is because when we start to differentiate or integrate hyperbolic functions it is often more straightforward to use the expressions involving exponentials and logarithms.

2.5 Limits and asymptotics

We have now spent quite a bit of time discussing different examples of functions and their properties, as well as how to plot them. If we consider the plot of the tangent function in fig. 2.7 you may have asked at the time, what happens when the red line goes off the top of the page and reappear at the bottom? If we followed both lines we would see that they are getting closer and closer to the point $x = \pi/2$. This idea of zooming in

on a particular value $x = a$ and asking what happens to a function there is known as **taking a limit** as x approaches a and is denoted $\lim_{x \rightarrow a} f(x)$.

If the function is *well behaved* at the point a , then the limit is just the value of the function evaluated at a , $f(a)$. For example

$$\lim_{x \rightarrow 0} \cos(x) = 1 = \cos(0).$$

The value of the limit does not have to be finite, it can head off to infinity. A good example of this is the exponential function which just keeps getting larger as x increases. We denote this by

$$\lim_{x \rightarrow \infty} \exp(x) \rightarrow \infty.$$

Notice that since the value of the limit is infinite we do not write an equals sign but instead use \rightarrow to denote that the limit **diverges**⁹. There are lots of other examples of functions that diverge, and not always for large x . A nice example to have in mind is $y = 1/x^2$ which diverges as $x \rightarrow 0$.

In the case of $\tan(x)$ we can observe that $\lim_{x \rightarrow \pi/2} \tan(x)$ will take on different values depending on if x is approaching $\pi/2$ from above or below. This means that the limit **does not exist**. In section 2.6 we will discuss the consequences of this for the function in more detail. If we are taking the limit from above we often say $x \rightarrow a^+$, while the limit from below is denoted $x \rightarrow a^-$. In this case

$$\begin{aligned} \lim_{x \rightarrow \frac{\pi}{2}^+} \tan(x) &\rightarrow \infty, \\ \lim_{x \rightarrow \frac{\pi}{2}^-} \tan(x) &\rightarrow -\infty, \end{aligned}$$

so $\tan(x)$ diverges in different directions depending on how we approach $\pi/2$.

Mathematical Diversion 2.3. *Putting our mathematician's hats on for a brief moment, we would say that limit of a function $f(x)$ as x approaches a is L and write it as*

$$\lim_{x \rightarrow a} f(x) = L,$$

as long as we can make $f(x)$ as close to L as we want for all x sufficiently close to a on both sides, without taking $x = a$.

If this was a maths module we would be even more precise and give what is called an ϵ, δ definition. Fortunately for all of us this is not a maths module and we can focus on a more practical/ working definition rather than getting bogged down in the abstract details. If you are interested to know how to do this in detail, the section “The Definition of a Limit” in [Dawkins, 2025b] is a good place to look.

Remember that $x \rightarrow a$ does not mean that x becomes equal to a , it just means that x is getting close to a . It is also possible that $f(x)$ may never equal L , but also just get close to it. This is especially true when we take the limit near a point that the function is not defined at, e.g. $x \rightarrow \pi/2$ for $\tan(x)$.

Curve sketching is a useful way to understand the main features of a function and can help us to know if the limit exists and if we have the same limit when approaching from different directions.

Example 2.11. *Consider the function*

$$f(x) = \frac{1}{x-2}, \tag{2.9}$$

defined for $x \neq 2$.

*Notices that for $x \rightarrow \pm\infty$ $f(x) \rightarrow 0$, in fig. 2.12 we see that the plot approaches the horizontal axis **asymptotically**¹⁰. We can also observe that for $a \neq 2$ $f(x) \rightarrow f(a)$ for $x \rightarrow a$. The only point where we need to be*

⁹This is just the term that mathematicians use for a function that heads off to infinity.

¹⁰The definition of asymptotically is “approaching a given value or condition, as a variable or an expression containing a variable approaches a limit, usually infinity” [Collins]

careful is when x tends to 2 looking at fig. 2.12 we see that if we approach 2 from above that $f(x) \rightarrow \infty$ while if we approach 2 from below $f(x) \rightarrow -\infty$.

Note that the fact that $f(x)$ is not defined at $x = 2$ has no bearing on the limiting behaviour.

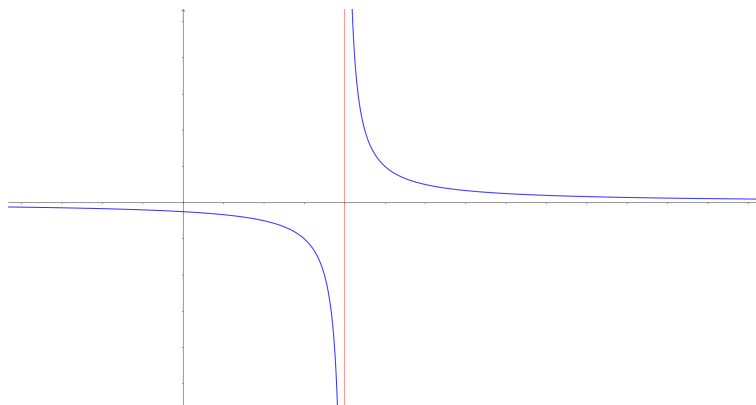


Figure 2.12: A plot of the function from eq. (2.9) produced using GeoGebra. The blue line is the function and the red line shows where the function diverges and has a vertical asymptote.

The general rule when curve sketching is to set $y = f(x)$ for the function of interest $f(x)$, and then to investigate the following:

- The **intercepts** of $f(x)$, these are the points where the function meets or crosses the coordinate axes, e.g. $f(x = 0)$ and $f(x) = 0$, as these position the graph relative to the coordinate axes.
- What happens to y as $x \rightarrow \pm\infty$?
- Which values of x make $y \rightarrow \pm\infty$?
- Does the function have any **symmetry**? For example does $f(-x) = f(x)$ (f is even), or $f(-x) = -f(x)$ (f is odd).
- Is the function periodic?

Note that a function can have several of these interesting features. e.g. the trig functions are all periodic, while $\sin(x)$ and $\tan(x)$ are odd functions and $\cos(x)$ is an even function.

Mathematical Diversion 2.4. An asymptote to a curve is a straight line that becomes close to the curve as either x or y tend to $\pm\infty$. In our above example the curve of eq. (2.9) has two asymptotes, the horizontal x -axis for $x \rightarrow \pm\infty$ and the vertical line $x = 2$ for $x \rightarrow 2$.

When sketching a curve it is often useful to start by drawing the asymptotes in as they help you to know how the function will behave in certain regions of the plot. In some books you may see $f(x) \asymp mx + c$ as $x \rightarrow \infty$ to signify that the straight line $y = mx + c$ is an asymptote to the graph of $f(x)$ as x tends to infinity.

Exercise 2.12. Apply the above rules to produce a sketch of the function

$$f(x) = \frac{x+2}{x-1}.$$

Example 2.13. Consider the rational function

$$f(x) = \frac{x^2 + 4x - 12}{x^2 - 2x}$$

and compute its limit as $x \rightarrow 2$.

Looking at the function you may think that it will diverge as $x \rightarrow 2$ since the denominator vanishes there. However the first step is always to look at if there are any common factors between the numerator and denominator. Factorising both we have

$$\begin{aligned} f(x) &= \frac{x^2 + 4x - 12}{x^2 - 2x} \\ &= \frac{(x+6)(x-2)}{x(x-2)} \\ &= \frac{x+6}{x}. \end{aligned}$$

So the denominator does not vanish at $x = 2$. We can now evaluate the limit to be

$$\lim_{x \rightarrow 2} f(x) = \lim_{x \rightarrow 2} \frac{x+6}{x} = \frac{8}{2} = 4.$$

The limit is the same regardless of which direction we approach 2 from.

Exercise 2.14. Estimate the limit from the above example by constructing a table of values of x and $f(x)$ for x approaching 2 from both above and below.

As a warning, this approach of tabulating the values does not work if the function is oscillating. For example, if you try to evaluate the limit as $x \rightarrow 0$ of $\cos(\pi/x)$ it can look like it is tending to a constant if we are not careful with the values that we pick. However, if we plot the function we see that it is highly oscillatory around zero.

Example 2.15. Consider the rational function

$$f(x) = \frac{x^3 + x^2 - 5x - 2}{2x^3 - 7x^2 + 4x + 4}.$$

It has three interesting limits, $x \rightarrow 0, \infty, 2$. We can evaluate the first two of these here, but need to leave the third, $x \rightarrow 2$ until we have learnt about L'Hôpital's rule in chapter 8. We will do $x \rightarrow 0$ first. As usual we check that the numerator and denominator both make sense as $x \rightarrow 0$ and then can evaluate the limit

$$\begin{aligned} \lim_{x \rightarrow 0} f(x) &= \lim_{x \rightarrow 0} \frac{x^3 + x^2 - 5x - 2}{2x^3 - 7x^2 + 4x + 4} \\ &= \frac{-2}{4} \\ &= -\frac{1}{2}. \end{aligned}$$

Next we do the $x \rightarrow \infty$ limit. Before we do this we need to multiply $f(x)$ by x^{-3}/x^{-3} so that it becomes

$$\begin{aligned} \lim_{x \rightarrow \infty} f(x) &= \lim_{x \rightarrow \infty} \left(\frac{x^{-3}}{x^{-3}} \frac{x^3 + x^2 - 5x - 2}{2x^3 - 7x^2 + 4x + 4} \right) \\ &= \lim_{x \rightarrow \infty} \left(\frac{1 + x^{-1} - 5x^{-2} - 2x^{-3}}{2 - 7x^{-1} + 4x^{-2} + 4x^{-3}} \right) \\ &= \frac{1}{2}. \end{aligned}$$

Remember that when evaluating a limit we want to look for as many cancellations as possible to simplify the calculation. This can be looking for common factors between the numerator and denominator, but can also mean fully expanding out all terms as there may be cancellations hidden by the way the function has been written.

Often it is useful to plot a function when we are estimating the limit, while this is not compulsory it can be very helpful, particularly if as in the case of $\tan(x)$, the value of the limit depends on the direction of approach.

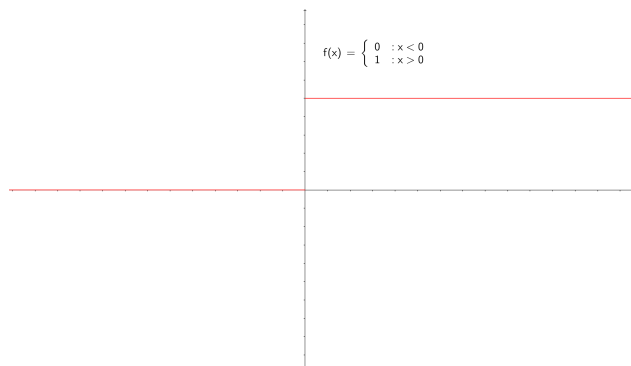


Figure 2.13: A plot of a step function from eq. (2.10) produced using GeoGebra.

Example 2.16. Consider the step function

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0, \end{cases} \quad (2.10)$$

shown in fig. 2.13. By looking at the graph we see that the limit of $f(x)$ as $x \rightarrow 0$ will depend on the direction of approach. If we start from below zero it is clear that $f(x) \rightarrow 0$, while starting above zero it is clear that $f(x) \rightarrow 1$. In contrast to the case of $\tan(x)$ the function is not diverging in the limit, but we still end up with a **One Sided Limit**¹¹

Mathematical Diversion 2.5. Note that it can be dangerous to use ∞ in calculations as if it were a number. This is because there are certain limits and other expressions that we may see which do not make sense:

$$\infty^0, \quad \infty - \infty, \quad \frac{0}{0}, \quad 0^0, \quad \frac{\infty}{\infty}, \quad \frac{\infty}{0}, \quad \frac{1}{0}.$$

Just because we see one of these expressions coming out a limit does not necessarily mean that the limit does not exist. It can instead mean that we need to be very careful how we evaluate the limit. In chapter 8 we will discuss L'Hôpital's rule which is a techniques for evaluating limits that initially look like they do not make sense.

There are a few limits that we need to know that we will not prove here, time permitting we will prove these in chapter 8, but you can also look at the “Proof of Trig Limits” section of Dawkins [2025b] to see one approach to proving them. These limits are

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1, \quad (2.11)$$

$$\lim_{x \rightarrow 0} \frac{\cos x - 1}{x} = 0, \quad (2.12)$$

$$\lim_{h \rightarrow 0} \frac{e^h - 1}{h} = 1, \quad (2.13)$$

$$\lim_{h \rightarrow 0} \frac{\ln(1 + h)}{h} = 1. \quad (2.14)$$

Note that eq. (2.13) is equivalent to eq. (2.1) which we discussed above as one of the definitions of the exponential function.

2.6 Continuity and differentiability

Now that we have some examples of functions and understand how to take limits, we can define two properties of a function which will be very important later in the course: **continuity** and **differentiability**.

¹¹This just means that the function is well behaved in the limit as long as we only look at one side.

A function $f(x)$ is said to be continuous at a point $x = a$ if

$$\lim_{x \rightarrow a} f(x) = f(a). \quad (2.15)$$

If X is the domain of $f(x)$, we say that f is continuous on X if it is continuous at each point in X , often we would just say that f is a continuous function. As a rule of thumb, we can say that a function is continuous if its graph can be drawn from start to finish without taking your pen off the paper. Functions which are not continuous will have jumps or divergences at the point that fails to be continuous.

If we have a rational function then we can find where it is not continuous, called being **discontinuous**, by finding the roots of the denominator.

Mathematical Diversion 2.6. *If we are being careful we need to give three parts to the definition of continuity:*

- 1) *The limit of $f(x)$ as $x \rightarrow a$ exists.*
- 2) *The value of $f(x)$ is defined at $x = a$. i.e. $f(a)$ is defined and is finite.*
- 3) *The limit of $f(x)$ as $x \rightarrow a$ agrees with the value of $f(a)$.*

This is another place where as this is not a course for mathematicians we can combine all three of these into the one statement in eq. (2.15). This is another example of being able to use a working definition and not needing to get sidetracked by all of the technical details.

Example 2.17. *We have already met several examples of continuous and discontinuous functions:*

- *The trig functions $\cos(x)$ and $\sin(x)$ are continuous on all of \mathbb{R} .*
- *The exponential function $\exp(x)$ is continuous on all of \mathbb{R} .*
- *The natural logarithm is continuous on the positive real numbers \mathbb{R} as, currently we have not defined it for negative x , and it diverges in the limit $x \rightarrow 0$.*
- *The tangent function $\tan(x)$ are not continuous on \mathbb{R} due to its divergences at $\pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \dots$. However, it is continuous on its domain¹² $(-\frac{\pi}{2}, \frac{\pi}{2})$*
- *The step function of eq. (2.10) is not continuous at $x = 0$.*
- *The function in eq. (2.9) has a discontinuity at $x = 2$*

Exercise 2.18. *Use the definition of continuity to determine if the function*

$$g(x) = \frac{4x + 10}{x^2 - 2x - 15} \quad (2.16)$$

is continuous and if not find the points where it has discontinuities.

Related to continuity is differentiability. A function $f(x)$ is differentiable at a point a if the limit

$$\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}$$

exists. We call this limit the derivative of the function at the point a ,

$$f'(a) = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}, \quad (2.17)$$

the notation $\frac{df}{dx}(a)$ is sometimes used instead of $f'(a)$. A function is called continuously differentiable if its derivative is also continuous.

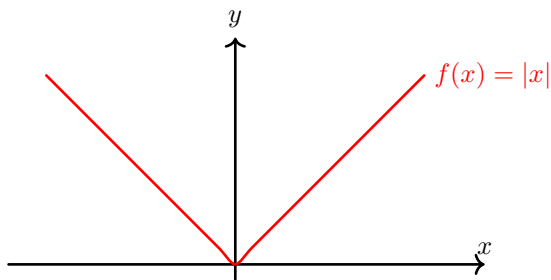


Figure 2.14: The absolute value function $f(x) = |x|$ is continuous at every point, but it is not differentiable at the point $x = 0$.

It is important to note that differentiability at a implies continuity at a , but continuity does not imply differentiability. For example the absolute value function shown in fig. 2.14 is continuous everywhere, but is only differentiable everywhere except at $x = 0$.

Example 2.19. Consider the modulus or absolute value function shown in fig. 2.14. We have said that this is a continuous function which is not differentiable at $x = 0$, but how do we show this? We do it by checking all of the limits.

Checking continuity at $x = 0$ is left as an exercise. For differentiability we need to check the limits as $x \rightarrow 0^+$ and $x \rightarrow 0^-$.

In the first one we have:

$$\lim_{x \rightarrow 0^+} \frac{f(x) - f(0)}{x} = \lim_{x \rightarrow 0^+} \frac{x - 0}{x} \lim_{x \rightarrow 0^+} 1 = 1,$$

while approaching from the other direction gives:

$$\lim_{x \rightarrow 0^-} \frac{f(x) - f(0)}{x} = \lim_{x \rightarrow 0^-} \frac{(-x) - 0}{x} \lim_{x \rightarrow 0^-} -1 = -1.$$

These do not agree, so the limit does not exist and the modulus function is not differentiable.

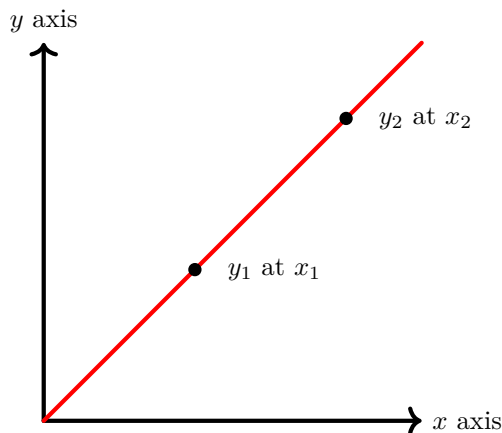


Figure 2.15: A plot of the straight line $y = x$ and example of $y = mx + c$ with gradient $m = 1$ and y -intercept $c = 0$, with two points on the line marked.

The fraction in eq. (2.17) may look familiar to you. If we had a straight line $y = f(x) = mx + c$ where m is the gradient of the straight line and c is the y -intercept, then calculating this fraction gives

$$\frac{f(x) - f(a)}{x - a} = \frac{mx + c - (ma + c)}{x - a} = \frac{m(x - a)}{x - a} = m,$$

¹²Strictly speaking this is the principle domain of $\tan(x)$ its domain is really $\{x \in \mathbb{R} | x \neq \pi/2 + n\pi, n \in \mathbb{Z}\}$

which is the gradient of the curve. For a function that is not a straight line, this procedure gives the gradient of the straight line between x and a . In the limit that $x \rightarrow a$ this fraction becomes the gradient of the **tangent** line¹³ to the curve at a . See fig. 2.16 for the example of the tangent to parabola $y = x^2$.

Mathematical Diversion 2.7. *The fraction*

$$\frac{f(x) - f(a)}{x - a}$$

is sometimes referred to as the Newton quotient of the function $f(x)$ at the point a . This is after Isaac Newton because when calculating a derivative we are calculating this quotient for smaller and smaller differences $x - a$.

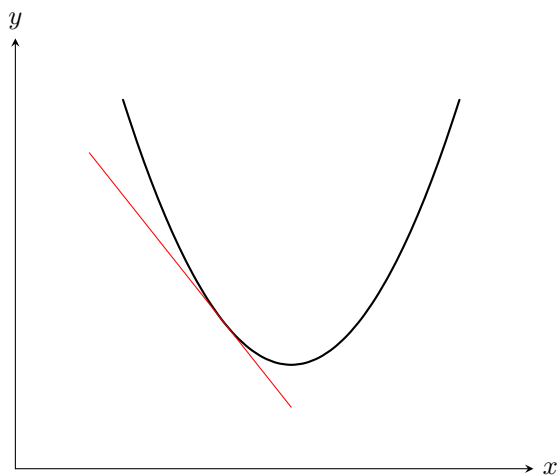


Figure 2.16: A plot of a parabola in black and its tangent in red

¹³A tangent is a line that touches a curve at one point

Chapter 3

Differentiation

You take a function of x and you
call it y .
Take any x_0 that you care to try.
You make a little change and call
it Δx .
The corresponding change in y is
what you find next.

*The Derivative Song by Tom
Lehrer*

3.1 Differentiation from first principles

In the previous chapter, we encountered the derivative of a function at a point in eq. (2.17) and its interpretation as the tangent to a curve at the point a . With a small amount of rewriting, setting $x = a + h$ for x “near” a , this becomes

$$\begin{aligned} f'(a) &= \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \\ &= \lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{a + h - a} \\ &= \lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{h}. \end{aligned}$$

This is still an expression for the tangent to a curve at a specific point $x = a$. However, if we are interested in the gradient at an arbitrary point x , then we can rewrite it as

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{h}. \quad (3.1)$$

The formula in eq. (3.1) is the definition of the **derivative** of the function $f(x)$. Not that only functions where the derivative exists for all points are called **differentiable**. There is a range of notation¹ and terminology used to denote derivative. The notation used above, $f'(x)$ is known as the **Lagrange** or **Euler** notation².

¹We will only use the two most common notations in this course. However, another notation that you may see in books is $f_x(x)$, where the subscript shows what we are differentiating with respect to. This notation shows up a lot if we have functions of more than one variable where we need to make it clear which variable we are differentiating with respect to. In the assessed part of this course we will only care about functions of one variable so do not need this notation.

²In an example of Sigler’s law of eponymy, this is most commonly called Lagrange’s notation even though it was first used by Euler. It is also quite close to the original notation that Newton used when he discovered calculus and referred to it as the **method of fluxions**.

The other very common notation is due to **Leibniz** where we write

$$\frac{df}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}, \quad (3.2)$$

this notation is particularly useful when we learn about integration as in certain contexts we can treat the derivative like a fraction.

Mathematical Diversion 3.1. In eq. (3.2) it looks like the right hand side is a fraction. This is not true, other than in certain very specific circumstances, we will not see df or dx appearing on their own. In Newton's approach the Newton quotient is sometimes written as

$$\frac{\Delta f}{\Delta x} = \frac{f(x) - f(a)}{x - a},$$

which does make sense as a fraction. The limit where this becomes the derivative is when the change in x , $\Delta x = x - a$, goes to zero. In this limit if we really had a fraction it would look like $0/0$, which is one of the nonsense expressions that we mentioned earlier. The power of calculus is that it enables us to make sense of this limit, but what we lose is the ability to treat it as a fraction. When we discuss integration and differential equations later on in the module we will return to this idea.

If we use eq. (3.1) to calculate the derivative of a function this is called **differentiation by first principles**. As you might expect, when the function $f(x)$ becomes more complicated calculating the derivative in this way becomes more complicated as well. Fortunately, there are certain standard rules and techniques that we can learn to simplify matters. With a little work all of these can be proved from the definition of the derivative, some of these proofs will be given here but others are left as an exercise to the interested reader.

As a warm up we will use eq. (3.1) to calculate the derivative of a straight line.

Example 3.1. Consider $f(x) = mx + c$ and calculate the derivative from first principles:

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{m(x+h) + c - (mx + c)}{h} \\ &= \lim_{h \rightarrow 0} \frac{mx + mh + c - mx - c}{h} \\ &= \lim_{h \rightarrow 0} m \frac{h}{h} \\ &= \lim_{h \rightarrow 0} m = m. \end{aligned}$$

So derivative of a straight line is a constant, the gradient of the line. We already knew this, but it is a good consistency check to ensure that our definition of the derivative is working as expected.

Example 3.2. Consider the function $f(x) = 4x^2 - 6x + 2$, using eq. (3.1) we calculate its derivative as follows:

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{4(x+h)^2 - 6(x+h) + 2 - (4x^2 - 6x + 2)}{h} \\ &= \lim_{h \rightarrow 0} \frac{4(x^2 + 2xh + h^2) - 6x - 6h + 2 - 4x^2 + 6x - 2}{h} \\ &= \lim_{h \rightarrow 0} \frac{8xh + 4h^2}{h} \\ &= \lim_{h \rightarrow 0} (8x + 4h) = 8x. \end{aligned}$$

Notice that since the curve is no longer a straight line the derivative, and thus the gradient of the tangent to the curve, depends where the point is along the curve.

Remember that if the limit in eq. (3.1) does not exist at a particular value of x , then the derivative does not exist. In other words, the definition of the derivative only makes sense for functions which satisfy the condition of differentiability.

Example 3.3. Consider the function

$$g(x) = \frac{1}{x+1},$$

and calculate its derivative. Note that this function has a discontinuity at $x = -1$ so it will not be differentiable at that point.

Calculating $g'(x)$ is good practice as we need to be careful when we have fractions.

$$\begin{aligned} g'(x) &= \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\frac{1}{x+h+1} - \frac{1}{x+1} \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\frac{x+1}{(x+h+1)(x+1)} - \frac{x+h+1}{(x+1)(x+h+1)} \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\frac{x+1-x-h-1}{(x+h+1)(x+1)} \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\frac{-h}{(x+h+1)(x+1)} \right) \\ &= \lim_{h \rightarrow 0} \frac{-1}{(x+h+1)(x+1)} \\ &= -\frac{1}{(x+1)^2}. \end{aligned}$$

We can also calculate the derivatives of some of the special functions from first principles.

Example 3.4. Consider $f(x) = \sin(x)$, this is a continuous function so we can hope that the derivative exists. Note that we can use the addition formula for $\sin(x)$ to expand $\sin(x+h)$ as

$$\sin(x+h) = \sin(x)\cos(h) + \sin(h)\cos(x).$$

Thus eq. (3.1) becomes

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\sin(x)\cos(h) + \sin(h)\cos(x) - \sin(x)}{h} \\ &= \lim_{h \rightarrow 0} \left(\sin(x) \frac{\cos(h) - 1}{h} + \cos(x) \frac{\sin(h)}{h} \right) \\ &= \sin(x) \lim_{h \rightarrow 0} \frac{\cos(h) - 1}{h} + \cos(x) \lim_{h \rightarrow 0} \frac{\sin(h)}{h} \\ &= \cos(x), \end{aligned}$$

where we have used the trig limits from eqs. (2.11) and (2.12)

Exercise 3.5. Show that the derivative of $f(x) = \cos(x)$ is $f'(x) = -\sin(x)$ using differentiation from first principles.

Example 3.6. Consider $f(x) = e^x$, its derivative is

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{e^{x+h} - e^x}{h} \\ &= \lim_{h \rightarrow 0} \frac{e^x e^h - 1}{h} \\ &= e^x \lim_{h \rightarrow 0} \frac{e^h - 1}{h} \\ &= e^x. \end{aligned}$$

Where we use eq. (2.13) to evaluate the limit in the final line. This result that the derivative of e^x is equal to e^x , is sometimes taken to be a definition of the exponential function.

The other standard derivative that you need to know is the derivative of the natural logarithm.

Example 3.7. Consider $f(x) = \ln(x)$, we calculate its derivative as

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\ln(x+h) - \ln(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \ln \left(\frac{x+h}{x} \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \ln \left(1 + \frac{h}{x} \right), \end{aligned}$$

now we can let $\epsilon = h/x$ which is tending to zero as h tends to zero. Thus the derivative becomes

$$\begin{aligned} f'(x) &= \lim_{\epsilon \rightarrow 0} \frac{1}{x\epsilon} \ln(1 + \epsilon) \\ &= \frac{1}{x} \lim_{\epsilon \rightarrow 0} \frac{\ln(1 + \epsilon)}{\epsilon} \\ &= \frac{1}{x}, \end{aligned}$$

where we have made use of eq. (2.14)

Using the first principles definition to calculate derivatives is hard work and involves careful manipulation of limits. You will be please to know that we will only use it in certain relatively simple cases. For more complicated expressions we can use a range of other techniques and formulas. Eventually you will internalise some of the standard derivative expressions or use a formula sheet like that in chapter 13.

3.2 Differentiation techniques

Properties of the derivative

In this section we will see the various formulae and rules in both the Lagrange and Leibniz notation so you should be familiar with both and use the notation that you feel most comfortable with.

The first thing that we need is to know is how to differentiate the sum of two functions and the product of a

function with a number. The proof of these results will be given in chapter 12. These formulae are:

$$\frac{d}{dx} (f(x) + g(x)) = \frac{df(x)}{dx} + \frac{dg(x)}{dx}, \quad (3.3)$$

$$\frac{d}{dx} (f(x) - g(x)) = \frac{df(x)}{dx} - \frac{dg(x)}{dx}, \quad (3.4)$$

$$\frac{d}{dx} (cf(x)) = c \frac{df(x)}{dx}, \quad (3.5)$$

where c is any number.

In the Lagrange/Euler notation these formulae are:

$$(f(x) + g(x))' = f'(x) + g'(x), \quad (3.6)$$

$$(f(x) - g(x))' = f'(x) - g'(x), \quad (3.7)$$

$$(cf(x))' = cf'(x). \quad (3.8)$$

These are useful formulae that we will frequently use in calculations as it enables us to split up the functions that we are differentiating into smaller tractable parts. The other shortcuts that we have is that the derivative of a constant is zero,

$$\frac{dc}{dx} = 0, \quad (3.9)$$

and that the derivative of a **monomial** is

$$\frac{d}{dx} (x^n) = nx^{n-1}. \quad (3.10)$$

Sometimes the second formula is referred to as the **power rule**, and is one of the most important rules that you can learn as it enables you to differentiate any polynomial.

The formula in eq. (3.9), which says that the derivative of a constant is zero makes sense since the derivative measures the rate of change of a function. A constant is, by definition, not changing so its rate of change is zero.

We can now use these rules to calculate some example derivatives.

Example 3.8. Consider the function

$$f(x) = 15x^{20} - 3x^5 + 2x + 4.$$

We calculate its derivative as follows:

$$\begin{aligned} \frac{df}{dx} &= \frac{d}{dx} (15x^{20} - 3x^5 + 2x + 4) \\ &= 15 \frac{d}{dx} (x^{20}) - 3 \frac{d}{dx} (x^5) + 2 \frac{d}{dx} (x) + \frac{d}{dx} (4) \\ &= 15 \times 20x^{19} - 3 \times 5x^4 + 2 + 0 \\ &= 300x^{19} - 15x^4 + 2. \end{aligned}$$

Example 3.9. Consider the function

$$g(x) = \frac{6}{x^2} - 4x^2 + 2x.$$

Its derivative is calculated as follows

$$\begin{aligned} \frac{dg}{dx} &= \frac{d}{dx} \left(\frac{6}{x^2} - 4x^2 + 2x \right) \\ &= 6 \frac{d}{dx} (x^{-2}) - 4 \frac{d}{dx} (x^2) + 2 \frac{d}{dx} (x) \\ &= 6 \times (-2)x^{-3} - 4 \times 2x + 2 \\ &= -\frac{12}{x^3} - 8x + 2. \end{aligned}$$

Exercise 3.10. Calculate the derivative of

$$y = 8x^2 + 2x - \frac{1}{x}.$$

Product rule

So far we have not discussed differentiating the product of two functions, unless you count $x^{a+b} = x^a x^b$. A very useful formula is the **Leibniz** or **product rule** which tells us how to differentiate the product of two functions. The product rule is

$$\frac{d}{dx}(f(x)g(x)) = \frac{df}{dx}g(x) + f(x)\frac{dg}{dx}. \quad (3.11)$$

You may think that it is disappointing that the derivative of a product is not just the product of the derivatives. However, you will come to appreciate the product rule as you make use of it. It will become particularly useful once we have discussed more about how to differentiate trig functions and exponentials.

Example 3.11. For the product of two functions $\sqrt{x^3}\sin(x)$ we find the derivative as follows. Let $f(x) = \sqrt{x^3}$ and $g(x) = \sin(x)$ and calculate the individual derivatives to be

$$f'(x) = (x^{\frac{3}{2}})' = \frac{3}{2}x^{\frac{3}{2}-1} = \frac{3}{2}\sqrt{x}, \quad g'(x) = (\sin(x))' = \cos(x).$$

Then we use product rule to calculate

$$\begin{aligned} \left(\sqrt{x^3}\sin(x)\right)' &= (f(x)g(x))' \\ &= f'(x)g(x) + f(x)g'(x) \\ &= \frac{3}{2}\sqrt{x}\sin(x) + \sqrt{x^3}\cos(x). \end{aligned}$$

Example 3.12. We can use the product rule to calculate the derivative of $f(x) = \sin^2(x)$. In this case the two functions are the same so we can evaluate the derivative as follows

$$\begin{aligned} f'(x) &= (\sin^2(x))' \\ &= 2\sin(x)(\sin(x))' \\ &= 2\sin(x)\cos(x) \\ &= \cos(2x), \end{aligned}$$

where we have used one of the trig double angle identities in the last line.

Exercise 3.13. Use the product rule to calculate the derivative of

$$f(x)g(x) = e^x \sin(x).$$

Note that the product rule applies if we have the product of more than two functions, we just have to iterate it for each pair of functions.

Quotient rule

If instead of a product of functions we have the ratio of two functions then there is a rule for that, the **quotient** rule³. In Leibniz notation the quotient rule is

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{\frac{df}{dx}g(x) - f(x)\frac{dg}{dx}}{g^2(x)}. \quad (3.12)$$

³You may be looking at these two formula and thinking that we could just use the product rule with $f(x)$ and $(g(x))^{-1}$. If you do this you will get an answer that looks just like the quotient rule. There are some technical differences between this approach and the quotient rule we give here, but as we are not mathematicians here we do not need to worry about them.

Remember that in most circumstances you can use either the product rule or the quotient rule, it depends how you want to approach the problem.

Example 3.14. Consider the function

$$F(x) = \frac{\sin(x)}{e^x}.$$

Note that we could use the product rule on $e^{-x} \sin(x)$ but will instead use the quotient rule here. Split $F(x)$ into the two functions $f(x) = \sin(x)$ and $g(x) = e^x$, applying the quotient rule eq. (3.12) gives

$$\begin{aligned} \frac{d}{dx} \left(\frac{f(x)}{g(x)} \right) &= \frac{\frac{df}{dx}g(x) - f(x)\frac{dg}{dx}}{g^2(x)} \\ &= \frac{e^x \cos(x) - \sin(x)e^x}{e^{2x}} \\ &= e^{-x} (\cos(x) - \sin(x)). \end{aligned}$$

Example 3.15. Consider

$$F(x) = \frac{1}{\sin^2(x)},$$

and split it into two functions $f(x) = 1$ and $g(x) = \sin^2(x)$, which have derivatives

$$\frac{df}{dx} = 0, \quad \frac{dg}{dx} = 2 \sin(x) \cos(x) = \sin(2x).$$

The quotient rule thus gives that

$$\begin{aligned} \frac{d}{dx} \left(\frac{f(x)}{g(x)} \right) &= \frac{\frac{df}{dx}g(x) - f(x)\frac{dg}{dx}}{g^2(x)} \\ &= \frac{0 - 2 \sin(x) \cos(x)}{\sin^4(x)} \\ &= -2 \frac{\cot(x)}{\sin^2(x)} \\ &= -2 \cot(x) \csc^2(x). \end{aligned}$$

Derivatives of trig functions

Armed with the quotient rule we can now give the derivatives of the six trig functions

$$\frac{d}{dx} \sin(ax) = a \cos(ax), \tag{3.13}$$

$$\frac{d}{dx} \cos(ax) = -a \sin(ax), \tag{3.14}$$

$$\frac{d}{dx} \tan(ax) = a \sec^2(ax), \tag{3.15}$$

$$\frac{d}{dx} \cot(ax) = -a \csc^2(ax), \tag{3.16}$$

$$\frac{d}{dx} \sec(ax) = a \sec(ax) \tan(ax), \tag{3.17}$$

$$\frac{d}{dx} \csc(ax) = -a \csc(ax) \cot(ax). \tag{3.18}$$

We have seen how to prove two of these and can now calculate the derivative of $\tan(x)$ the other three will be left as exercises.

Example 3.16. Consider $f(x) = \tan(x)$, which can be expressed as

$$f(x) = \tan(x) = \frac{\sin(x)}{\cos(x)}.$$

Using the quotient rule we have that

$$\begin{aligned} \frac{d}{dx} \tan(x) &= \frac{d}{dx} \left(\frac{\sin(x)}{\cos(x)} \right) \\ &= \frac{1}{(\cos(x))^2} \left(\cos(x) \frac{d}{dx} \sin(x) - \sin(x) \frac{d}{dx} \cos(x) \right) \\ &= \frac{1}{\cos^2(x)} (\cos(x) \cos(x) - \sin(x)(-\sin(x))) \\ &= \frac{1}{\cos^2(x)} (\cos^2(x) + \sin^2(x)) \\ &= \frac{1}{\cos^2(x)} \\ &= \sec^2(x). \end{aligned}$$

Where we have used the identity that $\cos^2(x) + \sin^2(x) = 1$ and the definition of $\sec(x) = 1/\cos(x)$.

Exercise 3.17. Compute the derivatives of $\cot(x)$, $\sec(x)$, $\csc(x)$.

Chain rule

If we have a function of a function, e.g. $f(x) = \exp(x^2 + x)$ or $g(x) = \cos(x + c)$, none of the rules that we have given so far will work. We could go back to first principles, which is exactly what we did when calculating the derivative of $1/(x + 1)$, but this would mean that we had to do lots of long and tricky calculations. Fortunately this is not necessary.

Consider the function

$$f(x) = \sqrt{4x + 2}, \tag{3.19}$$

we can write this as the **composition** of two functions if we think of $g(x) = \sqrt{x}$ and $h(x) = 4x + 2$,

$$f(x) = (g \circ h)(x) = g(h(x)) = g(4x + 2) = \sqrt{4x + 2}.$$

The **chain rule** is the, fairly, simple method to differentiate such compositions of functions. As long as both functions are differentiable, we have that if $f(x) = (g \circ h)(x)$ then

$$f'(x) = g'(h(x))g'(x). \tag{3.20}$$

An alternative way of writing the chain rule, that is easier to understand in Leibniz notation, works when $y = f(u)$ and $u = g(x)$ then

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}. \tag{3.21}$$

This second way of phrasing the chain rule makes it look like we are treating the derivative like a fraction, we are not, it is just a coincidence that the formula looks like this.

Armed with the chain rule, eq. (3.20) we can return to eq. (3.19) and calculate its derivative

Example 3.18. Consider the function

$$f(x) = \sqrt{4x + 2},$$

and let $g(x) = \sqrt{x}$ and $h(x) = 4x + 2$. The rules we already know about differentiation tell us that

$$g'(x) = \frac{1}{2\sqrt{x}}, \quad h'(x) = 4.$$

Thus applying the chain rule gives

$$\begin{aligned} f'(x) &= g'(h(x))h'(x) \\ &= g'(4x+2)(4) \\ &= \frac{4}{2\sqrt{4x+2}} \\ &= \frac{2}{\sqrt{4x+2}}. \end{aligned}$$

While we have kept track of the function composition here, this is not how we proceed in general, particularly since this can become very complicated if we have a matryoshka doll like set up with many nested functions. Usually we will just think about an *inside* and *outside* function and then differentiate the *outside* function using the rules that we already know for powers, trig, exponential, or logarithmic functions, then multiply this by the derivative of the *inside* function.

This may still sound quite complicated, and as is always the case in mathematics, the best way to get to grips with the concept is by solving lots of examples.

Example 3.19. Consider the function

$$g(x) = \cos(x+c).$$

Its derivative is given by

$$\begin{aligned} g'(x) &= -\sin(x+c)(x+c)' \\ &= -\sin(x+c). \end{aligned}$$

Example 3.20. Consider the function

$$f(x) = (2x^2 + \cos(x))^2.$$

Its derivative is given by

$$\begin{aligned} f'(x) &= 2(2x^2 + \cos(x))(2x^2 + \cos(x))' \\ &= 2(2x^2 + \cos(x))(4x - \sin(x)) \end{aligned}$$

Exercise 3.21. Calculate the derivative of

$$f(x) = \exp(x^2 + x).$$

3.3 Multiple derivatives

So far we have only talked about calculating the derivative once. However, for a function $f(x)$ its derivative is also a function $f'(x)$ which we could differentiate again. The notation for taking the second derivative is

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

We can keep doing this and the notation for the n 'th derivative is

$$f^{(n)} = \frac{d^n f}{dx^n}.$$

Example 3.22. Consider the function $f(x) = 2x^3 + 4x$, its second derivative is

$$\begin{aligned} \frac{d^2 f}{dx^2} &= \frac{d}{dx} \left(\frac{df}{dx} \right) \\ &= \frac{d}{dx} (6x^2 + 4) \\ &= 12x. \end{aligned}$$

The third derivative is

$$\begin{aligned}\frac{d^3 f}{dx^3} &= \frac{d}{dx} \left(\frac{d^2 f}{dx^2} \right) \\ &= \frac{d}{dx} (12x) \\ &= 12.\end{aligned}$$

We will not do much with second, or higher, derivatives, other than when discussing optimisation problems in chapter 8. However, you may come across them in your future studies so it is worth having some exposure to them.

3.4 Applications of differentiation

So, now we have met the derivative and learnt some rules for using it. Some of you reading this are likely to be saying “Ok, we can calculate derivatives. So what?” Well, now we will look at a few applications of differentiation.

Finding critical points

A critical point of a function $f(x)$ is a point a where the derivative, $f'(a)$, vanishes. Sometimes critical points are called **stationary points**⁴.

Mathematical Diversion 3.2. *Strictly speaking there are a couple of extra points to consider. First we need that at the point a we have that $f(a)$ exists. Then a is a critical point if either $f'(a) = 0$ or $f'(a)$ does not exist. In [Dawkins, 2025b] there is a discussion of this emphasising the fact that the point a needs to be in the domain of the function f .*

If we have a function like $f(x) = 4x^2 + 3x - 2$ we can ask, does it have critical points, if so what are they and what does the function look like near them. The way that we answer the first two parts of this is to calculate the the derivative of the function, set it to zero, and then solve the algebraic equation that we get. In other words, finding a critical point boils down to solving an algebraic equation. The easiest way to understand this is by considering an explicit example.

Example 3.23. *Consider the function*

$$f(x) = 4x^2 + 3x - 2.$$

We find its critical points by calculating $f'(x)$ and setting it to zero. The derivative is

$$\begin{aligned}\frac{df}{dx} &= \frac{d}{dx} (4x^2 + 3x - 2) \\ &= 8x + 3.\end{aligned}$$

Setting this equal to zero gives

$$\begin{aligned}8x + 3 &= 0 \\ x &= -\frac{3}{8}.\end{aligned}$$

This polynomial is plotted in fig. 3.1 with its critical point marked and its derivative plotted in green. Notice that the critical point corresponds to the minimum of the function.

⁴This is because of the main examples of a first derivative is speed being the derivative of the position in mechanics. In this case a vanishing derivative means that the speed is zero so the object is stationary.

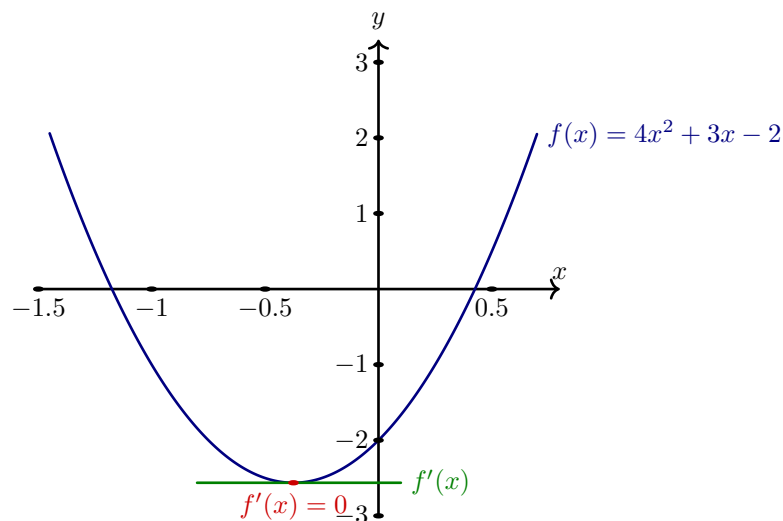


Figure 3.1: A plot of the function $f(x) = 4x^2 + 3x - 2$ with its critical point marked.

This is a general observation, critical points are related to where a function changes direction or stops moving. One way to understand this is that the derivative measures how “fast” a function is changing, it is the **rate of change** of the function.

As a check to see that this makes sense consider the plots of sine and cosine in fig. 2.6. Notice that when sine has its maxima and minima cosine is zero and the other way around. Recall that cosine is the derivative of sine, and that the derivative of cosine is minus sine, so the maxima and minima are critical points.

We need to be aware that not all critical points are maxima or minima. Even if the function does not change direction at a point, its derivative can still vanish there. These are called **points of inflection**, sometimes they are called **saddle points** as for functions of more than one variable they often have the shape of a saddle. See fig. 3.2 for an example of a point of inflection.

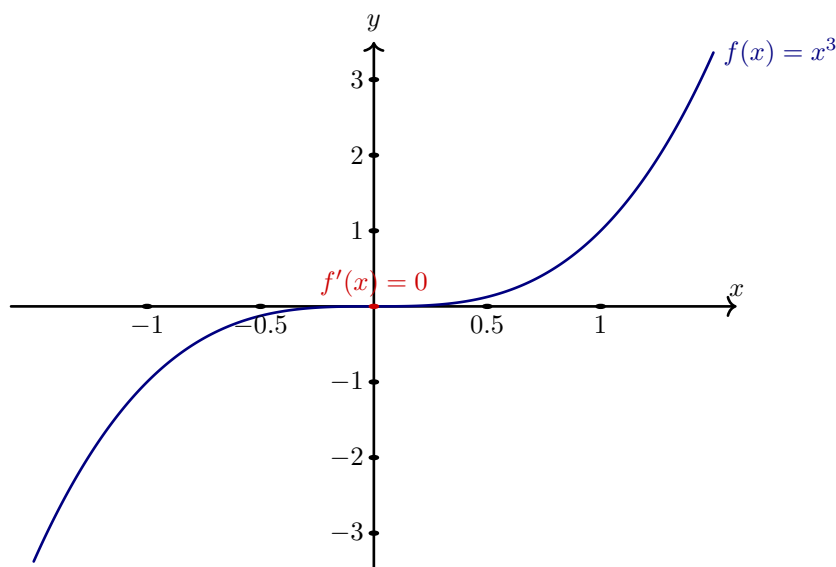


Figure 3.2: A plot of the function $f(x) = x^3$. The function has a critical point at $x = 0$ but it does not change direction there.

The other point to be aware of is that we are detecting critical, which can be **local** maxima or minima rather than just the **global** maxima and minima. For example if $f(x) = x^3/2 + 4x^2/5$, which is plotted in fig. 3.3, then we find the critical points as follows:

$$\begin{aligned} 0 &= f'(x) \\ &= \frac{3}{2}x^2 + \frac{8}{5}x \\ &= x^2 + \frac{16}{15}x. \end{aligned}$$

This is a quadratic equation which we solve either by using the quadratic formula or by extracting a common factor as follows,

$$\begin{aligned} 0 &= x^2 + \frac{16}{15}x \\ &= x \left(x + \frac{16}{15} \right), \end{aligned}$$

so the critical points are at $x = 0$ and $x = -16/15 \simeq -1.067$. By inspecting the graph we see that $x = 0$ is a local minima and $x = -16/15$ is a local maxima. They are local since for $x < -1.6$ $f(x)$ is below $f(x)$, and similarly for positive x $f(x)$ reaches values larger than at $x = -16/15$. The true maxima and minima are only reached asymptotically.

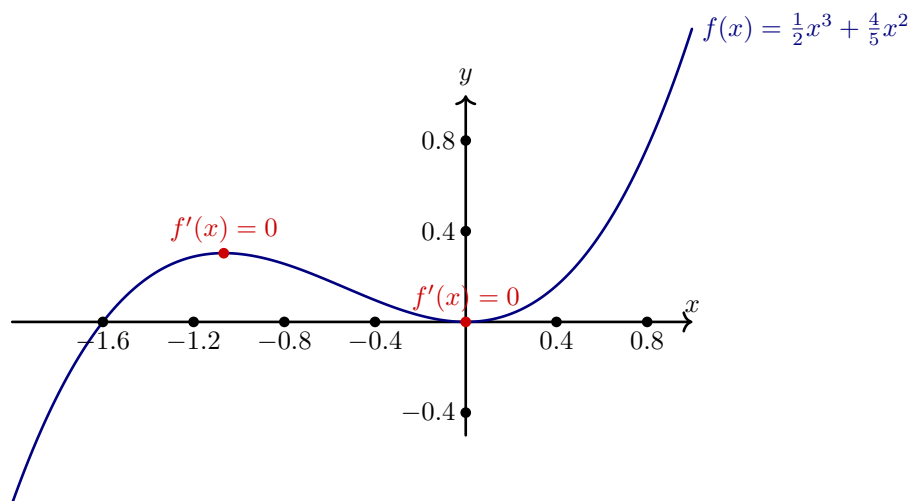


Figure 3.3: A plot of the function $f(x) = x^3/2 + 4x^2/5$. The function has a local minima at $x = 0$ but the global minima is at $x \rightarrow -\infty$.

Above we worked out whether a critical point was a maxima or a minima by consulting the graph. However, we can do this systematically without sketching the graph, which is handy since maxima and minima are useful to know before sketching a function. There are two ways to do this:

1. After finding a critical point a substitute it into the function to find its value $f(a)$, then pick two values of x , one just less than a , a_- and the other just bigger than a , a_+ . Then substitute these into $f(a)$.
 - If $f(a)$ is larger than both $f(a_-)$ and $f(a_+)$ then a is a local maxima.
 - If $f(a)$ is lower than both $f(a_-)$ and $f(a_+)$ then a is a local minima.
 - If we find that $f(a_-) > f(a) > f(a_+)$ or $f(a_+) > f(a) > f(a_-)$ then a is a point of inflection.
2. The other approach involves calculating the second derivative. If the derivative tells you about the rate of change of the function, then the second derivative tells you about the rate of change of the derivative.
 - If the derivative is decreasing at a critical point a , e.g. $f''(a) < 0$ then the point is a maxima.

- If $f''(a) > 0$ is the a is a local minima.
- If $f''(a) = 0$ then a could be a a maxima, minima, or point of inflection and we need to examine the sign of $f'(x)$ on either side of a .

In this module you can follow whichever approach you want if you are asked to identify the critical values.

Example 3.24. Consider the function

$$f(x) = x^3 - 3x,$$

we find and classify the critical points as follows. For the critical points we find the derivative

$$f'(x) = (x^3 - 3x)' = 3x^2 - 3,$$

and set it to zero,

$$\begin{aligned} 0 &= 3x^2 - 3 \\ &= x^2 - 1, \\ x^2 &= 1, \\ x &= \pm 1. \end{aligned}$$

So there are two critical points. To find the nature of the critical points we need the second derivative

$$f''(x) = (3x^2 - 3)' = 6x.$$

At the critical points this is

$$\begin{aligned} f''(1) &= 6, \\ f''(-1) &= -6. \end{aligned}$$

So $f''(1)$ is positive meaning that $x = 1$ is a minimum, while $f''(-1)$ is negative so $x = -1$ is a maxima. When we plot the graph in fig. 3.4 we see that these are only local minima and maxima.

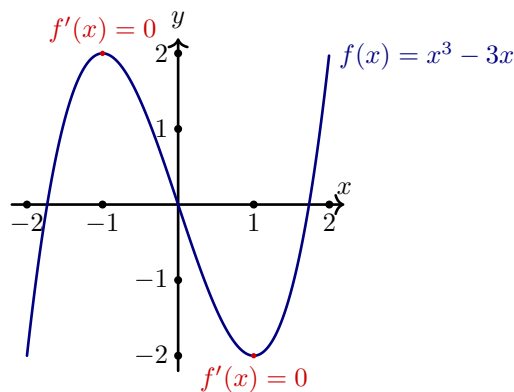


Figure 3.4: A plot of the function $f(x) = x^3 - 3x$ with its local minima and maxima marked.

Exercise 3.25. Find and classify the critical points of

$$f(x) = x^4 - 3x^2 + 2.$$

Linear approximations

Now we are going to look at how to approximate functions using their tangent lines. This is called the **linear approximation** to the function, and can be good near the point that the line is tangent to. However, the

approximation will get worse and worse the further away we go from the point⁵. There is a nice, if brief, discussion of linear approximation in [Dawkins, 2025b] which we draw on for the discussion here.

For a function $f(x)$, its tangent line at $x = a$ is given by the line

$$L(x) = f(a) + f'(a)(x - a). \quad (3.22)$$

For $f(x) = x^2/2 + 1/2$ the function and its linear approximation at $x = 1$ are plotted in fig. 3.5. We see that near the point $x = 1$ the tangent line is a good approximation to the full function as the graphs almost lie on top of each other.

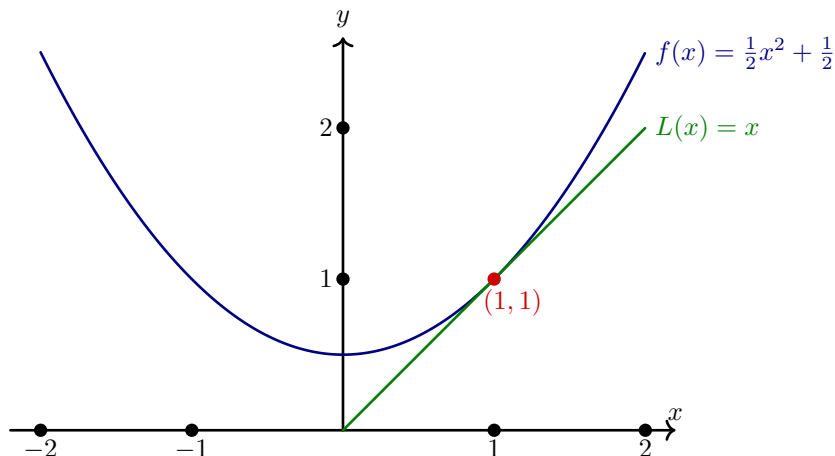


Figure 3.5: A plot of the function $f(x) = x^2/2 + 1/2$ with its linear approximation at $x = 1$ marked in red.

This means that if we want to find the values of $f(x)$ near the point a we can use eq. (3.22) instead of the full function. You may think that this does not seem very useful since we can fairly easily work out the value of $f(x)$ for many of the functions that we have considered in this module. The main advantage comes when we have to work out lots of values of a function, such as if we are simulating something on a computer. Every time the function is evaluated this takes time and it is much quicker to use an approximation than to use the full function.

Example 3.26. Consider $f(x) = \sin(x)$, its linear approximation at $x = 0$ is given by eq. (3.22). To calculate this we need that

$$f'(x) = \cos(x),$$

so we can calculate $f'(0) = 1$. This means that eq. (3.22) becomes

$$\begin{aligned} L(x) &= \sin(0) + \cos(0)(x - 0) \\ &= 0 + 1(x) \\ &= x. \end{aligned}$$

So for small angles we can use the approximation that $\sin(x) \simeq x$. This is an incredibly useful approximation, particularly if you are interested in studying the physics of pendulums.

If we repeated the above calculation for $\cos(x)$ we would find that $L(x) = 1$ at $x = 0$.

⁵We could make the approximation better by using higher order derivative terms, but the approximation would no longer be linear and would now be via a polynomial. Considering this would lead us to the **Taylor** or **Maclaurin series** of a function. You may have met this in A-level maths if not we may meet this in chapter 8 depending on which of the advanced topics we have time for.

Chapter 4

Integration

Integration: The inverse process to differentiation, i.e. the process of finding a function with a derivative that is a given function.

*Penguin Dictionary of
Mathematics*

4.1 Integration and antiderivatives

As the above quote from the *Penguin Dictionary of Mathematics* says, integration is the reverse process to differentiation. If differentiation is all about finding the rates of change of functions then we can think of integration as being about the areas under curves. As with differentiation it is a process that was discovered to handle problems from physics so you can find plenty of examples of applications by looking into topics from mechanics. We may meet some of those at the end of this chapter, but for now we will focus on the mathematics.

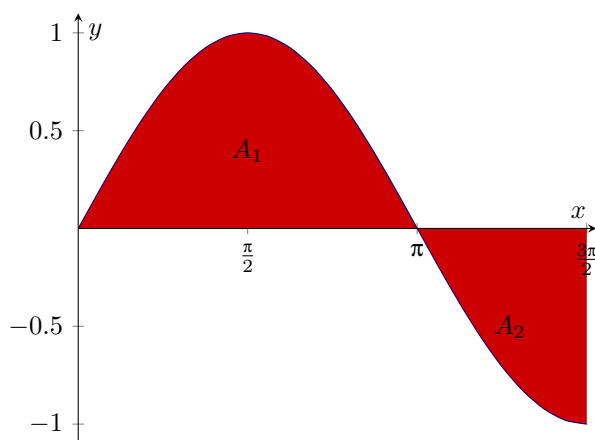


Figure 4.1: The graph of the sine function $f(x) = \sin(x)$ with the area between the curve and the x -axis between $x = 0$ and $x = 3\pi/2$ shaded. Note that the total shaded area is $A = A_1 + A_2$.

The integral of a function over an interval is a measure of the signed area between the curve and the x -axis, with the convention that areas above the x -axis are positive and areas under the x -axis are negative. This is shown for $f(x) = \sin(x)$ between 0 and $3\pi/2$ in fig. 4.1. When we discuss numerical methods of integration in chapter 6 we will understand how to approximate these areas using thin rectangles as is shown in fig. 1.4.

For now we will approach integration as the reverse of differentiation. In other words, if we have a function $f(x)$ we want to find the function whose derivative is $f(x)$.

Example 4.1. *If we have the function*

$$f(x) = x^4 + 2x^2 + 2$$

what function is this the derivative of?

To find this we need to remember our rules for differentiation, in particular the monomial rule of eq. (3.10), which says that we need to find an x^n , x^m and an x^p such that

$$\begin{aligned}\frac{d(ax^n)}{dx} &= anx^{n-1} = x^4, \\ \frac{d(bx^m)}{dx} &= bmx^{m-1} = 2x^2, \\ \frac{d(cx^p)}{dx} &= cpx^{p-1} = 2.\end{aligned}$$

By observation we find that $n = 5$, $a = 1/5$, $m = 3$, $b = 2/3$, $p = 1$ and $c = 2$. Thus if we have the function

$$F(x) = \frac{1}{5}x^5 + \frac{2}{3}x^3 + 2x,$$

its derivative is $F'(x) = f(x)$.

This looks like we are done. However, recall that in eq. (3.9) we said that the derivative of a constant is zero. Thus we could add any constant to $F(x)$ and still have a function whose derivative is $f(x)$. This means that we should write

$$F(x) = \frac{1}{5}x^5 + \frac{2}{3}x^3 + 2x + c,$$

where c can be any constant such as $0, 10, \pi, 11/25, -2, \dots$, when we have our brief discussion in chapter 5, we will see that often there are conditions that we need to apply which determine the constant.

Notice that since the constant in the example is arbitrary there are actually an infinite number of $F(x)$'s such that their derivative is the given $f(x)$. We call such functions $F(x)$ **antiderivatives** of $f(x)$.

As an aside, you might be wondering why a constant shows up here, but we did not mention anything about a constant when giving the intuition of an integral as the area under a curve. Surely the area is a fixed concept and cannot be arbitrarily changed by a constant? This is true. The difference is because when we talked about the area under a curve, we specified a region of the x -axis that we were interested in.

This leads to what is called a **definite integral** which we will study in more detail later. When we are constructing an antiderivative like $F(x)$, we are not specifying any limits which is why we need to add the constant. You will gain more experience with this as we look at some more examples.

In general we say that given a function $f(x)$, an antiderivative of $f(x)$ is any function $F(x)$ such that

$$\frac{dF}{dx} = f(x). \quad (4.1)$$

When we are explicitly including the arbitrary constant c we refer to the **indefinite integral** of $f(x)$ denoted by

$$\int f(x)dx = F(x) + c. \quad (4.2)$$

The symbol \int is the **integral** or **integration** symbol and in the above expression $f(x)$ is called the **integrand**, the function being integrated, while x is called the **integration variable** and c is the **constant of integration**.

As with many things in this module, the easiest way to get to grips with the content and the new terminology is through solving example problems.

Example 4.2. *The indefinite integral*

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

is calculated as in the same way as the antiderivative above. Thus we have

$$\int (x^4 + 2x^2 + 2) \, dx = \frac{1}{5}x^5 + \frac{2}{3}x^3 + 2x + c.$$

It is important to include the dx after the \int so that it is clear what is being integrated over. I also like to put brackets round the integrand when it involves multiple terms to make it unambiguous what is being integrated over.

Warning! If you are asked to calculate an indefinite integral you **must** include the constant of integration for your answer to be counted as correct.

Note that the integral has several nice properties which we will not prove here:

$$\int cf(x) \, dx = c \int f(x) \, dx, \quad (4.3)$$

$$\int (-f(x)) \, dx = - \int f(x) \, dx, \quad (4.4)$$

$$\int (f(x) \pm g(x)) \, dx = \int f(x) \, dx \pm \int g(x) \, dx. \quad (4.5)$$

These are analogous to eqs. (3.3) to (3.5) for the derivative. We have implicitly been using them already.

Note that for all of the functions that we know the derivative of we now know the integral of by reversing the identities so we have that

$$\begin{aligned} \int \cos(x) \, dx &= \sin(x) + c, \\ \int \sin(x) \, dx &= -\cos(x) + c, \\ \int \frac{1}{x} \, dx &= \ln(x) + c, \\ \int e^x \, dx &= e^x + c. \end{aligned}$$

We can also invert the identity that

$$\frac{dx^n}{dx} = nx^{n-1}$$

to be

$$\int x^n \, dx = \frac{1}{n+1}x^{n+1} + c \quad (4.6)$$

which holds for $n \neq -1$.

Example 4.3. We know that the derivative of $\tan(x)$ is $\sec^2(x)$ which means that the antiderivative of $\sec^2(x)$ is

$$F(x) = \tan(x),$$

and the indefinite integral is thus

$$\int \sec^2(x) dx = \tan(x) + c.$$

Exercise 4.4. Show that these trig functions have the claimed indefinite integrals:

$$\int \csc^2(x) dx = -\cot(x) + c,$$

$$\int \sec(x) \tan(x) dx = \sec(x) + c,$$

$$\int \csc(x) \cot(x) dx = -\csc(x) + c.$$

The process of calculating an integral by finding the antiderivative as the reverse of differentiation is sometimes called **integration by inspection**. This is very useful when you have a simple enough function that you already know what you differentiate to get it.

Example 4.5. The integral

$$\int dx,$$

is found by inverting the derivative identity $(x)' = 1$ to give

$$\int dx = \int 1 dx = x + c.$$

Example 4.6. Using some of the standard integrals that we have given above we evaluate

$$\int (x^2 + x^{-3} + x^{-1}) dx$$

as follows:

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

Note that here we have used eq. (4.6) for both the positive and negative power terms.

4.2 Techniques for integration

Integration by substitution

Now that we know the basics of how to calculate indefinite integrals we can introduce some common techniques. The first of these is **integration by substitution**, which is a way to integrate more complicated expressions by making them look like expressions that we already know how to integrate.

The idea behind integration by substitution is to consider an integral like

$$I = \int 3x^2 \sqrt[3]{x^3 + 4} dx.$$

We then observe that if we let $u = x^3 + 4$ and differentiate u with respect to x we get

$$\frac{du}{dx} = 3x^2.$$

Then we see that we can rewrite I as

$$I = \int \frac{du}{dx} \sqrt[3]{u} dx.$$

Remember back in chapter 3 we said that in certain circumstances we could treat a derivative as a bit like a fraction, well we can do this under an integral sign. It turns out that if we define a function $u(x)$ then integrating over u is related to integrating over x in the following way

$$\int f(u) du = \int f(u(x)) \frac{du}{dx} dx. \quad (4.7)$$

This means that in our above integral we can rewrite $\frac{du}{dx} dx = du$ so that

$$I = \int \frac{du}{dx} \sqrt[3]{u} dx = \int \sqrt[3]{u} du = \frac{3}{4} u^{\frac{4}{3}} + c = \frac{3}{4} (x^3 + 4)^{\frac{4}{3}} + c.$$

This is the same recipe that we will always use when integrating by substitution:

- Simplify the integrand as much as possible.
- Guess at a simplifying substitution $u = u(x)$, or use the substitution that you are given.
- Calculate the derivative du/dx and turn the integral over x into an integral over u .
- Evaluate the integral over u

Not all substitutions are as obvious as the above one, this is particularly true for trig substitutions where identifying the correct substitution to use takes practice.

Example 4.7. Consider the integral

$$I = \int \frac{1}{x^2 + 1} dx.$$

We can evaluate this using the substitution $x = \tan(u)$, then since $\tan(u) = \sin(u)/\cos(u)$ we use the quotient rule to evaluate the derivative of $\sin(u)$

$$\begin{aligned} \frac{d}{du} \tan(u) &= \frac{d}{du} \left(\frac{\sin(u)}{\cos(u)} \right) \\ &= \frac{1}{\cos^2} (\cos^2(u) + \sin^2(u)) \\ &= 1 + \tan^2(u). \end{aligned}$$

Recalling that $x = \tan(u)$ we have that

$$\frac{dx}{du} = 1 + x^2,$$

which means that

$$\begin{aligned} I &= \int \frac{1}{x^2 + 1} dx \\ &= \int \frac{1}{x^2 + 1} \frac{dx}{du} du \\ &= \int \frac{1}{x^2 + 1} (x^2 + 1) du \\ &= \int du \\ &= u + c \\ &= \arctan(x) + c. \end{aligned}$$

Approaches like the above example are how we can calculate integral expressions for all of the inverse trig functions. We just need to know the correct substitution.

Integration by parts

Integration by parts is the integral equivalent of the product rule from differentiation. We can use it whenever we have an integral which is the product of two functions and we know how to differentiate one of them and how to integrate the other.

Recall that the product rule says that given two functions of x , $f(x), g(x)$ the derivative of their product is

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

If we integrate this expression we have that

$$\begin{aligned}\int \frac{d}{dx}(fg) dx &= f(x)g(x) + c, \\ &= \int \left(\frac{df}{dx}g(x) + f(x)\frac{dg}{dx} \right) dx \\ &= \int \frac{df}{dx}g(x)dx + \int f(x)\frac{dg}{dx}dx.\end{aligned}$$

rearranging this and setting c to zero gives

$$\int f(x)\frac{dg}{dx}dx = f(x)g(x) - \int \frac{df}{dx}g(x)dx. \quad (4.8)$$

The reason that we can set c to zero is that we still have some other indefinite integrals that have not been evaluated and these contain a constant of integration. Thus we can absorb c into the remaining integral.

The formula in eq. (4.8) is known as the integration by parts formula. We will apply it for indefinite integrals first and then turn to definite integrals.

Example 4.8. Consider the integral

$$\int \ln(x)dx = \int \ln(x)1dx = \int \ln(x)\frac{dx}{dx}dx$$

this can be evaluated using integration by parts with $f(x) = \ln(x)$ and $g(x) = x$.

Applying eq. (4.8) gives

$$\begin{aligned}\int \ln(x)dx &= \int \ln(x)\frac{dx}{dx}dx \\ &= \ln(x)x - \int \left(\frac{d}{dx} \ln(x) \right) x dx \\ &= x \ln(x) - \int \frac{1}{x} x dx \\ &= x \ln(x) - x + c.\end{aligned}$$

The trick when applying integration by parts is to look for one function that is easy to integrate, which will be our $dg(x)/dx$, and another that is easy to differentiate, our $f(x)$.

Example 4.9. Consider the integral

$$\int x \sin(x)dx = \int x \frac{d}{dx}(-\cos(x)) dx.$$

We integrate this by parts taking $f(x) = x$ and $g(x) = -\cos(x)$, so that

$$\begin{aligned}\int x \sin(x) dx &= \int x \frac{d}{dx} (-\cos(x)) dx \\ &= x(-\cos(x)) - \int \left(\frac{dx}{dx}\right) (-\cos(x)) dx \\ &= -x \cos(x) + \int \cos(x) dx \\ &= -x \cos(x) + \sin(x) + c.\end{aligned}$$

Exercise 4.10. Use integration by parts to evaluate

$$I = \int x e^{2x} dx.$$

In this module we will only come across problems where you need to apply integration by parts once. However, some integrals require us to apply integration by parts more than once, and some require other tricks.

Mathematical Diversion 4.1. As an example of repeated integration by parts consider

$$I = \int x^2 \sin(x) dx = \int x^2 \frac{d}{dx} (-\cos(x)) dx.$$

We can carry out an integration by parts taking $f(x) = x^2$ and $g(x) = -\cos(x)$ to get

$$\begin{aligned}I &= \int x^2 (-\cos(x)) dx \\ &= -x^2 \cos(x) - \int 2x (-\cos(x)) dx \\ &= -x^2 \cos(x) + 2 \int x \cos(x) dx \\ &= -x^2 \cos(x) + 2 \int x \frac{d}{dx} (\sin(x)) dx.\end{aligned}$$

The integral in the last line is one that we have to do by parts taking $f(x) = x$ and $g(x) = \sin(x)$. This means that

$$\begin{aligned}I &= -x^2 \cos(x) + 2 \int x \frac{d}{dx} (\sin(x)) dx \\ &= -x^2 \cos(x) + 2 \left(x \sin(x) - \int \sin(x) dx \right) \\ &= -x^2 \cos(x) + 2(x \sin(x) + \cos(x) + c) \\ &= -x^2 \cos(x) + 2x \sin(x) + 2 \cos(x) + \tilde{c},\end{aligned}$$

where in the last line we have redefined the constant of integration to be $\tilde{c} = 2c$.

Partial fractions

Another integration technique that is useful for integration fractions is known as **Partial fractions**. It is an example of a simplification method and is not examinable within this module. However, it may be useful for you to have seen it at least once.

The idea is to take an expression like $1/(x^2 - 1)$ and express it as the sum of two fractions

$$\frac{1}{x^2 - 1} = \frac{1}{(x - 1)(x + 1)} = \frac{A}{x + 1} + \frac{B}{x - 1},$$

for A and B constants that need to be determined. This is done by multiplying through by $(x+1)(x-1)$ to get

$$1 = A(x-1) + B(x+1),$$

which rearranges to

$$(A+B)x - A + B - 1 = 0.$$

For this to be true we need to match up coefficients of powers of x on both sides of the equation so that

$$\begin{aligned} A+B &= 0, \\ B-A-1 &= 0, \end{aligned}$$

which is solved by

$$\begin{aligned} A &= -\frac{1}{2}, \\ B &= \frac{1}{2}. \end{aligned}$$

Which gives

$$\frac{1}{x^2-1} = -\frac{1}{2} \left[\frac{1}{x+1} - \frac{1}{x-1} \right].$$

These are both expressions that we can integrate to get

$$\begin{aligned} \int \frac{1}{x^2-1} dx &= -\frac{1}{2} \left[\int \frac{1}{x+1} dx - \int \frac{1}{x-1} dx \right] \\ &= -\frac{1}{2} \ln(x+1) + \frac{1}{2} \ln(x-1) + c \\ &= \ln \left[\left(\frac{x-1}{x+1} \right)^{\frac{1}{2}} \right] + c. \end{aligned}$$

Note that in terms of what we know at this stage, the right hand side is only valid for $x > 1$.

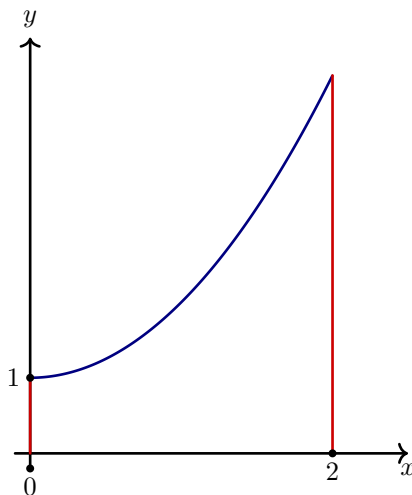
The general procedure is as follows:

- Factorise the denominator in terms of its roots.
- Write down a sum of the reciprocals of the roots with arbitrary coefficients. This step becomes complicated if there are repeated roots.
- Multiply through by the denominator on the left hand side and express it as a polynomial in x .
- Solve the system of equations for the coefficients A, B, \dots .
- Substitute the coefficients into the partial fractions expression and carry out the integral making use of integration by parts or substitutions as necessary.

As mentioned in the second step, if there are repeated roots this becomes trickier, since this is already a non-examinable diversion we will not give the details here.

4.3 Definite integrals

In the previous section we focussed on antiderivatives and indefinite integrals, now we return to the original motivation for the integral, as the measure of the area under a curve. This is called a definite integral and returns a number rather than a function as it involves us carrying out a definite integral and then evaluating the result at the end points. This may sound confusing at first but we will learn how to do it by seeing some examples.

Figure 4.2: A plot of our favourite function $f(x) = x^2 + 1$.

To start consider the function $f(x) = x^2 + 1$ on the interval $[0, 2]$, the full details of this example are given in Dawkins [2025b]. The idea is that we want to find the area enclosed between the curve and the x -axis, shown in fig. 4.2. Once we can integrate we will be able to compute this exactly, but for now we should consider how to approximate this area.

A shape that we understand how to calculate the area of really well is a rectangle. If a rectangle has height Δy and base length Δx then its area is $\Delta x \times \Delta y$. The aim is then to find the area under the curve by breaking it up into rectangles with a regular width. We do this by splitting up the interval, $[0, 2]$ in this case, into n subintervals. This will give us an approximation of the area which will improve the more sub intervals we take, in other words the larger that n is. The width of a subinterval is then

$$\Delta x = \frac{2 - 0}{n} = \frac{2}{n}.$$

In general it will be $\Delta x = (b - a)/n$, where the interval is $[a, b]$.

We will choose the height of the intervals so that they agree with the value of f at the right hand side of the subinterval. For example if $n = 4$ we have four subintervals of length $1/2$,

$$\left[0, \frac{1}{2}\right], \quad \left(\frac{1}{2}, 1\right], \quad \left(1, \frac{3}{2}\right], \quad \left(\frac{3}{2}, 2\right].$$

In each sub interval we take the height of our rectangle to be

$$\begin{aligned} f\left(\frac{1}{2}\right) &= \frac{5}{4}, \\ f(1) &= 2, \\ f\left(\frac{3}{2}\right) &= \frac{13}{4}, \\ f(2) &= 5. \end{aligned}$$

Plotting the rectangles we get the picture in fig. 4.3, note that the area of the rectangles is larger than the area under the curve. As an exercise consider the case where we take the rectangle to have the height of the function on the left boundary of the integral, what changes?

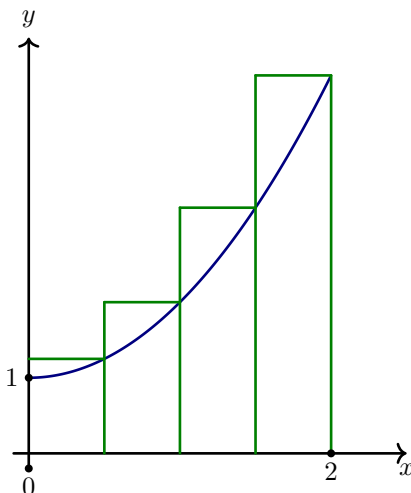


Figure 4.3: A plot of $f(x) = x^2 + 1$ with rectangles approximating the area under the curve.

Now let's calculate the area of each of these integrals by multiplying the width, $1/2$ for all four, by the height:

$$\begin{aligned} A_1 &= \frac{1}{2} \times \frac{5}{4} = \frac{5}{8}, \\ A_2 &= \frac{1}{2} \times 2 = 1, \\ A_3 &= \frac{1}{2} \times \frac{13}{4} = \frac{13}{8}, \\ A_4 &= \frac{1}{2} \times 5 = \frac{5}{2}. \end{aligned}$$

Adding all of these together gives

$$\begin{aligned} A_r &= A_1 + A_2 + A_3 + A_4 \\ &= \frac{5}{8} + 1 + \frac{13}{8} + \frac{5}{2} \\ &= \frac{23}{4} = 5.75. \end{aligned}$$

If we had taken the heights at the left end points we would have found $A_l = 3.75$, which is an underestimate of the area, again I invite you to check the details for yourself, or look at Dawkins [2025b] which goes through it.

Finally, we could make a third choice where we take the height of the rectangle to be the value of the function at the midpoint, shown in fig. 4.4, then we find that the area is $A_m = 37/8 = 4.625$.

You will see that the value of the area that we calculate depends on the size of our rectangles. This should not be surprising, but suggests that we need a more systematic way to do this, which enables us to find a unique answer. This is what integration will give us. Once we have the machinery of integration we can calculate the actual area under the curve to be $A = 14/3 = 4.67$. So the mid point method above is the closest to the true value.

This leads us to the concept of the **Riemann sum**¹. Consider a function plotted over an interval. Split the integral into n subintervals each of width Δx as above. Then pick a point x_i^* in each sub interval and use the rectangle of height $f(x_i^*)$ as our approximation for the area of the function over that sub interval. The total

¹Bernhard Riemann was a German Mathematician who contributed to many different areas of mathematics and is maybe most famous for the Riemann hypothesis about the distribution of prime numbers.

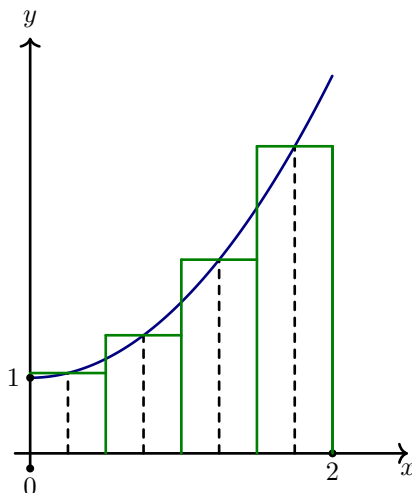


Figure 4.4: A plot of $f(x) = x^2 + 1$ with rectangles approximating the area under the curve.

area is then

$$A \simeq f(x_1^*)\Delta x + f(x_2^*)\Delta x + \cdots + f(x_i^*)\Delta x + \cdots + f(x_n^*)\Delta x = \sum_{i=1}^n f(x_i^*)\Delta x. \quad (4.9)$$

This approximation gets better and better the more subintervals we take, as they will become narrower and narrower, and their height becomes a better match for the functions value over that interval. If we could take the limit of $n \rightarrow \infty$ the sub intervals will become a single point and the height is the value of the function at that point. The sum would then become the exact area,

$$A = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*)\Delta x. \quad (4.10)$$

It is important to note that if our function was below the x -axis, then we would get a negative answer for the area. This is because we take the heights of the rectangles to be the value of the value of the function at some point in the interval. So if the function takes negative values, we will find a negative area.

This is a feature and not a bug! The integral will be a signed count of the area under a curve. This does mean that if a function crosses the axis in the interval that we are interested in, we will need to be careful whether we are interested in the net area under the curve, the difference between the positive and negative areas, or the total area, which does not care about the sign.

Exercise 4.11. Consider the function $x^2 - 4$ on the interval $[0, 2]$. Take $n = 8$ and use the midpoint approach to estimate the area between the function and the x -axis. You should find $A_m = -171/32$.

The sum in eq. (4.10) is taken as the definition of the definite integral:

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*)\Delta x. \quad (4.11)$$

This is the integral of the function $f(x)$ over the interval $[a, b]$. The terminology that is usually used is that a is called the lower limit and b the upper limit.

Mathematical Diversion 4.2. If this was a module for mathematicians, we would spend a bit of time here discussing how to calculate this using the sum definition given on the right hand side of eq. (4.11). Doing this

involves knowing some standard expressions for summations such as:

$$\begin{aligned}\sum_{i=1}^n 1 &= n, \\ \sum_{i=1}^n i &= \frac{n(n+1)}{2}, \\ \sum_{i=1}^n i^2 &= \frac{n(n+1)(2n+1)}{6}.\end{aligned}$$

Armed with these and a few other expressions one can evaluate the Riemann sum expression for a wide variety of functions. Including $f(x) = x^2 + 1$ that we discussed above. If you are interested in trying this it is worth doing at least once, though is an exercise for maths enthusiasts only as you are unlikely to need to use the Riemann sum definition of an integral directly.

The properties that we discussed for indefinite integrals still hold here, but now we get some new properties:

$$\begin{aligned}\int_a^b f(x)dx &= -\int_b^a f(x)dx, \quad \text{swapping limits introduces a minus sign.} \\ \int_a^a f(x)dx &= 0, \quad \text{if the limits agree the integral is zero.} \\ \int_a^b f(x)dx &= \int_a^c f(x)dx + \int_c^b f(x)dx, \quad \text{where } c \text{ is any number between } a \text{ and } b.\end{aligned}$$

Recall that the antiderivative $F(x)$ for the function $f(x)$ is defined as any function such that

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

For a definite integral antiderivatives make an appearance through the fundamental theorem of calculus. This states that for a function $f(x)$ which is continuous on $[a, b]$ and has antiderivative $F(x)$ then:

$$\int_a^b f(x)dx = [F(x)]_a^b = F(b) - F(a). \quad (4.12)$$

In other words, a definite integral is just the difference between the antiderivative evaluated at the endpoints of the interval. This result is proved in Dawkins [2025b], and a proof may be added to chapter 12 at some stage.

Example 4.12. Consider the function $f(x) = x^2 + 1$, we know that an antiderivative for this is

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

where we have carried out the indefinite integral of $f(x)$. Using eq. (4.12) we get that the definite integral over the interval $[0, 2]$ is

$$\begin{aligned}\int_0^2 (x^2 + 1) dx &= \left[\frac{1}{3}x^3 + x \right]_0^2 \\ &= \frac{1}{3}(2)^3 + 2 - \left(\frac{1}{3}(0)^3 + 0 \right) \\ &= \frac{8}{3} + 2 - 0 \\ &= \frac{14}{3}\end{aligned}$$

which is what we claimed was the exact value above.

Note that we can use any antiderivative since the constant piece will cancel out in the difference for example if we took

$$F_2(x) = \frac{1}{3}x^3 + c,$$

then

$$\left[\frac{1}{3}x^3 + x + c \right]_0^2 = \frac{14}{3} + c - (0 + c) = \frac{14}{3} + c - c = \frac{14}{3}.$$

This is why we always take the simplest antiderivative when evaluating definite integrals.

Example 4.13. Consider the function $f(x) = x^4 + 2x^2 + 2$ and calculate its integral over the interval $[0, 1]$. Recall that an antiderivative of this function is

$$F(x) = \frac{1}{5}x^5 + \frac{2}{3}x^3 + 2x,$$

so using eq. (4.12) the definite integral is

$$\begin{aligned} \int_0^1 (x^4 + 2x^2 + 2) dx &= \left[\frac{1}{5}x^5 + \frac{2}{3}x^3 + 2x \right]_0^1 \\ &= \frac{1}{5} + \frac{2}{3} + 2 - 0 \\ &= \frac{43}{15}. \end{aligned}$$

As with derivatives you will not evaluate every integral by finding an antiderivative. You will become familiar with a range of standard integrals, and techniques for evaluating integrals so that you do not have to calculate an antiderivative every time. The techniques of substitution and integration by parts that we met above are very useful here. Though we now need to be careful about what happens to the limits when we perform a substitution and when dealing with the total derivative part of integration by parts.

Example 4.14. Consider the integral

$$I = \int_0^1 \frac{1}{x^2 + 1} dx.$$

We evaluated the indefinite integral above using the substitution $x = \tan(u)$ which meant that

$$\int \frac{1}{x^2 + 1} dx = \int du.$$

The new step is that we need to look at what happens to the limits of the integral upon substitution. e.g. find the values of u such that $\tan(u) = 0$ and $\tan(u) = 1$, remembering that we need to work in radians. In this case $\tan(0) = 0$ and $\tan(\pi/4) = 1$ so we have that

$$I = \int_0^1 \frac{1}{x^2 + 1} dx = \int_0^{\pi/4} du = [u]_0^{\pi/4} = \frac{\pi}{4}.$$

Any time you use integration by substitution for a definite integral you similarly need to keep track of how the limits transform under the substitution.

With integration by parts the identity in eq. (4.8) is modified to become

$$\int_a^b f(x) \frac{dg}{dx} dx = [f(x)g(x)]_a^b - \int_a^b \frac{df}{dx} g(x) dx. \quad (4.13)$$

In other words we need to evaluate the total derivative at the limits of integration.

Example 4.15. Consider the integral

$$I = \int_1^2 \ln(x) dx.$$

We saw above that integration by parts gave that

$$\int \ln(x) dx = x \ln(x) - x + c.$$

When we calculated this we took $f(x) = \ln(x)$ and $g(x) = x$ so that

$$\begin{aligned} I &= \int_1^2 \ln(x) dx \\ &= \int_1^2 \ln(x) \frac{dx}{dx} \\ &= [\ln(x)x]_1^2 - \int_1^2 1^2 dx \\ &= 2 \ln(2) - \ln(1) - [x]_1^2 \\ &= 2 \ln(2) - (2 - 1) \\ &= 2 \ln(2) - 1. \end{aligned}$$

Improper integrals

An important type of definite integrals are the **improper integrals**. These are definite integrals

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

where either $f(x)$ becomes infinite for $a \leq x \leq b$, or one of the limits goes to infinity.

The following integrals are improper:

•

$$\int_1^\infty \frac{1}{x^2} dx,$$

because the upper limit is infinity,

•

$$\int_0^1 \frac{1}{\sqrt{x}} dx$$

because $1/\sqrt{x}$ becomes infinite for $x = 0$.

•

$$\int_1^\infty \sin(x) dx,$$

because the upper limit is infinity,

It turns out that some improper integrals can still give sensible results, we just need to interpret them properly.

First up, we reinterpret them using a limit so that

$$\begin{aligned} \int_1^\infty \frac{1}{x^2} dx &= \lim_{X \rightarrow \infty} \int_1^X \frac{1}{x^2} dx \\ &= \lim_{X \rightarrow \infty} \left[-\frac{1}{x} \right]_1^X \\ &= \lim_{X \rightarrow \infty} \left(1 - \frac{1}{X} \right) \\ &= 1. \end{aligned}$$

Thus the improper integral gives us a sensible answer.

Next consider

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{x}} dx &= \lim_{X \rightarrow 0} \int_X^1 \frac{1}{\sqrt{x}} dx \\ &= \lim_{X \rightarrow 0} [2\sqrt{x}]_X^1 \\ &= \lim_{X \rightarrow 0} 2(1 - \sqrt{X}) \\ &= 2.\end{aligned}$$

When we can reinterpret improper integrals using limits and get a finite answer we say that the integral is **convergent**. However, this does not always work as sometimes the improper integral is **divergent**.

Example 4.16. Consider the integral

$$I = \int_0^\infty \sin(x) dx.$$

If we try to interpret it as above we find that

$$\begin{aligned}I &= \int_0^\infty \sin(x) dx \\ &= \lim_{X \rightarrow \infty} \int_0^X \sin(x) dx \\ &= \lim_{X \rightarrow \infty} [-\cos(x)]_0^X \\ &= \lim_{X \rightarrow \infty} (1 - \cos(x))\end{aligned}$$

which does not converge as $\cos(x)$ oscillates between -1 and 1 .

Example 4.17. Consider the integral

$$I = \int_0^\infty x dx.$$

We can attempt to evaluate this as follows:

$$\begin{aligned}I &= \int_0^\infty x dx \\ &= \lim_{X \rightarrow \infty} \int_0^X x dx \\ &= \lim_{X \rightarrow \infty} \left[\frac{1}{2} x^2 \right]_0^X \\ &= \lim_{X \rightarrow \infty} \frac{1}{2} X^2,\end{aligned}$$

which tends to infinity, so this is another divergent integral.

Example 4.18. Consider the integral

$$I = \int_0^1 \frac{1}{x} dx.$$

We attempt to evaluate this as follows:

$$\begin{aligned}I &= \int_0^1 \frac{1}{x} dx \\ &= \lim_{X \rightarrow 0^+} \int_X^1 \frac{1}{x} dx \\ &= \lim_{X \rightarrow 0^+} [\ln(x)]_X^1 \\ &= \lim_{X \rightarrow 0^+} (-\ln(X)),\end{aligned}$$

which again tends to infinity.

Splitting the range

Another useful technique for evaluating definite integrals is to split the range of integration. For example if we are interested in the integral of

$$|x| = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0, \end{cases}$$

between -1 and 1 . Because the function changes when the interval goes from negative to positive we need to split the integral using the identity that we learnt above:

$$\begin{aligned} \int_{-1}^1 |x| dx &= \int_{-1}^0 |x| dx + \int_0^1 |x| dx \\ &= -\int_{-1}^0 x dx + \int_0^1 x dx \\ &= -\frac{1}{2} [x^2]_{-1}^0 + \frac{1}{2} [x^2]_0^1 \\ &= \frac{1}{2} + \frac{1}{2} = 1. \end{aligned}$$

Exercise 4.19. Evaluate the integral

$$I = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin(|x|) dx$$

by splitting the range of the integral.

Symmetries of functions

Before moving on to some applications of integration we have time to learn one more useful trick. We call a function $f(x)$ even if

$$f(-x) = f(x) \tag{4.14}$$

and we call it odd if

$$f(-x) = -f(x). \tag{4.15}$$

In other words an even function does not care about the sign of its argument, while an odd function changes sign if its argument does.

An example of an even function is $\cos(x)$ while $\sin(x)$ is an odd function.

Exercise 4.20. Can you come up with more examples of even and odd functions?

When we integrate an even function over a symmetric range, $[-a, a]$ we have that

$$\int_{-a}^a f(x) dx = 2 \int_0^a f(x) dx. \tag{4.16}$$

while if we integrate an odd function over the same region it will vanish,

$$\int_{-a}^a f(x) dx = 0. \tag{4.17}$$

Exercise 4.21. Prove that these formulae are true. Hint: You will want to use the range splitting that we learnt about previously.

4.4 Applications of integrals

As with the derivative, there are plenty of applications of integrals. We will touch on some of them here, but as usual there is a wealth of further details in both [Dawkins, 2025b] and [Riley et al., 2006].

Averaging functions

The average value of a continuous function $f(x)$ can be calculated by integration. For a discrete set of n numbers $\{x_1, x_2, \dots, x_n\}$ the average is given by

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}.$$

If we went back to the Riemann sum definition of the integral in eq. (4.10), then it can be shown that for a continuous function on the interval $[a, b]$ that

$$f_{\text{avg}} = \frac{1}{b-a} \int_a^b f(x) dx. \quad (4.18)$$

Mathematical Diversion 4.3. *In this module you do not need to know where the identity in eq. (4.18) comes from, but if you are interested a proof is included in the Proof of Various Integral Properties section of Dawkins [2025b]. Going through the proof of the result shows you that this is the natural extension of the concept of an average to a continuous function. In a module for mathematicians, we would also be learning some interesting results on upper bounds of the value of an integral, called the **ML Lemma** in many places, where we can show that under some assumptions that*

$$\left| \int_a^b f(x) dx \right| \leq ML,$$

where $L = b - a$ is the length of the interval, and M is an upper bound on the value of $|f(x)|$ over the interval.

Example 4.22. We can find the average of $f(x) = \sin(x)$ on the interval $[0, \pi]$ as follows:

$$\begin{aligned} f_{\text{avg}} &= \frac{1}{b-a} \int_a^b f(x) dx \\ &= \frac{\pi - 0}{\int_0^\pi} \sin(x) dx \\ &= \frac{1}{\pi} [-\cos(x)]_0^\pi \\ &= \frac{2}{\pi} \end{aligned}$$

Exercise 4.23. Find the average value of $\cos(x)$ over the interval $[0, 2\pi]$.

In some applications it is the average value of the function so you can use this approach to calculate it.

Areas between curves

This subsection is currently lacking pictures, this will be added eventually.

When we introduced the integral we gave the intuition to think of it as the area between the curve given by $y = f(x)$ and the x -axis. Now the x -axis is just the curve given by $y = 0$, so a natural question to ask is if we can use an integral to calculate the area between two curves. The answer is yes.

Consider two functions $f(x), g(x)$ on the interval $[a, b]$. The area is given by the integral of the difference between these two function $f(x) - g(x)$ or $g(x) - f(x)$, but we need to decide which way round to calculate the difference. The convention is to look at the graphs of the functions and subtract the lower function from the upper function.

This means that if $f(x) \geq g(x)$ for all $a \leq x \leq b$, the area between the curves is

$$A = \int_a^b (f(x) - g(x)) \, dx. \quad (4.19)$$

If we took the functions the other way around we would get the same numerical answer but with a negative sign. It is important that you always check which function is greater over the interval. This can be checked by producing a plot of the functions. If the functions cross in the interval, then you should split the integral in to two regions and evaluate these separately to find the total area enclosed between the two curves. There are some examples of problems like this in the Area Between Curves section of Dawkins [2025b].

Example 4.24. Consider the two functions $f(x) = x^2$ and $g(x) = x^3$ shown in fig. 4.5. Over the interval $[0, 1]$ $f(x) \geq g(x)$ so the area is given by

$$A = \int_0^1 (x^2 - x^3) \, dx.$$

Evaluating this integral we find that

$$\begin{aligned} A &= \int_0^1 (x^2 - x^3) \, dx \\ &= \left[\frac{1}{3}x^3 - \frac{1}{4}x^4 \right]_0^1 \\ &= \frac{1}{3} - \frac{1}{4} - 0 \\ &= \frac{1}{12}. \end{aligned}$$

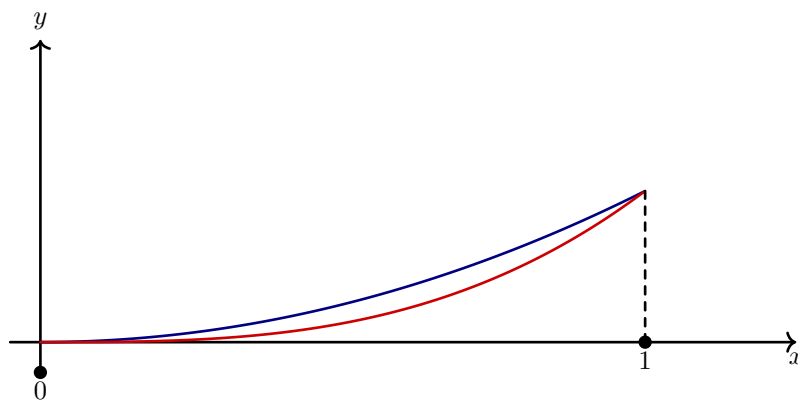


Figure 4.5: A plot of the functions $f(x) = x^2$ and $g(x) = x^3$ over the interval $[0, 1]$.

Exercise 4.25. Calculate the area between the curves $\cos(x)$ and $\sin(x)$ over the interval $[0, \pi/4]$.

Chapter 5

Differential Equations

This is a non-examinable section and is not taught every year. It is included in the indicative content of the module but is not assessed by the learning outcomes. No content has been included here yet, but when it is it can all be considered one big mathematical deviation.

Chapter 6

Numerical Methods

6.1 Solving equations numerically

It frequently happens that when confronted with a problem you end up with a mathematical problem or expression that cannot be evaluated analytically. This may be a differential equation that we cannot solve, an integral that has no closed form solution, or even an algebraic equation which does not yield an explicit solution. This is where we have to turn to computer packages or follow a numerical method. Nowadays many computer programs can implement these approaches as standard. However, it is a good idea to understand how to implement a few of these approaches by hand as you may end up writing one of your own at some stage.

Chapter 27 of [Riley et al., 2006] contains a lot of information about this topic and is a good place to go for more details than we will discuss here.

Before moving on to numerical differentiation and integration we will first discuss how to solve algebraic equations numerically. This is all about finding the real roots of an equation $f(x) = 0$. This equation can either be algebraic, with $f(x)$ being a polynomial, or transcendental, if $f(x)$ includes trig, log, or exponential terms. We will be using iterative schemes to solve the equations, where we make successive approximations to the true solution, which converge to the true solution. For a method to be successful we need both that the approximations converge to the true solution, and also that we only have a finite number of solutions.

It is important to note that different iteration methods will be better at solving certain kinds of problems, in practice some combination of methods is likely to be used. We will see four methods, **Rearrangement**, **Linear interpolation**, and **Bisection**, before discussing **Newton's** method, which is the most important method to know in this module.

Rearrangement

To explain what we mean by an iterative scheme it is useful to consider the first example method, recasting our equation $f(x) = 0$ as

$$x = \phi(x),$$

for some slowly varying function $\phi(x)$. Our iteration scheme starts from a guess, x_0 which will not solve $f(x_0) = 0$ but should be picked so that it is close to zero¹. We then have that $x_1 = \phi(x_0)$ and can keep repeating this process, i.e. $x_{n+1} = \phi(x_n)$. The solution we are looking for will be a value of x that is left unchanged by applying ϕ .

¹It is useful to plot or roughly sketch the function $f(x)$ as then it can be easier to pick a starting point.

The different methods are different ways of building $\phi(x)$ so that we can start the iteration scheme.

As an example of how to use the rearrangement iteration consider

$$f(x) = x^5 - 3x^3 + 2x - 4 = 0. \quad (6.1)$$

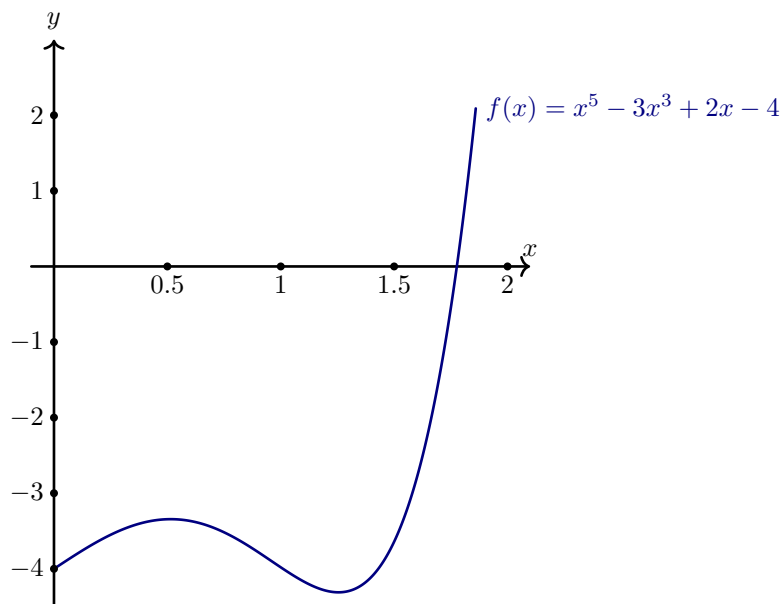


Figure 6.1: A plot of the function $f(x) = x^5 - 3x^3 + 2x - 4$ for $0 \leq x \leq 1.86$.

By rearranging this equation we have that

$$x = (3x^3 - 2x + 4)^{\frac{1}{5}}.$$

From the plot we in fig. 6.1 we see that the root is between $x = 1.5$ and $x = 2$ and to see how the method works we will not attempt to start from that close to the true root. Here we take $x_0 = 1.8$ and then iterate using

$$x_{n+1} = (3x_n^3 - 2x_n + 4)^{\frac{1}{5}}.$$

The successive values of x_n and $f(x_n)$ are shown in table 6.1 and they are closing in on

$$x = 1.757632748, \quad (6.2)$$

which is the value of the root to 9 decimal places. We see from the table that once we got to x_5 we were only differing from the precise answer in the third decimal places, so are within two parts in 10^3 . We can also see that lots of iterations are needed to get an accurate answer. The precise number will depend on the specific problem.

Exercise 6.1. *Implement this iteration method numerically for eq. (6.1) using a programming language or computer program of your choice. I would suggest trying Python or Matlab, but you can use whatever you feel most comfortable with.*

Linear interpolation

The next method to look at is **linear interpolation**. The idea here is to pick two points A_1, B_1 on the graph $y = f(x)$ that lie on either side of the root, i.e. so that $f(A_1)$ and $f(B_1)$ have opposite signs. We can then think about the straight line² joining these points ($A_1, f(A_1)$) to ($B_1, f(B_1)$). An example of this is shown in

²This straight line is known as a chord.

Table 6.1: Successive approximations to the root of eq. (6.1) using the rearranged equation. As the convergence is relatively slow we jump several steps at the end to show you it getting closer.

n	x_n	$f(x_n)$
0	1.8	0.99968
1	1.7805378	0.52247969
2	1.7700175	0.27736234
3	1.7643295	0.14849136
4	1.761254	0.07986324
5	1.7595909	0.043059716
12	1.75765592	0.00058017842
18	1.7576331	7.8434818×10^{-6}

fig. 6.2 where $A_1 = 1.5$ and $B_1 = 1.8$.

The intersection of this straight line with the x -axis is given by

$$x_1 = \frac{A_1 f(B_1) - B_1 f(A_1)}{f(B_1) - f(A_1)},$$

We then replace A_1 or B_1 by x_1 , find the point on the curve corresponding to $x = x_1$ and then iterate this process using the interpolation formula

$$x_n = \frac{A_n f(B_n) - B_n f(A_n)}{f(B_n) - f(A_n)}. \quad (6.3)$$

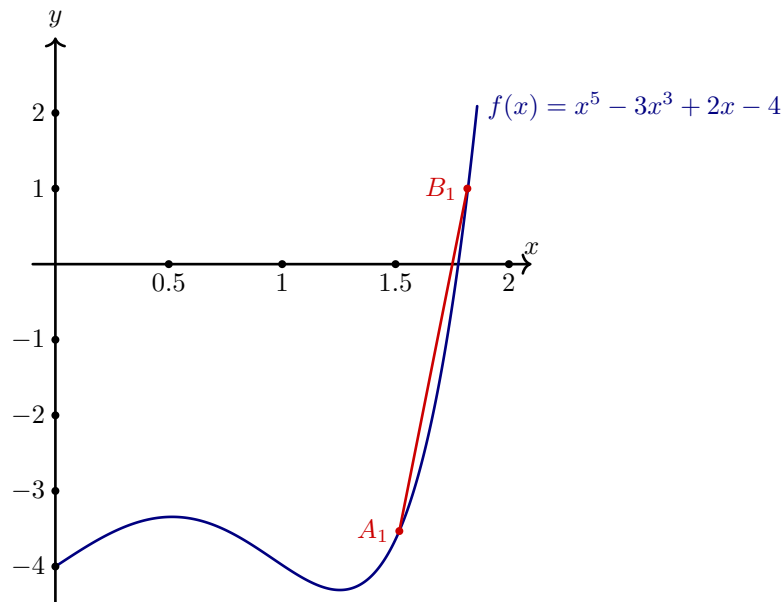


Figure 6.2: A plot of the function $f(x) = x^5 - 3x^3 + 2x - 4$ for $0 \leq x \leq 1.86$ with the first linear interpolation chords shown in red.

With each iteration the chord will become shorter and the end points will move until the value of x_n converges

to the precise value of the root. The first few steps of this interpolation starting from $A_1 = 1.5$ and $B_1 = 1.8$ are shown in table 6.2.

Table 6.2: Successive approximations to the root of eq. (6.1) using linear interpolation. This is converging very slowly and changes by around 0.0003 each iteration.

n	A_n	$f(A_n)$	B_n	$f(B_n)$	x_n	$f(x_n)$
1	1.5	-3.5315	1.8	0.99968	1.50003	-3.5310
2	1.50003	-3.5310	1.8	0.99968	1.50006	-3.53082
3						
4						
5						

We can see from this example that the linear interpolation method is much slower than rearrangement method, this will not always be true, but is your first hint that you should play around with several approaches before settling on the one that works the best. If you are writing code for these problems, then it is a good idea to have code that can implement any of the iteration schemes that we are discussing here.

Bisection method

Another method for finding roots is the Bisection method. Consider a root a of a function $f(x)$. The idea here is to guess two points, one on either side of the root a , and then keep shrinking the size of the interval until we have an estimate of the value of the root. Since the function is zero at the root, we can check that we have actually picked points on either side of the root by checking that the function has different signs on either side of the interval.

Example 6.2. Consider the function $f(x) = x^4 + 2x - 2$, we start by guessing that the root is contained in the interval $[1/2, 1]$. We can check this by computing that $f(1/2) = -0.9375$ while $f(1) = 1$. Since the function changes sign across the interval, we know that it has a zero in the interval. We call the left end of the interval a_0 and the right end b_0 , and for each step of refining the interval we will have an a_r and a b_r . To narrow down the position of the root we find the mid point of the interval and evaluate the function there:

$$c_1 = \frac{a_1 + b_1}{2} = \frac{1/2 + 1}{2} = \frac{3}{4} = 0.75.$$

and $f(0.75) = -0.1836$. Since this is still negative we then know that the root lies in the interval $[0.75, 1]$ and again we can find the new mid point:

$$c_2 = \frac{a_2 + b_2}{2} = \frac{0.75 + 1}{2} = 0.875, .$$

with $f(0.875) = 0.3362$. Since this is positive we can again restrict the interval to $[0.75, 0.875]$. We can keep repeating this until we get the desired precision. Though it is much easier to do this in a table, as seen in table 6.3

If we want an answer to two decimal places we find that the root is at $x = 0.80$. If we kept going we would find that to four decimal places the root is at $x = 0.7976$.

Exercise 6.3. The function $f(x) = x^3 + 3x - 5$ has a root between 1 and 2, use the method of bisection four times to find the root.

Newton's method

The main numerical method that is needed in this module is **Newton's method**, sometimes called the **Newton-Raphson** method. Newton's method finds the root by constructing the tangent to a curve through

Table 6.3: Successive approximations to the root of $x^4 + 2x - 2$ using the bisection method.

r	a_r	sign of $f(a_r)$	b_r	sign of $f(b_r)$	c_r	sign of $f(c_r)$
1	0.500000	—	1.000000	+	0.750000	—
2	0.750000	—	1.000000	+	0.875000	+
3	0.750000	—	0.875000	+	0.812500	+
4	0.750000	—	0.812500	+	0.781250	—
5	0.781250	—	0.812500	+	0.796875	—
6	0.796875	—	0.812500	+	0.804688	+
7	0.796875	—	0.804688	+	0.800781	+
8	0.796875	—	0.800781	+	0.798828	+

the initial point x_0 , and then taking the next point x_1 to be where the tangent line intersects the x -axis.

Table 6.4: Successive approximations to the root of eq. (6.1) using Newton's method. For this example, Newton's method converges much quicker than the other methods.

n	x_n	$f(x_n)$
0	1.8	0.99968
1	1.760530638	0.06382986248
2	1.757647406	0.00032122929
3	1.757632748	$8.267167839 \times 10^{-9}$
4	1.757632748	$-8.881784197 \times 10^{-16}$
5	1.757632748	$-8.881784197 \times 10^{-16}$

In other words we start from a guess, x_0 , then construct the tangent line

$$y(x) = (x - x_0) f'(x_0) + f(x_0).$$

Once we have the tangent we find the value of x such that $y(x) = 0$ and call this x_1 . Doing this for a general point in the iteration x_n the iteration scheme is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \quad (6.4)$$

An important observation here is that if the points x_n get close to a critical point of $f(x)$ then the scheme will break down as then $f'(x_n)$ will be approaching zero. This means that we need to be very careful about picking our starting point, if we pick one that is on the opposite side of a critical point from a root then we may not be able to get to the root. Figure 6.3 shows what the first step in Newton's method looks like graphically.

For the specific equation in eq. (6.1) we have that

$$x_{n+1} = x_n - \frac{x_n^5 - 3x_n^3 + 2x_n - 4}{5x_n^4 - 9x_n^2 + 2}. \quad (6.5)$$

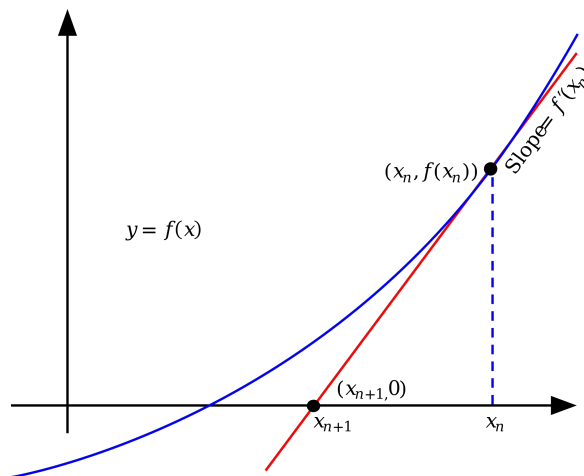


Figure 6.3: The first step of Newton's iteration method to find the root of $f(x)$. We pick a point x_0 then find the tangent to $f(x)$ at x_0 and solve for the point x_1 where this tangent intersects the x -axis. This is repeated until we converge on the root.

Starting from $x_0 = 1.8$ we get the sequence in table 6.4. This iteration scheme converges to the precise answer of eq. (6.2) much quicker than the other methods.

Newton's method is the approach that you need to be able to use during this module so we will treat a couple more examples using it. There are a couple more examples on [Dawkins, 2025b] if you want to see more.

Example 6.4. Use Newton's method to determine an approximation to the solution to $\sin x = -x$ that lie in the interval $[-1, 1]$. Find the approximation to 6 decimal places.

First we need to pick a starting point x_0 . Here we will take $x_0 = 0.5$, you can get an idea of what value to pick by plotting the function. For this example we have that

$$f(x) = \sin(x) + x = 0, \quad (6.6)$$

so the iteration equation for Newton's method eq. (6.4) becomes

$$x_{n+1} = x_n - \frac{\sin x_n + x_n}{\cos x_n + 1}.$$

The first approximation is then

$$x_1 = \frac{1}{2} - \frac{\sin \frac{1}{2} + \frac{1}{2}}{\cos \frac{1}{2} + 1} = -0.0216418$$

Table 6.5: Successive approximations to the root of eq. (6.6) using Newton's method.

n	x_n	$f(x_n)$
0	0.5	0.979426
1	-0.021642	-0.043282
2	2×10^{-6}	$3. \times 10^{-6}$
3	0	0

The further iteration steps are shown in table 6.5, and we see that to 6 decimal places, it only takes 4 iterative steps to find that the root is at $x = 0$. If we did any more steps we would see that these remain at $x = 0$.

Now we can consider one of the examples from [Dawkins, 2025b] which shows when Newton's method does not work.

Example 6.5. Starting from $x_0 = 1$ we will apply Newton's method to $\sqrt[3]{x}$.

Intuitively it is clear that the root is $x = 0$. However, Newton's method will not find this root. For this example eq. (6.4) becomes

$$x_{n+1} = x_n - \frac{\sqrt[3]{x}}{\frac{1}{3}x^{-\frac{2}{3}}} = x_n - 3x_n = -2x_n.$$

This already tells us that instead of converging to $x = 0$, Newton's method will diverge with $x_1 = -2$, $x_2 = 4$, $x_3 = -8$, $x_4 = 16$, \dots . This is the opposite of what we want. Fortunately, it became obvious quite quickly that the method was not working and we did not have to waste much time before discovering that we needed to use a different model.

Exercise 6.6. Write a computer program to implement Newton's method and use it to find the roots in the examples above, where Newton's method works.

Secant method

A variant of Newton's method is the **Secant** method, where we start with two estimates of the root, x_0 and x_1 and estimate the derivative as being the Newton quotient

$$Q(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0},$$

so that

$$x_2 = x_1 - \frac{f(x_1)}{Q(x_0, x_1)} = x_1 - \frac{f(x_1)}{\frac{f(x_1) - f(x_0)}{x_1 - x_0}}.$$

This means that the iteration scheme is given by

$$x_{n+1} = x_n - \frac{f(x_n)}{Q(x_{n-1}, x_n)} = x_n - \frac{f(x_n)}{\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}} \quad (6.7)$$

Geometrically, the secant method can also be viewed as constructing a line nearby our function and moving to the point where this line cuts the x -axis. However, this line is no longer a tangent line, but is a secant.

Example 6.7. For

$$f(x) = x^2 - 10,$$

and taking $x_0 = 2$, $x_1 = 3$ we find x_2 using the secant method as follows:

$$\begin{aligned} x_2 &= x_1 - \frac{f(x_1)}{\frac{f(x_1) - f(x_0)}{x_1 - x_0}} \\ &= 3 - \frac{-1}{\frac{-1 - (-6)}{3 - 2}} \\ &= 3 + \frac{1}{7} \\ &= 3.14286. \end{aligned}$$

This is an equation that we can solve directly by the rearrangement method since $f(x) = 0$ corresponds to $x^2 = 10$ so the exact answers are $x = -\sqrt{10} \simeq -3.16228$ and $x = \sqrt{10} \simeq 3.16228$. We can see that x_2 is already approaching this exact answer.

Convergence and errors

The absolute error of a numerical method is the difference between the true solution x and the approximate solution ξ , $\epsilon = |x - \xi|$. It can be hard to estimate the absolute error unless we already know the exact solution.

We say that an iteration scheme converges when it subsequent iterations do not change the value of x_n , at least to the accuracy that we are looking for. Not every method will converge for a given problem, We have already seen that Newton's method does not converge for the cube root, $\sqrt[3]{x}$. You will gain experience of which methods works best for which type of problem. In this module, you will mainly be using Newton's method so you need to remember that it can fail.

6.2 Approximating functions

Through much of this module we have been discussing functions and how to work with them. Now we will look at a method for approximating functions using polynomials. This is one of the reasons that we spent so long discussing polynomials at the start. If we can understand polynomials then we can get a good idea of how many functions behave.

In essence this function is all about fitting data points, or establishing if there is a relationship or trend between known data. We are not framing it like this here but if you go on to take any statistics or data analysis modules then the approximation ideas here will prove useful to you.

If we have time, in chapter 8 we will discuss the continuous analogue of this which goes by the name of the **Taylor** or **Maclaurin** series. However, to understand these properly, we need to know how to calculate multiple derivative, which is just at the edge of what is included in this module. In this section we will be doing this using two discrete methods, one known as the **Newton interpolating polynomial**, or **Newton's divided difference method**, and the other **Lagrange's method**.

Newton's method

The idea is that idea is that given some values of a function, we can find a polynomial which passes through them. The example given in [Lissamen and West, 2004] is to consider the growth rate per year for a group of young people. The results are given in table 6.6 and can be plotted as five points on a graph. The question asked, and answered, in [Lissamen and West, 2004] is can a formula be found which fits all of these points. The formula given in [Lissamen and West, 2004] is

$$f(x) = 4.13984375 + 3.34791667x - 0.91315104x^2 + 0.08177083x^3 - 0.00232747x^4.$$

The intuition behind this is that we first guess that the function has the form of a straight line $f(x) = mx + c$, notice that this does not fit the data, then add higher order monomials such as x^2, x^3, x^4 with appropriate sized coefficients until we get the correct shape and the curve passes through, or close to, the points.

By the end of this section you will know how to calculate the approximate function which fits the given data.

Table 6.6: Measurements of the average height increase per year at various ages taken from [Lissamen and West, 2004].

Age in years	2	6	10	14	18
Rate of Growth in cm per year	7.8	6.0	4.8	7.0	1.1

Following [Lissamen and West, 2004], we will use the notation in table 6.7. We start our labelling from the left most point and moving rightwards until we get to the last point. Here we are taking x values that have a

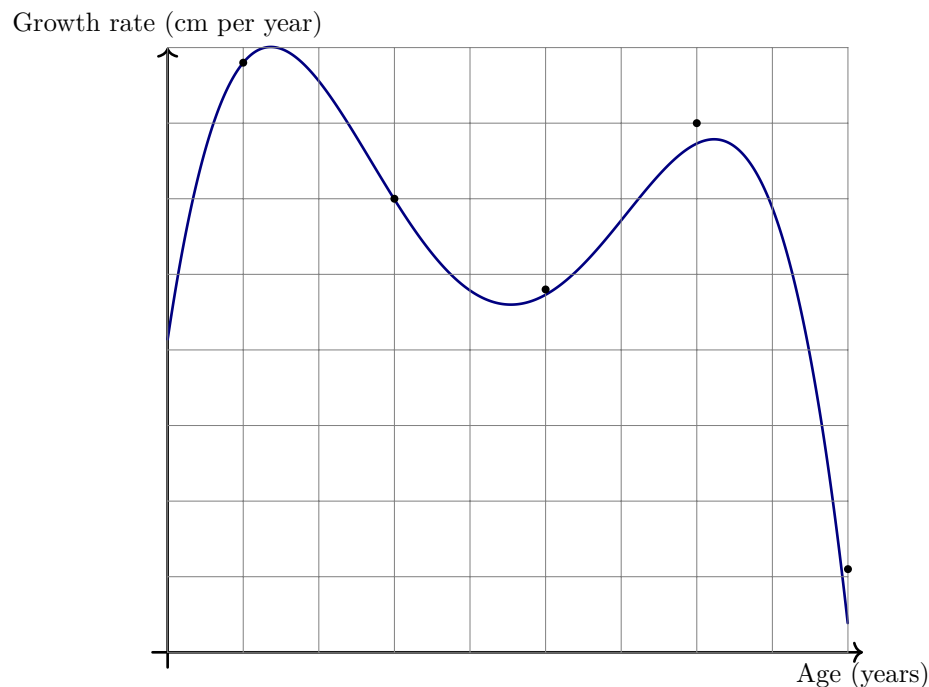


Figure 6.4: A plot of the function which fits the heights from table 6.6 with an approximate polynomial fitted. The polynomial does not quite go through all of the points but passes close to all of them, this is an artefact of the plot being scaled, This will be fixed soon.

constant spacing which we denote by h^3 .

Table 6.7: Our standard notation for putting a list of n points and the values the true function takes at these points into a table.

x	x_0	x_1	x_2	x_3	x_4	\dots	x_n
$f(x)$	$f(x_0)$	$f(x_1)$	$f(x_2)$	$f(x_3)$	$f(x_4)$	\dots	$f(x_n)$

One approach to finding the polynomial is via a finite difference table. This is a means of keeping track of the points, the values of the function at the points, the differences between the values of the function at neighbouring points, the differences between the differences, and so on.

As usual we take Δ to mean the change in some quantity. Here it is technically being used to denote the **forward difference operator**,

$$\Delta f_0 = f_1 - f_0, \quad \Delta f_1 = f_2 - f_1, \quad \dots, \Delta f_{n-1} = f_n - f_{n-1}. \quad (6.8)$$

This is called the forward difference because it is defined as $f_1 - f_0$, e.g we subtract the value of the function at the point on the left from the value at the point on the right. There is a corresponding **backward difference operator** coming from considering $f_0 - f_1$. We will not consider backwards differences here, but you can come up with an alternative method using them. If you go into numerical analysis, or solving real world problems by writing computer programmes then you will need to be familiar with both methods.

³If this seems suggestive of our notation when doing differentiation from first principles, that is on purpose. The separation between points is precisely what we want to take to zero to get the derivative.

The differences between differences are then denoted

$$\Delta^2 f_0 = \Delta f_1 - \Delta f_0, \quad \Delta^2 f_1 = \Delta f_2 - \Delta f_1, \dots \quad (6.9)$$

If we wanted to we can keep going with differences of differences of differences etc. How far you go will depend on the specific problem, as well as how precise you want your approximation to be.

A convenient way to keep track of everything is to put it all in a table as in table 6.8 which has four points and works up to third order differences $\Delta^3 f_i$. If we wanted to include more differences then we would need to add more columns.

Table 6.8: A finite difference table with four points working up to the third differences.

x_i	f_i	Δf_i	$\Delta^2 f_i$	$\Delta^3 f_i$
x_0	f_1			
		$\Delta f_0 = f_1 - f_0$		
x_1	f_1		$\Delta^2 f_0 = \Delta f_1 - \Delta f_0$	
		$\Delta f_1 = f_2 - f_1$		$\Delta^3 f_0 = \Delta^2 f_1 - \Delta^2 f_0$
x_2	f_2		$\Delta^2 f_1 = \Delta f_2 - \Delta f_1$	
		$\Delta f_2 = f_3 - f_2$		
x_3	f_3			

Example 6.8. If we consider the data in table 6.6 and work to fourth order differences we get the table in table 6.9. This finite difference table is what we can use to construct the approximate function.

Table 6.9: A finite difference table with four points working up to the third differences for the data in table 6.6.

x_i	f_i	Δf_i	$\Delta^2 f_i$	$\Delta^3 f_i$
2	7.8			
		$6 - 7.8 = -1.8$		
4	6		$0 - (-1.8) = 1.8$	
		$6 - 6 = 0$		$0.5 - 1.8 = -1.3$
6	6		$0.5 - 0 = 0.5$	
		$6.5 - 6 = 0.5$		
8	6.5			

Example 6.9. Consider the quadratic function $f(x) = x^2 + 1$, working up to fourth order differences, this gives the finite difference table in table 6.10. Notice that the third and fourth differences all vanish.

Exercise 6.10. Construct the finite difference table, up to third order differences, for the linear function $f(x) = 2x + 1$.

Exercise 6.11. Construct the finite difference table, up to third order differences, for a cubic function of your choice. How many differences do you think you need to include?

It is a general property of polynomials that for a polynomial with highest order term x^n , the n 'th differences will be constant and all higher order differences will vanish.

Table 6.10: A finite difference table with four points working up to the fourth order differences for the function $f(x) = x^2 + 1$.

x_i	f_i	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$
0	1				
		1			
1	2		2		
		3		0	
2	5		2		0
		5		0	
3	10		2		
		7			
4	17				

The polynomial built from the finite difference table is known as the interpolating polynomial. This gives the interpolating, or approximating function as

$$f(x) = f_0 + \frac{x - x_0}{h} \Delta f_0 + \frac{(x - x_0)(x - x_1)}{2!h^2} \Delta^2 f_0 + \frac{(x - x_0)(x - x_1)(x - x_2)}{3!h^3} \Delta^3 f_0 + \dots, \quad (6.10)$$

with h the constant spacing between the x values.

This may look complicated, but there is a set pattern to the terms, and the sum will terminate, either when all of the $\Delta^a f_0 = 0$ or because of the number of data points we are given.

When we only have two points this method is called **linear interpolation** as only the first two terms in eq. (6.10) are non-zero.

Example 6.12. Consider the finite difference table in table 6.10. The corresponding Newton interpolating polynomial is found as follows:

- $h = 1$,
- $\Delta f_0 = 1$,
- $\Delta^2 f = 2$,
- $\Delta^3 f = 0$,

and we know what the x_i are, so we can substitute everything in to find:

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

This is the polynomial that we started with which is a good sign.

Exercise 6.13. Consider the finite difference table in table 6.9, and find the Newton interpolation polynomial for this data.

Table 6.11: A table of data that you need to find the Newton polynomial for.

x	2	3	4	5
$f(x)$	1	1	2	2

Exercise 6.14. Consider the data in table 6.11 and find the polynomial that fits the data.

Then use your approximation to find the value of $f(4.3)$.

Note that if all of the differences of a given order are approximately constant then we can sometimes get away with a lower order polynomial approximation, even if the higher order terms do not actually vanish.

Lagrange's method

There is an alternative method due to Lagrange, where we do not need to start with equally spaced points. The easiest way to understand this method is to start by considering the linear case, then the quadratic case, before giving the general expression. This is also discussed briefly in [Lissamen and West, 2004].

The linear case of Lagrange's equation is very similar to eq. (6.3) that we met when discussing linear interpolation as a method of root finding.

Given two points (a, A) and (b, B) , the straight line that passes through them is given by

$$f(x) = \frac{A(x-b)}{a-b} + \frac{B(x-a)}{b-a}$$

where we need that $a \neq b$.

If we have three points, $(a, A), (b, B), (c, C)$ then they lie on a quadratic given by

$$f(x) = \frac{A(x-b)(x-c)}{(a-b)(a-c)} + \frac{B(x-c)(x-a)}{(b-c)(b-a)} + \frac{C(x-a)(x-b)}{(c-a)(c-b)},$$

where we now need that a, b, c are non equal.

This approach can be generalised to a polynomial going through n points. This will be a polynomial of degree $n-1$ for n points.

Exercise 6.15. Find the expression for the degree 3 polynomial going through four points, then use the pattern that you can see between the line, quadratic, and cubic, cases to find the general expression for the polynomial passing through n points.

6.3 Numerical integration

This section needs more images added. There are graphs for the mid point rule case but not for the trapezium rule case. These will be added soon.

We now turn to numerical methods of integration. We have already carried out some numerical integration when we discussed definite integrals and the relationship between integrals and sums. In fact the first method we will use to carry out numerical integration, the **midpoint rule**, is one we have already seen. We will also discuss two other related rules that estimate the area under a graph, the **trapezium rule** and **Simpson's rule**. These are not independent methods, and Simpson's rule is actually a weighted average of the mid point

and trapezium rule which improves the accuracy.

In practice, when you ask a computer to carry out a definite integration, unless you use a symbolic computation method in **Maple**, **Mathematica**, or similar, the computer will be implementing one of these rules. This means that by understanding how they work and how to implement them by hand, you will be able to code up your own numerical integrator. I recommend this as a challenge to keen students who want to make practical use of what you are learning in this module and I will put an exercise at the end of each section suggesting that you do this for each of the rules.

Midpoint rule

The mid point rule is the approximation approach that we met in chapter 4 when approximating the area under the curve for the function $f(x) = x^2 + 1$. Let's recap it here.

Consider the figure in fig. 6.5, as we saw earlier we will be approximating the area between the curve and the x -axis, using rectangles of width Δx and height Δy .

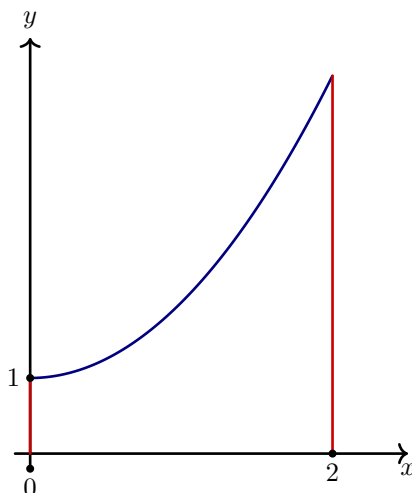


Figure 6.5: A plot of our favourite function $f(x) = x^2 + 1$.

As before, we do this by splitting up the interval, $[0, 2]$ in this case, into n subintervals. This will give us an approximation of the area which will improve the more subintervals we take, in other words the larger that n is. The width of a subinterval is then

$$\Delta x = \frac{2 - 0}{n} = \frac{2}{n}.$$

In general it will be $\Delta x = (b - a)/n$, when the interval is $[a, b]$.

Before we said that we could pick the height of the intervals to be any value the function takes within the subinterval. As the numerical approach that we are following is called the midpoint rule, unsurprisingly, we take the height to be the value of the function at the midpoint of the subinterval. Recall that for an interval $[a, b]$ the midpoint is $c = (a + b)/2$. This method is shown in ??.

As we are taking four subintervals they are

$$\left[0, \frac{1}{2}\right], \quad \left[\frac{1}{2}, 1\right], \quad \left[1, \frac{3}{2}\right], \quad \left[\frac{3}{2}, 2\right],$$

which all have width $\Delta x = 1/2$. The heights are

$$f\left(\frac{1}{4}\right) = \frac{17}{16}, \quad f\left(\frac{3}{4}\right) = \frac{25}{16}, \quad f\left(\frac{5}{4}\right) = \frac{41}{16}, \quad f\left(\frac{7}{4}\right) = \frac{65}{16}.$$

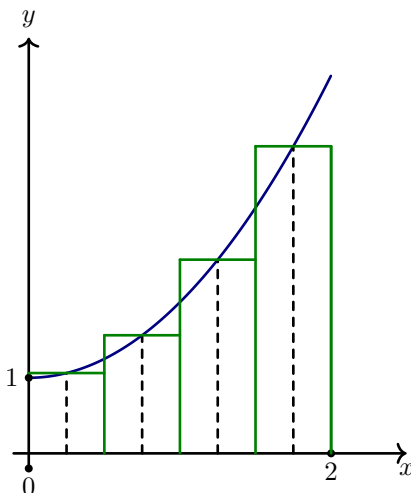


Figure 6.6: A plot of $f(x) = x^2 + 1$ with rectangles approximating the area under the curve.

The area of each rectangle is then

$$\begin{aligned} A_1 &= \Delta x \times f\left(\frac{1}{4}\right) = \frac{1}{2} \times \frac{17}{16} = \frac{17}{32}, \\ A_2 &= \Delta x \times f\left(\frac{3}{4}\right) = \frac{1}{2} \times \frac{25}{16} = \frac{25}{32}, \\ A_3 &= \Delta x \times f\left(\frac{5}{4}\right) = \frac{1}{2} \times \frac{41}{16} = \frac{41}{32}, \\ A_4 &= \Delta x \times f\left(\frac{7}{4}\right) = \frac{1}{2} \times \frac{65}{16} = \frac{65}{32}. \end{aligned}$$

The total area, A_m since we are using the midpoint rule, is thus

$$A_m = A_1 + A_2 + A_3 + A_4 = \frac{17}{32} + \frac{25}{32} + \frac{41}{32} + \frac{65}{32} = \frac{148}{32} = \frac{37}{8} \simeq 4.625.$$

This was a result that we quoted in chapter 4, but did not give all the details for. We also commented on the fact that this was pretty close to the true value of $A = 14/3 = 4.67$ found by carrying out the definite integral.

This shows that we do not always need to use that many sub intervals to get a decent estimate of the integral using the midpoint rule.

We have given one example here, and you have checked at least one more on the problem sheets. However, this idea is general.

Given a function $f(x)$ and an interval $[a, b]$ you should use the following steps:

- pick a number of sub intervals, or use the number given in the question, n , and split your interval into n sub intervals $[a_0, a_1], [a_1, a_2], \dots, [a_{n-1}, a_n]$, of fixed width $\Delta x = (b - a)/n$.
- find the mid point of each sub interval $c_i = (a_i + a_{i-1})/2$ and take the height of the rectangle to be the value of the function at this point, $\Delta y_i = f(c_i)$.
- Calculate the area of each of the rectangles as $A_i = \Delta x \times \Delta y_i$.

- Sum up all of the areas of the rectangles and get your estimate of the total area

$$A_m = A_1 + \dots + A_n = \sum_{i=1}^n A_i.$$

and use this as your approximation to the definite integral.

In essence we are treating the problem the opposite way round to in chapter 4, where we used this approximate picture and discussed how this became the definite integral in the limit of an infinite number of very thin rectangles. Here we are saying that we can use a finite number of rectangles to estimate the definite integral.

Note that in many of the cases you are asked about here you can also compute the integrals directly. This enables you to check how close your answer is to the true answer. The real power of this method is that you can approximate the value of integrals that cannot be calculated directly, or which are very hard to calculate directly.

Example 6.16. Consider the definite integral

$$I = \int_0^1 \sqrt{x^3 + 1} dx,$$

we can evaluate this with the midpoint rule for $n = 5$ strips as follows. Since $n = 5$ the width of each strip is $\Delta x = (1 - 0)/5 = 1/5 = 0.2$.

Exercise 6.17. Consider the function $f(x) = x^2 - 2$ on the interval $[0, 2]$. Take $n = 8$ and use the midpoint rule to approximate the definite integral

$$I = \int_0^2 (x^2 - 2) dx.$$

Exercise 6.18. Consider the function $f(x) = x^3 + x^2 - x - 1$ on the interval $[-1, 1]$. Take $n = 6$ and use the midpoint rule to approximate the definite integral

$$I = \int_{-1}^1 (x^3 + x^2 - x - 1) dx.$$

How does this compare to calculating the definite integral directly?

Exercise 6.19. Consider the function $f(x) = \sin(x)$ on the interval $[0, \pi]$. Take $n = 8$ and use the midpoint rule to approximate the definite integral

$$I = \int_0^\pi \sin(x) dx.$$

Trapezium rule

The next rule is known as the trapezium rule. It is very similar to the midpoint rule in that we are approximating the area under the curve using shapes that we know how to calculate the area of. However, here we use trapezia instead of rectangles.

A trapezium is like a rectangle glued onto a triangle, in general it is a four sided shape where at least one pair of sides are parallel. So squares and rectangles are examples of trapezia.

Consider a function $f(x)$ on the interval $[a, b]$. If the length of the left edge of the trapezium is $f(a)$ and the length of the right edge is $f(a + \Delta x)$, while its width is Δx , then the area is

$$A_{\text{trap}} = \frac{\Delta x}{2} (f(a) + f(a + \Delta x)).$$

If we have an interval $[a, b]$ that we split into n sub intervals of width $\Delta x = (b - a)/n$ as before, then we use n -trapezia, one for each sub intervals. The side lengths of the trapezia are:

- $f_0 = f(a)$ for the left most trapezium,
- $f_1 = f(a + \Delta x)$ for the right side of the first sub interval and the left side of the second,
- $f_2 = f(a + 2\Delta x)$ for the right side of the second sub interval and the left side of the third, and so on up to
- $f_n = f(a + n\Delta x) = f(b)$ for the right side of the n 'th sub interval.

The area of the trapezia are thus

$$\begin{aligned} A_1 &= \frac{\Delta x}{2} (f_0 + f_1), \\ A_2 &= \frac{\Delta x}{2} (f_1 + f_2), \\ &\dots\dots \\ A_n &= \frac{\Delta x}{2} (f_{n-1} + f_n). \end{aligned}$$

Summing these up gives

$$\begin{aligned} A_T &= A_1 + A_2 + \dots + A_n \\ &= \frac{\Delta x}{2} (f_0 + f_1) + \frac{\Delta x}{2} (f_1 + f_2) + \dots + \frac{\Delta x}{2} (f_{n-1} + f_n) \\ &= \frac{\Delta x}{2} (f_0 + 2(f_1 + f_2 + \dots + f_{n-1}) + f_n). \end{aligned} \tag{6.11}$$

Thus to estimate the integral using the trapezium rule we just need to know the value of the function at the end points of the intervals, and the width of the intervals.

Example 6.20. Consider $f(x) = \cos(x)$ over the interval $[0, \pi/2]$. In [Lissamen and West, 2004], they discuss this example for 1, 2, 4 and 8 sub intervals, here we will just treat the case of $n = 4$, so $\Delta x = \pi/8$. The easiest way to keep track of the information we need is in a table of points, the boundaries of the interval, and the value of the function at these points. This table is shown in table 6.12. taking these five values of $f(x)$ and substituting into eq. (6.11) we get

Table 6.12: A table of data that you need to implement the trapezium rule.

x	0	$\frac{\pi}{8}$	$\frac{\pi}{4}$	$\frac{3\pi}{8}$	$\frac{\pi}{2}$
$f(x)$	1	0.92	$\frac{1}{\sqrt{2}}$	0.38	0

$$\begin{aligned} A_T &= \frac{\Delta x}{2} (f_0 + 2(f_1 + f_2 + f_3) + f_4) \\ &= \frac{\pi}{16} \left(1 + 2(0.92 + \frac{1}{\sqrt{2}} + 0.38) + 0 \right) \\ &= 0.98, \end{aligned}$$

which is close to the true value of 1.

Note that both the trapezium rule and the midpoint rule can be used to find the area under a curve even if we do not know the form of the function.

Example 6.21. Consider that we are given the data in table 6.13, using just this information we can find

$$\int_1^3 f(x) dx,$$

Table 6.13: A table of data that you need to implement the trapezium rule.

x	1	1.5	2	2.5	3
$f(x)$	0	0.41	0.69	0.92	1.10

without knowing the form of $f(x)$ using the trapezium rule.

This works as follows. We have $n = 4$ sub intervals defined by the five values of x we are given. Thus we can use eq. (6.11), and $\Delta x = (3 - 1)/4 = 1/2$ to find

$$\begin{aligned} A_T &= \frac{\Delta x}{2} (f_0 + 2(f_1 + f_2 + f_3) + f_4) \\ &= \frac{1}{4} (0 + 2(0.41 + 0.69 + 0.92) + 1.10) \\ &= 1.29. \end{aligned}$$

Note that we could use Newton's divided difference method to find an approximation to this function, but we do not need to do this to calculate the integral.

Exercise 6.22. Use the trapezium rule with 6 sub intervals to calculate

$$I = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin(x) dx.$$

You can do this either by hand, or set up a spread sheet or computer programme to solve this for you.

Exercise 6.23. Use the trapezium rule with 4 steps to approximate the definite integral

$$I = \int_0^2 (x^2 + 1) dx$$

and compare it to the answer that we got using the mid point rule in chapter 4.

Exercise 6.24. Write a computer programme to implement the trapezium rule and use it to check the examples and exercises in this section.

Simpson's rule

The final numerical method that we will discuss is Simpson's rule. This is a weighted average of the trapezium rule and the mid point rule. We could actually consider any weighted average of the the two methods to get an approximation to the definite integral. However, if we were analysing the error we would see that there is one particular choice which minimises the error and gets as close as possible to the true value of the integral. This is Simpson's rule,

$$A_S = \frac{2A_m + A_T}{3}. \quad (6.12)$$

Note that here we would use the same number of intervals for both the midpoint rule and the trapezium rule.

Example 6.25. Let's return to one of our favourite examples

$$I = \int_0^{\frac{\pi}{2}} \cos(x) dx$$

and calculate this for $n = 4$ sub intervals using Simpson's rule.

The data to calculate this is given in table 6.12, though we need to find the value of the function at the midpoints for the midpoint rule. We already know that $A_T = 0.98$ so we just need to find A_m . The four mid points and the value of the function at them are:

- $c_0 = (\pi/8 + 0)/2 = \pi/16$ and $\cos(c_0) = 0.98$
- $c_1 = (\pi/4 + \pi/8)/2 = 3\pi/16$ and $\cos(c_1) = 0.83$,
- $c_2 = (3\pi/8 + \pi/4)/2 = 5\pi/16$ and $\cos(c_2) = 0.556$
- $c_3 = (\pi/2 + 3\pi/8)/2 = 7\pi/16$ and $\cos(c_3) = 0.195$

We thus have that

$$\begin{aligned} A_m &= \Delta x (f(c_0) + f(c_1) + f(c_2) + f(c_3)) \\ &= \frac{\pi}{8} (0.98 + 0.83 + 0.556 + 0.195) \\ &= 1.0057. \end{aligned}$$

Simpson's rule then gives that

$$A_S = \frac{2A_m + A_T}{3} = 0.997 \simeq 1$$

where we have rounded to two decimal places.

Exercise 6.26. Go through the case from the example above,

$$I = \int_0^{\frac{\pi}{2}} \cos(x) dx,$$

and get the answer to 9 decimal places.

Exercise 6.27. Use Simpson's rule with $n = 6$ sub intervals to calculate

$$I = \int_0^2 \sqrt{1 + \sin(x) + \cos(x)} dx,$$

Give your answer to 6 decimal places.

6.4 Numerical differentiation

We now turn to numerical approaches to differentiation. We have seen hints of this in the background when discussing other topics, including as far back as when we discussed differentiation from first principles.

A major reason that we need a numerical approach to differentiation is that the definition of differentiation from first principles is very hard to implement on a computer. It is much easier for a computer to understand an approximate method.

Forward difference method

The key idea is to look at the definition of differentiation from first principles in eq. (3.1), and drop the limit part. e.g. we calculate an approximate gradient for points a small difference apart. In other words,

$$\frac{df}{dx} \simeq \frac{f(x+h) - f(x)}{h} = \frac{\Delta f}{h}, \quad (6.13)$$

for h a small step between neighbouring points. This is known as the **forward difference method**, since we are looking at the difference between the function at the point x and at the point $x+h$ one step ahead of x .

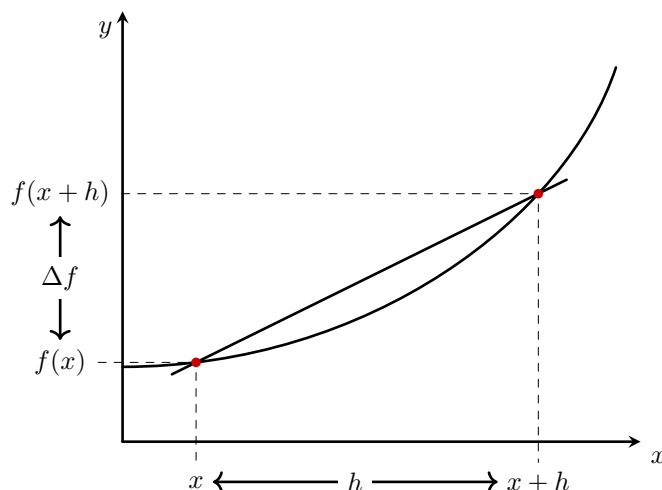


Figure 6.7: A schematic of the forward difference method showing how to approximate the tangent line to the curve at x

This is shown schematically in fig. 6.7.

Note that when we are numerically differentiation we will find the value of the derivative of the function at a specific point.

Example 6.28. Consider the function $f(x) = x^2 \cos(x)$, we can calculate the approximate derivative at $x = 2$ using eq. (6.13) and taking the step size to be $h = 0.1, 0.05, 0.025, 0.0125$. We do this as follows:

- $h = 0.1$

$$\frac{df}{dx}(2) \simeq \frac{f(2 + 0.1) - f(2)}{0.1} = \frac{-2.22637 + 1.66459}{0.1} = -5.61784,$$

- $h = 0.05$

$$\frac{df}{dx}(2) \simeq \frac{f(2 + 0.05) - f(2)}{0.05} = \frac{-1.93766 + 1.66459}{0.05} = -5.46141,$$

- $h = 0.025$

$$\frac{df}{dx}(2) \simeq \frac{f(2 + 0.025) - f(2)}{0.025} = \frac{-1.79914 + 1.66459}{0.025} = -5.38196,$$

$$h = 0.0125$$

$$\frac{df}{dx}(2) \simeq \frac{f(2 + 0.0125) - f(2)}{0.0125} = \frac{-1.73136 + 1.66459}{0.0125} = -5.34196,$$

The true value found by calculating the derivative is

$$\begin{aligned} \frac{df}{dx}(x) &= 2x \cos(x) - x^2 \sin(x), \\ \frac{df}{dx}(2) &= -5.30178, \end{aligned}$$

so the approximating is getting closer to this as we take a smaller step size.

Exercise 6.29. Use the forward difference method with $h = 0.05$ to approximate the derivative of $f(x) = 4e^{2x}$ at $x = 1$.

The forward difference method is not the only approach to approximating the gradient of the tangent line at a point. There are two other popular approaches, the **central difference method** and the **backwards difference method**.

Backwards difference method

The backwards difference method is the *reverse* of the forward difference method. The picture to have in mind, shown in fig. 6.8, is essentially the same as in fig. 6.7. The difference is that now we are comparing the value of the function at point with its value at the point one step earlier. As an equation this is

$$\frac{df}{dx} \simeq \frac{f(x) - f(x-h)}{h}. \quad (6.14)$$

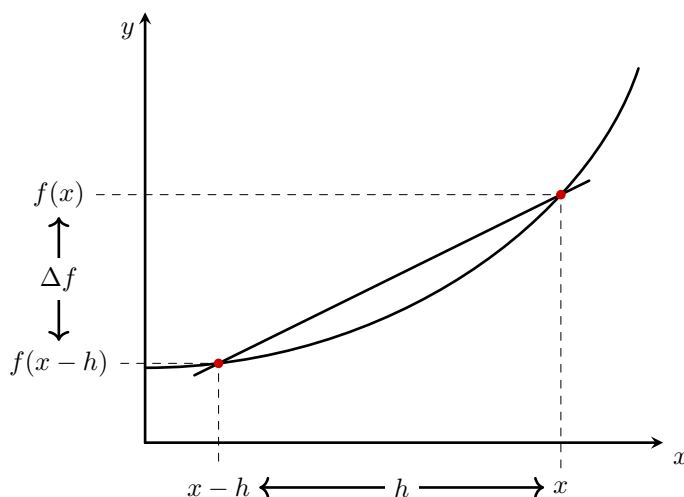


Figure 6.8: A schematic of the backwards difference method showing how to approximate the tangent line to the curve at x

Example 6.30. *Similar to in the previous case, consider the function $f(x) = x^2 \cos(x)$, we can calculate the approximate derivative at $x = 2$ using the backwards difference method and taking the step size to be $h = 0.1, 0.05, 0.025, 0.0125$. We do this as follows:*

- $h = 0.1$

$$\frac{df}{dx}(2) \simeq \frac{f(2) - f(2 - 0.1)}{0.1} = \frac{-1.66459 + 1.16708}{0.1} = -4.97512,$$

- $h = 0.05$

$$\frac{df}{dx}(2) \simeq \frac{f(2) - f(2 - 0.05)}{0.05} = \frac{-1.66459 + 1.40761}{0.05} = -5.13949,$$

- $h = 0.025$

$$\frac{df}{dx}(2) \simeq \frac{f(2) - f(2 - 0.025)}{0.025} = \frac{-1.66459 + 1.53406}{0.025} = -5.22093,$$

$$h = 0.0125$$

$$\frac{df}{dx}(2) \simeq \frac{f(2) - f(2 - 0.0125)}{0.0125} = \frac{-1.66459 + 1.59882}{0.0125} = -5.26143,$$

The true value we found before is

$$\frac{df}{dx}(2) = -5.30178,$$

so again the approximating is getting closer to this as we take a smaller step size, though this time we are underestimating the gradient and approaching it from below.

Central difference method

The central difference method is essentially a combination of the other two methods. In this case we approximate the gradient of the tangent line by using the straight line between the two points $f(x-h)$ and $f(x+h)$. In this case we are approximating the derivative by the gradient,

$$\frac{df}{dx} \simeq \frac{\Delta f}{\Delta x} = \frac{f(x+h) - f(x-h)}{(x+h) - (x-h)} = \frac{f(x+h) - f(x-h)}{2h}. \quad (6.15)$$

The central difference method is often preferred to the forward and backwards difference method, since it is a *higher order method*. This means that when we discuss the error we will see that it is quadratic in the small parameter h while the other methods are linear in h .

Example 6.31. Consider $f(x) = \tan(x)$ and use the central difference method with $h = 0.1, 0.05, 0.025, 0.0125$ to approximate the derivative at $x = 1$. This gives:

- $h = 0.1$

$$\frac{df}{dx}(2) \simeq \frac{f(2+0.1) - f(2-0.1)}{0.2} = 6.08625,$$

- $h = 0.05$

$$\frac{df}{dx}(2) \simeq \frac{f(2+0.05) - f(2-0.05)}{0.1} = 5.84915,$$

- $h = 0.025$

$$\frac{df}{dx}(2) \simeq \frac{f(2+0.025) - f(2-0.025)}{0.05} = 5.7929,$$

$$h = 0.0125$$

$$\frac{df}{dx}(2) \simeq \frac{f(2+0.0125) - f(2-0.0125)}{0.025} = 5.77901,$$

The true value of the derivative is 5.7744 which our approximations are closing in on.

Example 6.32. Use both the forward and central difference methods to approximate the derivative of $f(x) = \cos(x)$ at $x = \pi/4$ using the step sizes $h = 0.1$ and $h = 0.05$. Compare these to each other and to the exact answer.

Example 6.33. Pick your favourite programming language and write a function which implements the central difference method. Check it against the example that we computed above, then check a the cases that we explored using the forward and backwards difference method.

Understanding the error

We briefly mentioned above that the absolute error is the difference between the true value, x , and the value that our approximation returns, X , is given by

$$\text{Absolute error} = |X - x|. \quad (6.16)$$

The $||$ mean that this is the absolute value, and of the number within and is always positive. In other words we care about how far away our approximate value is from the true value, but not whether it is greater than or less than the true answer.

The absolute relative error is then

$$\text{Absolute relative error} = \left| \frac{X - x}{x} \right|, \quad (6.17)$$

and the percentage error is

$$\text{Percentage error} = \left| \frac{X - x}{x} \right| \times 100\% \quad (6.18)$$

There is a discussion of error and the related concept of convergence of an approximation, i.e. whether the approximation is closing in on the true value and how quickly, in [Lissamen and West, 2004].

Example 6.34. Consider the examples of approximating the derivative of $x^2 \cos(x)$ using both the backwards and forwards difference method with $h = 0.0125$. The relative and absolute errors are shown in table 6.14.

Table 6.14: A comparison of our approximations to the derivative of $x^2 \cos(x)$ at $x = 2$ and the percentage error in each case.

	Approximate derivative	Relative absolute error	Percentage error
Forward difference	-5.34196	0.00758	0.758%
Backward difference	-5.26143	0.00761	0.761%

Exercise 6.35. Go back to all of the examples of using the forward, backwards, and central difference equation and calculate the relative and percentage error in each case.

6.5 Numerical approaches to differential equations

This is a non-examinable section and is not taught every year. It is included in the indicative content of the module but is not assessed in the learning outcomes. Content will be added here at the same time as content is added to chapter 5 since that is the theoretical background for what will be discussed numerically here.

Chapter 7

Calculus in Computer Science

Machines take me by surprise with great frequency.

Alan Turing

In most of this module we have focussed on the mathematics, and even when trying to motivate the connection to computer science we have been fairly sketchy. In this chapter we are going to rectify that and discuss some concrete examples of where the material that you have studied here is of relevance to Computer Science, AI, and Machine Learning.

If you try to read this chapter before covering all of the basics you may find the examples hard to follow, but if you have finished everything up to chapter 4, then you should be fine.

This chapter is split into sections, at least one of which should be relevant to each of the degree schemes that this module is included on, **Computer Science**, **Computer Science and AI**, and **Robotics and AI**. This does not mean that you will not find examples of interest in other sections, it just means that I am trying to group the examples thematically. As with several other chapters, this is a work in progress and will be adapted and expanded over time. Be sure to check back frequently if you want to see the most up to date examples.

7.1 LLMs and AI

Large language models (LLMs) are basically¹ big functions and compositions of functions. Think of it like fig. 7.1, where we have input data x which the function sends to an output $f(x)$.

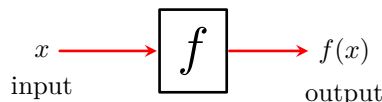


Figure 7.1: A schematic of a LLM as a function where the parameters are tuned to take an input and return a specific output.

Here we focus on the simplest case where there is one input, one output, and the function f has a single parameter. In a real system this will be much more complicated, but then you need to be able to use calculus

¹This is a real simplification. 3Blue1Brown has a playlist on Neural Networks, available here, which explains more about the applications of calculus to LLMs and what is really going on.

in several variables which is beyond this module. If you are interested in knowing more about the general story let me know and I can provide some further resources.

In this picture the function f is a model that tells us how the input is mapped to the output. The parameters in the model tell us how this happens.

Example 7.1. Consider the function which sends x to $f(x) = ax + b$, this has two parameters a and b .

When training a model we use training data where we already know the true output y that we want to get from a given input x . To do this we essentially want to find the “best” values of a and b , or whatever the parameters are, so that our functions output \hat{y} is as close as possible to the true value y .

If we are sticking to a simple one parameter model then given the input x we get out put $\hat{y}(w) = f(w, x)$, where $f(x)$ will be a function like $f(w, x) = x + w, wx, e^{wx}, \sin(wx), \dots$. For the training data we then measure the difference between $\hat{y}(w)$ and the true output y for the fixed input x .

This difference between $\hat{y}(w)$ and y is called an error term, it measures how far out our model is from the true answer. For example $y - \hat{y}(w)$ is a linear error term, while $(y - \hat{y}(w))^2$ is a quadratic error term. If we treat the error term as a function of the parameter w and find its critical points this will tell us the values of the parameter that minimise or maximise the error, and we want to minimise the error term.

This error function is often called the **loss function**. Loss functions appear generally in decision theory and the study of optimisation problems. They appear whenever you are training an AI or LLM, where they measure the deviation of the model’s output against the true value for the training data.

A conventional choice of loss function is the quadratic loss function

$$\lambda(w) = (\hat{y}(w) - y)^2 \quad (7.1)$$

Again the idea that given a function f with a single free parameter w we want to minimise $\lambda(w)$ for fixed input x and fixed true output y . The optimisation method used is called **gradient flow**. In a one dimensional case we are solving

$$\frac{d\lambda}{dw} = 0$$

for w , to get the optimal $\hat{y}(w)$ for a given x .

Real World Application 7.1. Consider the one parameter model $\hat{y}(w) = \sin(wx)$ and the quadratic loss function in eq. (7.1). If the input is $x = \pi$ and the desired output is $y = 0$ then the loss function becomes

$$\lambda(w) = (\sin(w\pi) - 0)^2 = (\sin(w\pi))^2.$$

Then

$$\begin{aligned} \frac{d\lambda}{dw} &= 2\pi \sin(w\pi) \cos(w\pi) \\ &= \pi \sin(2w\pi). \end{aligned}$$

This is zero when $w = \pm 1, \pm 2, \pm 3, \dots$ or $w = \pm 1/2, \pm 3/2, \pm 5/2, \dots$ i.e. w is either an integer or a half-integer.

Looking at the second derivative we see that

$$\frac{d^2\lambda}{dw^2} = 2\pi^2 \cos(2w\pi),$$

and that

$$\begin{aligned} \frac{d^2\lambda}{dw^2}(\pm 1, \pm 2, \dots) &> 0 \\ \frac{d^2\lambda}{dw^2}\left(\pm \frac{1}{2}, \pm \frac{3}{2}, \dots\right) &< 0 \end{aligned}$$

So integer values of w are minima and half integer values are maxima.

For all of these values $f : \pi \mapsto 0$ so you would need to look at more training data to fix a single value of w .

It is important to note that while optimisation can get the output closer to the true value, how close it can get depends on how good the choice of model f is.

Example 7.2. *Consider the model*

$$\hat{y}(w) = \frac{wx}{1 + w^2x^2},$$

alongside the quadratic loss function. If the target output for input $x = 1$ is $y = 2$ what is the optimal value of w ?

For $x = 1$ and $y = 2$ the loss function becomes

$$\hat{y}(w) = \left(\frac{w}{1 + w^2} - 2 \right)^2.$$

Note here that $w/(1 + w^2) < 1/2$ so we can never get $\lambda(w) = 0$, i.e. this model is a bad fit for the training data.

If we optimise anyway then we get

$$\begin{aligned} \frac{d\lambda}{dw} 2 \left(\frac{w}{1 + w^2} - 2 \right) \frac{d}{dw} \left(\frac{w}{1 + w^2} \right) \\ = 2 \left(\frac{w}{1 + w^2} - 2 \right) \frac{(1 - w^2)}{(1 + w^2)^2}. \end{aligned}$$

We know that the first bracket does not vanish so this is zero when $1 - w^2 = 0$ giving $w = \pm 1$. If $w = 1$ then $\lambda(+1) = 9/4$ while $\lambda(-1) = 25/4$ so $w = +1$ is the minimum. But for this choice of w the output is

$$\hat{y}(1) = \frac{1}{1 + 1} = \frac{1}{2} \neq 2$$

So we do not get a match for the desired output.

This example of optimisation not working is a useful warning as it shows us that while optimisation is a powerful and useful model you still need to start from a sensible model.

7.2 Robotics

To be added.

Chapter 8

Advanced Topics

The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift, which we neither understand nor deserve.

Eugene Wigner

This section is non examinable and is included so that you have a feel for how the material in this module can be taken further.

8.1 L'Hôpital's rule for evaluating limits

We saw in chapter 2 that some limits lead to nonsense expressions. For example, If we were confronted with $\lim_{x \rightarrow 0} \sin(x)/x$, this looks like it will result in $0/0$ which is does not make sense. Remember that not all limits that look indeterminate really are. For example

$$\lim_{x \rightarrow 4} \frac{x^2 - 16}{x - 4}$$

looks at first glance like it will have the form $0/0$ since both the numerator and denominator vanish for $x = 4$. However, the numerator can be factored as $x^2 - 16 = (x - 4)(x + 4)$ so the limit simplifies to

$$\lim_{x \rightarrow 4} \frac{x^2 - 16}{x - 4} = \lim_{x \rightarrow 4} \frac{(x - 4)(x + 4)}{x - 4} = \lim_{x \rightarrow 4} (x + 4) = 8.$$

This is why we said that you should try to expand and simplify the function that you are taking the limit of as much as possible.

L'Hôpital¹'s rule enable us to make sense of some of those limits which still look indeterminate after they have been simplified. To apply L'Hôpital's rule, we have to be taking the limit of a ratio of functions $f(x), g(x)$, where both are differentiable, the derivative of $g(x)$ does not vanish, and in the limit $f(x)$ and $g(x)$ either both go to zero or both go to infinity. Then we have that

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}, \quad (8.1)$$

so we can replace the ratio of the functions with the ratio of their derivatives. If both of the derivatives still

¹Originally spelt L'Hospital, with a silent s and no circumflex

vanish, or diverge, in the limit, then process can be repeated and we end up with more derivatives

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f^{(n)}(x)}{g^{(n)}(x)},$$

where the superscript (n) means that we are differentiating the functions n times.

Note that if the limit does not lead to an indeterminate then L'Hôpital's rule does not hold!

Example 8.1. *As a counter example consider the limit*

$$\lim_{x \rightarrow 1} \frac{f(x)}{g(x)} = \lim_{x \rightarrow 1} \frac{x+1}{2x+1}.$$

Both $\lim_{x \rightarrow 1} f(x) = 2$ and $\lim_{x \rightarrow 1} g(x) = 3$ are finite and non-zero. Evaluating the limit directly gives

$$\lim_{x \rightarrow 1} \frac{x+1}{2x+1} = \frac{2}{3}.$$

We can also calculate the limit of the ratio of derivatives,

$$\lim_{x \rightarrow 1} \frac{f'(x)}{g'(x)} = \lim_{x \rightarrow 1} \frac{1}{2} = \frac{1}{2} \neq \frac{2}{3}.$$

So the limit of the ratio does not match the limit of the ratio of derivatives.

As long as our original limit looks like it is indeterminate we can use L'Hôpital's rule.

Example 8.2. *The limit of $\sin(x)/x$ as $x \rightarrow 0$ is evaluated as follows*

$$\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = \lim_{x \rightarrow 0} \frac{\cos(x)}{1} = 1.$$

We can also evaluate other limits that may not at first look like a ratio of functions.

Example 8.3. *Consider the limit $\lim_{x \rightarrow -\infty} xe^x$ this looks like it becomes the indeterminate $(-\infty)(0)$. If we recall that $1/e^x = e^{-x}$ then we can rewrite the limit as*

$$\lim_{x \rightarrow -\infty} xe^x = \lim_{x \rightarrow -\infty} \frac{x}{e^{-x}},$$

which looks like the indeterminate $-\infty/\infty$. Now we can apply L'Hôpital's rule to get

$$\lim_{x \rightarrow -\infty} \frac{x}{e^{-x}} = \lim_{x \rightarrow -\infty} \frac{1}{-e^{-x}} = 0,$$

since $1/\infty$ is zero.

8.2 Trigonometric Derivatives in Degrees

At various points in this module I have said that you should always work with angles in radians as this makes life easier when differentiating trig functions. Here I want to discuss how to take the derivative if we are measuring the angle in degrees.

First we need some notation, we will use the subscript rad when our function has an argument in degrees and deg when our function has an argument in degrees. Then we can assume the result that we were given in chapter 3, but rewritten in this new notation, that

$$\frac{d}{dx} \sin_{\text{rad}}(x) = \cos_{\text{rad}}(x).$$

Now recall that if x is an angle in radians then the corresponding angle, θ in degrees is

$$\theta = \frac{180^\circ}{\pi}x, \quad \text{or } x = \frac{\pi}{180^\circ}\theta.$$

Now we can calculate the derivative of \sin_{deg} using the chain rule

$$\begin{aligned} \frac{d}{d\theta} \sin_{\text{deg}}(\theta) &= \frac{d}{d\theta} \sin_{\text{rad}}\left(\frac{\pi}{180^\circ}\theta\right) \\ &= \frac{dx}{d\theta} \frac{d}{dx} \sin_{\text{rad}}(x) \\ &= \frac{\pi}{180^\circ} \cos_{\text{rad}}(x) \\ &= \frac{\pi}{180^\circ} \cos_{\text{rad}}\left(\frac{\pi}{180^\circ}\theta\right) \\ &= \frac{\pi}{180^\circ} \cos_{\text{deg}}(\theta). \end{aligned}$$

The derivatives of other trig functions can be calculated in similar ways. This is why we have to be careful to use radians as we do not want to have to keep track of factors of $\pi/180^\circ$ every time we differentiate.

There is an interesting discussion of this result and how to think about the difference between trig functions that take arguments in radians and degrees here

8.3 Functions of two variables

8.4 Multiple integrals

8.5 Optimisation Problems

In section 3.4 we discussed finding the critical points of a function by studying where the derivative vanishes. The example that I like to have in mind comes from physics. Consider a ball thrown straight up in the air, how do we find the maximum height that the ball reaches? We look at where the velocity vanishes, since as the ball is moving upwards towards it is being decelerated by gravity so the velocity decreases, then when it is going down after reaching its maximum height the velocity is getting larger but pointing in the opposite direction. For this to happen there has to be a point where the velocity is instantaneously zero. This is at the maximum height. Calculus comes into this as the velocity of an object is the derivative of its displacement, so looking for where the velocity vanishes is exactly the same as finding the critical points where the derivative of the displacement vanishes.

We also discussed that critical points correspond to the minima and maxima (and points of inflection) of a function. This means that if our function describes the area enclosed by a fence, or the time taken to run an algorithm that finding the critical points will tell us the maximum or minimum area that we can enclose, or the fastest and slowest that the algorithm can run. This is what is meant by **Optimisation**, we are optimising a function by finding the parameter values which make its output the largest or the smallest it can be.

Mathematical Diversion 8.1. *Note that the function can be constrained, e.g. if it is restricted a particular set of values that is not the full range of the function, then it may not achieve its global maximum or minimum. In that case you have to evaluate the function at the boundary of the region and see if the values it takes are larger or smaller than that taken at any critical points which satisfy the constraint.*

See section 4.8 of [Dawkins, 2025b] for more on constrained optimisation at the level relevant to this module.

In other words, optimisation problems are all about finding the maxima, minima, and points of inflection for functions, and interpreting the physical meaning of these points. A famous historical example is the **Dido problem**² which asks you to find the maximum area that can be enclosed by a shape with a given fixed perimeter. The solution to the Dido problem is the circle.

Another similar problem is asking what curve minimises the distance between two points. In flat space we all know the answer is a straight line, but have you ever thought about how this changes if the surfaces is curved? It turns out that the solution to the problem changes. This is why aeroplanes fly along sections of a circle, as these give the shortest distance on a globe.

8.6 Polynomial approximation

In chapter 6 we met Newton's divided difference method which enabled us to find a polynomial approximation to a function, making use of a table of values with a uniform step size³. The interpolating polynomial is expressed in terms of the data points, step size, and finite differences as

$$f(x) = f_0 + \frac{x - x_0}{h} \Delta f_0 + \frac{(x - x_0)(x - x_1)}{2!h^2} \Delta^2 f_0 + \dots + \frac{(x - x_0)(x - x_1) \dots (x - x_n)}{n!h^n} \Delta^n f_0. \quad (8.2)$$

The number of terms $n + 1$ is determined by the number of data points we start of with as if we only have one point, then we just get f_0 , two data points gives one first difference $\Delta f_0 = f_1 - f_0$, three data points gives one second difference, $\Delta^2 f_0 = \Delta f_1 - \Delta f_0$. Thus $n + 1$ points gives a series with $n + 1$ terms as shown in eq. (8.2).

The discussion of numerical differentiation in chapter 6 showed us that these finite differences are discrete are discrete versions of the derivative. For example

$$\frac{df}{dx}(a) \simeq \frac{f(a + h) - f(a)}{h} = \frac{\Delta f_a}{h}$$

in the forward difference scheme. It turns out that we can do the same for higher derivatives and see that

$$\frac{d^2 f}{dx^2}(a) \simeq \frac{\Delta^2 f_a}{h^2},$$

or in general

$$\frac{\Delta^n f_a}{h^n} = \frac{d^n f}{dx^n}(a)$$

²named after a queen of Carthage

³We saw that if the step size is not uniform we can still get a similar approximation.

Chapter 9

Background Mathematics

Try as you may you just can't get
away from Mathematics

*That's Mathematics by Tom
Lehrer*

9.1 Background and References

As this is a maths module there is a lot of assumed background. This means that to understand the material in the module and to be able to solve the tutorial problems, you need to have a experience with a variety of mathematical techniques such as:

- solving linear equation,
- solving quadratic equations
- using trigonometry,
- knowing some simple functions.

These topics will be familiar to many of you. However, it may have been a while since some of you studied mathematics and I am aiming to briefly introduce any new mathematical topics when we need them. However, I also want to link to some extra resources where you can brush up on your maths beforehand.

A great resource is the website Pauls Online Math Notes. The website contains notes for a variety of mathematics courses including algebra and calculus. The most useful background materials are the Algebra and trigonometry review linked here and the preliminaries section of the algebra notes, linked here.

As the module goes on I may add more background resources or add some examples.

9.2 Polynomials and roots

One of the most common examples of a function that we will meet is a polynomial. These are functions where the variable x is sent to a sum of powers of x ,

$$p(x) = a_n x^n + a_{n-1} x^{n-1} \cdots + a_1 x + a_0. \quad (9.1)$$

If we plot a polynomial, e.g. produce the graph $(x, y = p(x))$, then the locations where the plot cuts the x -axis, where $y = p(x) = 0$, are called the zeros or roots. For a linear equation we solve it via rearrangement, see section 9.5. For a quadratic equation like

$$ax^2 + bx + c = 0,$$

we solve it using the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (9.2)$$

As an example consider the quadratic equation

$$2x^2 - 4x - 2 = 0.$$

To apply the quadratic formula we compare this specific equation to the general quadratic equation above, and note that $a = 2, b = -4, c = -2$. The quadratic formula thus gives

$$\begin{aligned} x &= \frac{4 \pm \sqrt{4^2 + 4 \times 2 \times 2}}{4} \\ &= \frac{4 \pm \sqrt{16 + 16}}{4} \\ &= 1 \pm \frac{\sqrt{32}}{4} \\ &= 1 \pm \frac{\sqrt{16}\sqrt{2}}{4} \\ &= 1 \pm \sqrt{2}. \end{aligned}$$

So the two roots are $1 + \sqrt{2}, 1 - \sqrt{2}$.

Note that we can guess the roots by looking at the polynomial. A quadratic can always be written as

$$\begin{aligned} a(x - \alpha)(x - \beta) &= a(x^2 - (\alpha + \beta)x + \alpha\beta) \\ &= ax^2 - a(\alpha + \beta)x + a\alpha\beta. \end{aligned}$$

So the product of the roots is related to the constant term, and the sum of the roots is related to the linear term. This means that we can guess one of the roots by looking at the factors of the constant term. For quadratics we do not need to do this, but with higher order equations this is very useful.

For higher order polynomials we do not have a formula to find the solutions¹ In this case we use the higher order versions of the above identity that the constant term in the polynomial is related to the product of the roots. We can then use a technique like polynomial long division to reduce the order of the equation.

Consider the equation

$$x^3 + x^2 - x - 1 = 0$$

Since $(x - a)(x - b)(x - c)$ has a constant term $-abc$ we see that the product of the three roots is 1. This means that we can guess that 1 is one of the roots.

¹The exception is cubic equations. There is a cubic formula. However, it is fairly complicated and there is no point in trying to use it.

We can the perform polynomial long division to find that

$$\begin{array}{r}
 x^2 + 2x + 1 \\
 x - 1 \overline{) \begin{array}{r} x^3 + x^2 - x - 1 \\ - x^3 + x^2 \\ \hline 2x^2 - x \\ - 2x^2 + 2x \\ \hline x - 1 \\ - x + 1 \\ \hline 0 \end{array}}
 \end{array}$$

This means that we can rewrite

$$x^3 + x^2 - x - 1 = (x - 1)(x^2 + 2x + 1),$$

and then we can either factorise the quadratic directly or use the quadratic formula. Note that we could also solve this case using polynomial long division by noting that the product of the roots is 1 so we can guess that -1 is a repeated root and calculate

$$\begin{array}{r}
 x + 1 \\
 x + 1 \overline{) \begin{array}{r} x^2 + 2x + 1 \\ - x^2 - x \\ \hline x + 1 \\ - x - 1 \\ \hline 0 \end{array}}
 \end{array}$$

so we have that

$$x^3 + x^2 - x - 1 = (x - 1)(x + 1)^2.$$

This style of approach can be applied to any polynomial equation. There is a more complete description of polynomials and their roots given in [Dawkins, 2025a].

9.3 Trigonometry Primer

Since trigonometric functions come up a lot in this course a review of the basics is included here. If you are not familiar with it you can either check out the links suggested above or ask me to provide more background information.

Here we will only discuss trigonometry for right angled triangles, but in the main content of the module we will deal with general trig functions. **Trigonometry** is an area of mathematics related to the study of triangles and provides a way to compute the lengths and angles in a triangle provided that you now some of them already.

The typical mnemonic used to remember trigonometry is **SOH CAH TOA** which means Sine is opposite over hypotenuse

$$\sin \theta = \frac{c}{a}, \quad \sin \phi = \frac{b}{a},$$

Cosine is adjacent divided by hypotenuse

$$\cos \theta = \frac{b}{a}, \quad \cos \phi = \frac{c}{a},$$

and Tangent is opposite divided by adjacent

$$\tan \theta = \frac{c}{b}, \quad \sin \phi = \frac{b}{c}.$$

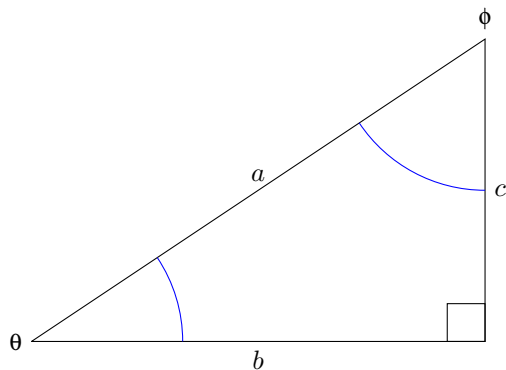


Figure 9.1: A right angle triangle with all the angles and sides marked.

There are associated inverse functions \arcsin , \arccos , \arctan which convert ratios of sides into angles. All of these functions are available on your calculator.

It is important to be careful with which units you are using to express angles. It is likely that you will have come across degrees where going around a full circle is represented by an angle of 360° . However, it is often convenient to work with a different measure of angles called radians, in this case we take a full circle to be 2π rad and express angles as a number between 0 and 2π . In this module you need to work in radians as otherwise the derivative expressions for trig functions will become more complicated. If you are unfamiliar with radians have a look at the next section which explains more about them and how to convert between degrees and radians.

9.4 Radians and Degrees

To work with trig functions we need to know how to describe angles, eg how round in a circle we have gone. Often when working with circles we do not describe angles in terms of degrees but instead use units called radians. You are likely to be familiar with describing angles in a degrees where a full rotation is 360° and a half rotation is 180° . **Radians** are an alternative, and some would say “better” method of measuring angles, they certainly simplify some computations, though they can seem confusing when we first meet them.

The basic idea of radians is to work in terms of the distance around the circle that the angle corresponds to. For a circle with radius r the circumference is $c = 2\pi r$, while an arc of the circle would have length s , shown in red in fig. 9.2. Radians are defined in terms of the length of arcs as

$$\theta = \frac{s}{r}, \quad (9.3)$$

in other words, an angle measured in radians is the ratio of the distance travelled on an arc round the circle to the radius of the circle. This means that a full rotation 360° corresponds to 2π radians as in this case $s = 2\pi r$, half a rotation or 180° then corresponds to π radians.

More generally the conversion rule is that

$$\text{radians} = \frac{\pi}{180^\circ} \times \text{degrees}.$$

The main advantage of radians comes when differentiating and expanding functions as working in degrees requires extra terms to be included and greater care to be taken. While we will not be doing any differentiation in this module we will come across some expressions where we need to work in radians for the quoted expressions to be valid. When this happens we will include a warning so that it is clear where you need to be careful.

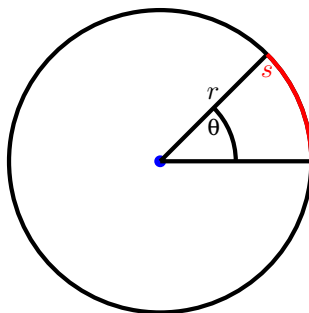


Figure 9.2: A circle of radius r showing how arc length around the circle corresponds to angles

One way to motivate working in radians is to think of unwinding the circle to get a straight line, for a circle of radius r this length will be the circumference, if we divide this length by r we will have a line of length 2π and the angle in radians will be the distance travelled along this line. In other words, working in radians is a bit like converting from circular motion to linear motion.

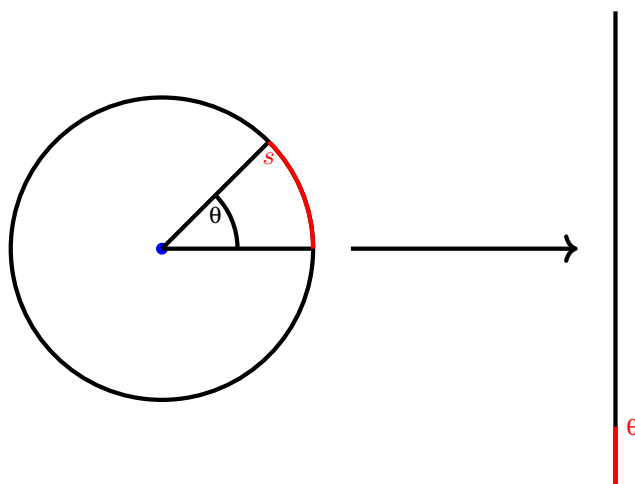


Figure 9.3: A circle of radius r showing how arc length around the circle corresponds to angles

As evidenced in figs. 9.2 and 9.3, it is common practice to use the Greek letter theta, θ , to denote an angle. Other Greek letters like phi, ϕ or φ , and psi, ψ , are sometimes used as well. It is worth spending some time familiarising yourself with these Greek letters so you do not think that they are just badly written English letters.

9.5 Rearranging Equations

A very important skill for solving mathematics problems, and finding the roots of functions, is to be able to rearrange equations. This is sometimes referred to as changing the subject of an equation. Newcastle University have a webpage, available [here](#) that goes through some examples of how to rearrange equations. The webpage also has some self test questions that you can look at if you want more practice. Some of the details are reviewed [here](#), along with examples for the specific equations that we have been using in this module.

In the equation

$$x = 5y + 4z,$$

x is called the subject, which is just a fancy way of saying that x is expressed in terms of the other variables. When we talk about rearranging an equation, we mean that we change the subject of the equation from x to another variable like y and z . We do this by performing a variety of mathematical operations to both sides of

the equation to swap some of the variables from one side to the other.

These operations can include: adding or subtracting a quantity from both sides, multiplying or dividing by a quantity, taking logarithms of or exponentiating both sides of the equation, raising both sides of the equation to any non-zero power.

Example 9.1. *Returning to the above equation $x = 5y + 4z$, we can rearrange it to make z the subject. Again, this means that we will perform mathematical operations on both sides of the equation to put it in the form $z = \dots$:*

$$\begin{aligned} x &= 5y + 4z, && \text{first subtract } 5y \text{ from both sides,} \\ x - 5y &= 5y + 4z - 5y = 4z, && \text{then divide both sides by } 4, \\ \frac{x - 5y}{4} &= \frac{4z}{4} = z. \end{aligned}$$

This gives us

$$z = \frac{x - 5y}{4},$$

with z now the subject of the equation.

Example 9.2. *As another example consider an equation from physics $v = u + at$, which says that if an object starts off with a velocity u and accelerates at a rate a then after time t its velocity will be v . Here we go through this step by step what we do when solving for a :*

$$\begin{aligned} v &= u + at, && \text{subtract } u \text{ from both sides,} \\ v - u &= u + at - u = at, && \text{divide both sides by } t, \\ \frac{v - u}{t} &= \frac{at}{t} = a, \end{aligned}$$

which gives us

$$a = \frac{v - u}{t}.$$

Example 9.3. *Another example is rearranging a different equation from physics,*

$$s = ut + \frac{1}{2}at^2,$$

to solve for either a or t .

To make a the subject we proceed as follows:

$$\begin{aligned} s &= ut + \frac{1}{2}at^2, && \text{subtract } ut \text{ from both sides,} \\ s - ut &= ut + \frac{1}{2}at^2 - ut = \frac{1}{2}at^2, && \text{multiply both sides by } 2, \\ 2(s - ut) &= 2\left(\frac{1}{2}at^2\right) = at^2, && \text{divide both sides by } t^2, \\ \frac{2(s - ut)}{t^2} &= \frac{at^2}{t^2} = a, \end{aligned}$$

thus the equation with a being the subject is

$$a = \frac{2(s - ut)}{t^2}.$$

If instead we wanted t to be the subject then it is easier to put it in the form of a quadratic equation and then use the quadratic formula:

$$\begin{aligned}s &= ut + \frac{1}{2}at^2, \quad \text{subtract } s \text{ from both sides,} \\ 0 &= \frac{1}{2}at^2 + ut - s, \quad \text{this is a quadratic equation for } t \text{ and is solved by} \\ t &= \frac{-u \pm \sqrt{u^2 + 2as}}{a}.\end{aligned}$$

There are a few special cases where we do not need to use the quadratic formula that it is worth being aware of. If $s = 0$ then we have

$$\begin{aligned}0 &= ut + \frac{1}{2}at^2, \quad \text{factor out the common factor,} \\ 0 &= t \left(u + \frac{1}{2}at \right),\end{aligned}$$

This has two solutions $t = 0$, which is the start of the motion, and

$$\begin{aligned}0 &= u + \frac{1}{2}at, \quad \text{subtract } u \text{ from both sides,} \\ -u &= \frac{1}{2}at, \quad \text{multiply both sides by 2,} \\ -2u &= at, \quad \text{divide both sides by } a, \\ -\frac{2u}{a} &= t.\end{aligned}$$

This case often appear when considering projectile motion and you want to calculate the total length of time that the projectile is in the air for.

The other common example is if $u = 0$, then:

$$\begin{aligned}s &= \frac{1}{2}at^2, \quad \text{multiply both sides by 2,} \\ 2s &= at^2, \quad \text{divide both sides by } a, \\ \frac{2s}{a} &= t^2, \quad \text{take the square root of both sides,} \\ \sqrt{\frac{2s}{a}} &= t.\end{aligned}$$

In the last line we have dropped the \pm that should be in front of the square root since we do not consider negative time. However, if you were just solving a quadratic equation then you would have to remember to include that. Both of these special cases can also be found by direct substitution into the quadratic formula of $s = 0$ or $u = 0$ respectively.

Example 9.4. The final example is to rearrange an equation involving a square root,

$$T = 2\pi\sqrt{\frac{l}{g}}.$$

If you are asked to make l the subject of this equation we proceed as follows:

$$\begin{aligned} T &= 2\pi\sqrt{\frac{l}{g}}, \quad \text{first divide both sides by } 2\pi, \\ \frac{T}{2\pi} &= \frac{2\pi}{2\pi}\sqrt{\frac{l}{g}} = \sqrt{\frac{l}{g}}, \quad \text{then square both sides of the equation,} \\ \left(\frac{T}{2\pi}\right)^2 &= \left(\sqrt{\frac{l}{g}}\right)^2 = \frac{l}{g}, \quad \text{now multiply both sides by } g, \\ g\left(\frac{T}{2\pi}\right)^2 &= g \times \frac{l}{g} = l. \end{aligned}$$

This leaves us with

$$l = g\left(\frac{T}{2\pi}\right)^2 = \frac{gT^2}{4\pi^2}.$$

This is the equation for a straight line $y = mx + c$ where l plays the role of y , the y -intercept $c = 0$, $\left(\frac{T}{2\pi}\right)^2$ plays the role of x , and $m = g$ is the gradient of the straight line. When analysing your data you would use plot your data and should observe a straight line whose gradient is g .

At the last step of this rearrangement we could instead make g the subject. To do this we proceed as follows:

$$\begin{aligned} \left(\frac{T}{2\pi}\right)^2 &= \left(\sqrt{\frac{l}{g}}\right)^2 = \frac{l}{g}, \quad \text{now multiply both sides by } g, \\ g\left(\frac{T}{2\pi}\right)^2 &= g \times \frac{l}{g} = l, \quad \text{divide both sides by } \left(\frac{T}{2\pi}\right)^2, \\ g &= \frac{l}{\left(\frac{T}{2\pi}\right)^2} = \frac{4\pi^2 l}{T^2}. \end{aligned}$$

If you want more examples there is a **Transposition of Formulae** workbook, designed by mathcentre, available [here](#).

9.6 More on Limits

This section contains some extra material on limits to help bridge the gap between the material that we need in this module and what you may have seen before if you have not done any mathematics since GCSE. There is a nice discussion of limits in the limits section of [Pierce, 2024] which I have adapted an example from.

The idea of a limit is that you are not looking at the value of a function at a given point, but are looking at what happens to the values of the function as they get close to the specified point. For example if we take

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

It looks like we cannot evaluate this at $x = 1$ since that would give $0/0$ which is not defined, sometimes called *indeterminate*, and usually means that either we have made a mistake in our calculation or we are trying to look at a function in a parameter region where it is not defined. However, just because the function looks to be undefined at $x = 1$ does not mean that we cannot ask what happens as we approach $x = 1$. The easiest

way to do that is to evaluate $f(x)$ at a range of values of x getting closer and closer to $x = 1$:

$$\begin{array}{lcl} x, & f(x) & \\ \frac{1}{2}, & \frac{3}{2} = 1.5, & \\ \frac{3}{4}, & \frac{7}{4} = 1.75, & \\ \frac{9}{10}, & \frac{19}{10} = 1.9 & \\ \frac{99}{100}, & \frac{199}{100} = 1.99 & \\ \frac{999}{1000}, & \frac{1999}{1000} = 1.999 & \\ \frac{9999}{10000}, & \frac{19999}{10000} = 1.9999 & \\ \frac{99999}{100000}, & \frac{199999}{100000} = 1.99999. & \end{array}$$

This shows that the closer x gets to 1 the closer that $f(x)$ becomes to 2. We call this process taking the limit of $f(x)$ as x goes to 1, and say that in this limit $f(x)$ goes to 2. The notation that we use for this is

$$\lim_{x \rightarrow 1} f(x) = \lim_{x \rightarrow 1} \left(\frac{x^2 - 1}{x - 1} \right) = 2.$$

We have a tension here between the value we get when we try to evaluate $f(x)$ at $x = 1$ and when we take the limit as $x \rightarrow 1$. If we could not resolve this then we would say that the function is not defined at $x = 1$ but the limit of $f(x)$ as $x \rightarrow 1$ is defined.

So far we have only look at what happens to $f(x)$ as x goes to 1 from below, i.e. starting from a value of x less than 1 and increasing it towards 1. What happens if we start with a value of x above 1 and lower it towards 1?

$$\begin{array}{lcl} x, & f(x) & \\ \frac{3}{2}, & \frac{5}{2} = 2.5, & \\ \frac{5}{4}, & \frac{9}{4} = 2.25, & \\ \frac{11}{10}, & \frac{21}{10} = 2.1 & \\ \frac{101}{100}, & \frac{201}{100} = 2.01 & \\ \frac{1001}{1000}, & \frac{2001}{1000} = 2.001 & \\ \frac{10001}{10000}, & \frac{20001}{10000} = 2.0001 & \\ \frac{100001}{100000}, & \frac{200001}{100000} = 2.00001. & \end{array}$$

This is also approaching 1, so the limit is the same from both sides. This is good as it means that the limit is defined.

In this case we can actually simplify $f(x)$ to clear up whether the function can be evaluated at $x = 1$ or not. Note that $x^2 - 1 = (x - 1)(x + 1)$ so

$$f(x) = \frac{x^2 - 1}{x - 1} = \frac{(x - 1)(x + 1)}{x - 1} = x + 1,$$

which can be evaluated at $x = 1$ to give $f(x) = 2$. Thus, the limit as x approaches 1 of $f(x)$ agrees with just evaluating the function at $x = 1$.

Warning! this will not always be true. It is possible that the function is not defined at the point, but the limit still exists.

As an example of this let us consider the function

$$f(x) = \frac{1}{x-1}$$

and ask what happens in the limit that x goes to infinity. The important point here is that infinity, denoted ∞ , is not really a number, but we can think of approaching ∞ by taking x larger and larger, and seeing what happens to $f(x)$. Here we get:

$$\begin{array}{ll} x, & f(x) = \frac{1}{x-1} \\ 2, & 1, \\ 3, & \frac{1}{2} = 0.5, \\ 5, & \frac{1}{4} = 0.25 \\ 11, & \frac{1}{10} = 0.1 \\ 101, & \frac{1}{100} = 0.01 \\ 1001, & \frac{1}{1000} = 0.001 \\ 10001, & \frac{1}{10000} = 0.0001. \end{array}$$

This shows that $f(x)$ is tending to zero as $x \rightarrow \infty$. Sometimes we will be careless with notation and write $1/\infty = 0$, but this is really short hand for $\lim_{x \rightarrow \infty} 1/x = 0$. Remember, we cannot evaluate the function $f(x)$ at ∞ but can say how the function behaves as we approach infinity.

Exercise 9.5. Consider the function

$$f(x) = \frac{1}{x-1}$$

again. By exploring what happens as x approaches 1 both from above and below find out if the limit of $f(x)$ as $x \rightarrow 1$ exists.

Chapter 10

Further Reading

Mathematics, you see, is not a spectator sport. To understand mathematics means to be able to do mathematics. And what does it mean doing mathematics? In the first place, it means to be able to solve mathematical problems.

How to Solve It by George Pólya

What we discuss in the lectures is just a guide to calculus its many uses. Depending on the modules that you take later on in your degree you will meet Calculus in different guises, particularly when you come across optimisation problems . In this section I will provide links to further resources and suggestions for further reading based on the material that we met each week. If you have any questions about the topics linked here or you want to get even more information then drop me an email.

It is important to remember that the lectures are their to introduce you to topics and signpost where to find out more information. If you are just attending the lectures and not doing any further reading or solving practice problems then you will struggle to pass the course. Each 20 credit module is considered 200 hours of work, only 36 hours of which are the lectures and seminars. You are expected to put in around 164 hours of self study during a 12 week module. The resources linked here will help with this self study.

10.1 Why Calculus Extra Reading

MIT have a course called **Calculus for Beginners and Artists** which contains some overlap with this module. The webpage for the course is available here and some sections of it, particularly **Chapter 0: Why Study Calculus?** are worth a read to complement what we will do in this module. I will occasionally suggest reading sections of these notes or sections of [Dawkins, 2025b] for a complementary explanation.

The book **Everyday Calculus** by Oscar Fernandez, [Fernandez, 2017], contains a nice overview of many of the places that calculus appears in our everyday lives, even when we do not realise it. If you are ever struggling for motivation with why you should care about calculus, then have a look in this book and explore some of the applications of calculus. The library has a digital copy of this book available so it is easy to dip into sections of it.

10.2 Functions Extra Reading

The first three chapters of [Nitecki, 2022] cover some complementary material to what we met in this module. In particular, there is a discussion of finding limits and convergence, as well as some useful material on continuous functions. This book used to be an essential part of the reading list of this module. However, it is pitched at a slightly higher level than the material that we cover here. In particular, it is aimed at students taking a second course in calculus, so is more appropriate for those on a mathematics degree. There is still some good material in the book, and it can be useful both for those of you looking for material to help cement the basics and those looking to go beyond the material in the module. Students have access to a digital copy of this book through the library so you are able to dip in and out of it as you want to.

If you are looking for more material on trig functions then there is a nice popular science book by Matt Parker called **Love Triangle**, [Parker, 2025], which talks about many of the areas that trigonometry and trig functions appear in the real world. The book also has the nice feature that the page numbers are represented by the sin of the regular page number.

Another nice popular science book is **Once Upon a Prime** by Sarah Hart, [Hart, 2023], which covers the relationship between mathematics and literature. While not directly related to functions, it is a nice read that can help you to discover the mathematics inside some of your favourite stories. It can be a really good way to remove the scare factor that is sometimes associated with mathematics.

If you want more on limits there is some nice content available on the maths is fun website [Pierce, 2024]. In particular, there is a nice discussion of limits available there which you might find useful to complement the discussion of limits in this module.

10.3 Differentiation Extra Reading

The main reference for this section and many of the other sections of these lecture notes is the wonderful book **Mathematical Methods for Physics and Engineering**, [Riley et al., 2006]. There are several copies of this book and its student solution manual in the library and I recommend that you try and have a look at this at some stage.

10.4 Integration Extra Reading

10.5 Differential Equations Extra Reading

10.6 Numerical Methods Extra Reading

The first port of call to see more on numerical methods at the level of this module is in [Lissamen and West, 2004], which contains a very readable discussion of how to approximate functions, integrals, and derivatives. It also spends more time on errors and convergence than we have done so far in this module. It was the main reference when chapter 6 was being written so it should be fairly easy to dip into if you have been reading through the lecture notes. There are physical copies of this book in the library.

10.7 Advanced Topics Extra Reading

A great book to look at for more on integration, both the case of one variable that we focussed on in this module and the more general case of multiple integrals, is the book **Inside Interesting integrals** by Paul Nahin, [Nahin, 2020]. This is a wonderful book if you want to be equipped to deal with any integral you come across. It goes far beyond the level of this module, and some of the examples are down right painful to compute, However, if you enjoy getting your hands dirty and solving every integral that you can get your hands on then this is the book for you. I am recommending it here rather than in the integration section purely because most of the book is so far above the level of this module that it is only useful if you are already comfortable with the basics of integration. Personally I love this book and think that it is incredibly useful.

Chapter 11

Tutorial Sheets

Here we collect all of the tutorial problems for the module. They are split into different weeks depending on the topic they relate to and when they were given out.

Many of these questions are taken from or adapted from the recommended books for the module or from some of the linked resources. These problems are to be attempted in the tutorial sessions and are there to help you familiarise yourself with the material that we have covered in the lectures.

Problems marked with a star, (\star) are particularly worth attempting. Problems marked with a dagger, (\dagger), are more challenging and often go beyond what we directly discussed in the lectures.

The challenge problem sections contain extra problems. Some of them are just there for extra practice, but others are significantly more difficult than what you need to be able to solve to pass the module. If you are finding the content too easy then have a go at the challenge problems. Sometime the challenge problems from one week will be quite similar to the ordinary problems of the next week. as the problems will become more accessible the more material that we cover.

11.1 Week 1

Functions

Problem 11.1.1 (\star). *Find the roots of the polynomial*

$$g(x) = x^2 - 2x - 12$$

and plot the function.

Problem 11.1.2. *Find the roots of the polynomial*

$$g(x) = x^3 + x^2 - x - 1$$

and plot the function.

Problem 11.1.3 (\star). *Consider the function*

$$g(x) = x^2 + 2x + 2,$$

produce a plot of the function by calculating its value at a selection of points. What do you notice as x gets very large? What happens for $x = 0$?

Problem 11.1.4 (\star). Consider the function

$$g(x) = \frac{1}{x-4},$$

and plot the function. what happens as x gets large? What happens as x approaches 4?

Plot the function and comment on its behaviour.

Problem 11.1.5. Draw a schematic of a one-to-one function between the sets

$$\begin{aligned} &\{1, 2, 3, 4, 5\} \\ &\{a, b, c, d, e, f, g\}. \end{aligned}$$

Is there only one way to do this?

Problem 11.1.6 (\star). Given the function

$$f(x) = x^3 - 2x^2 - x + 2$$

find:

- $f(0)$,
- $f(1)$,
- $f(-1)$,
- $f(2)$,
- $f(-2)$,
- $f(t)$,
- $f(x-1)$.

Problem 11.1.7. Consider the function

$$h(x) = \frac{x}{\sqrt{x^2 - 9}}.$$

Find the points where the denominator vanishes, then plot the function avoiding these points. What happens to the plot as the function approaches these points?

Problem 11.1.8. Given the functions

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

find:

- $(f \circ g)(2)$,
- $(g \circ f)(2)$,
- $(f \circ g)(x)$,
- $(g \circ f)(x)$.

Problem 11.1.9 (\star). Given the functions

$$\begin{aligned} f(x) &= 3x - 2 \\ g(x) &= \frac{x}{3} + \frac{2}{3}, \end{aligned}$$

find:

- $(f \circ g)(x)$,
- $(g \circ f)(x)$,
- What is the relationship between $f(x)$ and $g(x)$?

Problem 11.1.10 (\dagger). Given the function

$$h(x) = \frac{x+4}{2x-5},$$

identify when it has an inverse and calculate the inverse.

Challenge Problems

Problem 11.1.11. Consider the function

$$f(x) = \frac{x^2 - x - 12}{x - 1}.$$

Identify the points where the numerator and denominator vanish. Plot the function and explain what happens to the function near these points.

Problem 11.1.12. Given the function

$$f(x) = 2x - 3,$$

find the inverse function $f^{-1}(x)$.

Problem 11.1.13 ($\dagger\dagger$). Build a schematic of a bijection between the natural numbers

$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots\},$$

and the integers

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \dots\}.$$

Are there more integers than natural numbers?

Could you do the same for the real numbers \mathbb{R} ? If you find this interesting you may want to explore the work of Cantor.

11.2 Week 2

Polynomials

Problem 11.2.1 (\star). For the polynomial equation

$$x^3 - 3x + 2 = 0,$$

express it as a product of its factors.

Hint: this means write it as

$$(x - a)(x - b)(x - c),$$

where a, b, c are the roots of the polynomial.

Problem 11.2.2. Find the roots of the following polynomial equation

$$x^4 - 4x^3 + 6x^2 - 4x + 1 = 0.$$

Trig functions Remember to work in radians for any problems related to trigonometry.

Problem 11.2.3. Find the solutions to

$$\sqrt{2} \cos x = 1.$$

Problem 11.2.4. Identify any solutions to the equation

$$\sin(2x) = -2.$$

Problem 11.2.5 (\star). Solve

$$2x \sin x = x.$$

Exponentials and Logarithms

Problem 11.2.6 (\star). Solve the equation

$$x = xe^{4x}.$$

Problem 11.2.7 (\star). Solve

$$2 \ln x - \ln(x + 2) = 1.$$

Limits and Asymptotes

Problem 11.2.8 (\star). For the function

$$f(x) = \frac{x^2 + 4x + 12}{x^2 - 2x}$$

evaluate the limit

$$\lim_{x \rightarrow 2} f(x).$$

Problem 11.2.9 (\star). For the function

$$g(x) = \frac{2x^4 - x^2 + 8x}{7 - 5x^4}$$

evaluate the limits

$$\begin{aligned} \lim_{x \rightarrow -\infty} g(x), \\ \lim_{x \rightarrow \infty} g(x). \end{aligned}$$

Problem 11.2.10 (\dagger). For the function

$$h(x) = \frac{6e^{4x} - e^{-2x}}{8e^{4x} - 2e^{2x} + 3e^{-x}}$$

evaluate the limit

$$\lim_{x \rightarrow \infty} h(x).$$

Problem 11.2.11. Using an appropriate plot evaluate the limits

$$\begin{aligned} \lim_{x \rightarrow \frac{\pi}{2}} \tan x, \\ \lim_{x \rightarrow 0} \tan x. \end{aligned}$$

Problem 11.2.12 (\star). Determine where the function

$$g(x) = \frac{x^3 + x^2 - x - 1}{x^3 - 2x^2 - 4x + 8}$$

fails to be continuous. What do these points correspond to on a plot of $g(x)$?

Challenge Problems

Problem 11.2.13. By making a table of values of $f(x)$, x estimate the limit

$$\lim_{x \rightarrow 2} f(x)$$

from above and below for the function

$$f(x) = \frac{x^2 + 4x - 12}{x^2 - 2x}.$$

Problem 11.2.14. Use the definition of continuity to determine in the function

$$g(x) = \frac{4x + 10}{x^2 - 2x - 15}$$

is continuous and if not find the points where it has discontinuities.

Problem 11.2.15 (†). For the function $g(x)$ in the previous problem look at the definition of differentiability and work out where $g(x)$ fails to be differentiable.

Hyperbolic Trig functions

Problem 11.2.16 (†). Show that

$$\sinh(x + y) = \sinh x \cosh y + \cosh x \sinh y.$$

Problem 11.2.17 (†). Following some of the examples in section 2.4 show that

$$\cosh^{-1} x = \ln \left(x \pm \sqrt{x^2 - 1} \right).$$

11.3 Week 3**Differentiation**

Problem 11.3.1 (*). Find the derivative of

$$f(x) = 7x^2$$

a) using first principles,

b) using the rule for differentiating monomials.

Problem 11.3.2. Find the derivative of

$$f(x) = 3x^3 - 8x.$$

Challenge Problems

Chapter 12

Extra Proofs and Derivations

The material in this chapter is not examinable and is included for completeness. Here we will give a selection of proof and derivations for results and formulas that were used in the rest of the notes. You could consider this whole chapter to be one big mathematical deviation.

To prove the product rule from first principles we proceed as follows. Consider $f(x) = p(x)q(x)$ then

$$\begin{aligned}\frac{df}{dx} &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{p(x+h)q(x+h) - p(x)q(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{p(x+h)q(x+h) - p(x+h)q(x) + p(x+h)q(x) - p(x)q(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{p(x+h)(q(x+h) - q(x))}{h} + \lim_{h \rightarrow 0} \frac{(p(x+h) - p(x))q(x)}{h} \\ &= \lim_{h \rightarrow 0} p(x+h) \frac{q(x+h) - q(x)}{h} + q(x) \lim_{h \rightarrow 0} \frac{p(x+h) - p(x)}{h} \\ &= p(x) \lim_{h \rightarrow 0} \frac{q(x+h) - q(x)}{h} + q(x) \lim_{h \rightarrow 0} \frac{p(x+h) - p(x)}{h} \\ &= p(x) \frac{dq}{dx} + q(x) \frac{dp}{dx},\end{aligned}$$

which is the product rule.

For the quotient rule, we can either prove it from first principles, which is fiddly, or we can use the product rule on $f(x) = p(x)g(x)$ where $g(x) = 1/q(x)$. Here we will take the second approach as hopefully that will be easier to follow. Consider $f(x) = p(x)/q(x)$ and let $g(x) = 1/q(x)$ then applying the product rule means that

$$\begin{aligned}\frac{df}{dx} &= \frac{d}{dx} (p(x)g(x)) \\ &= p(x) \frac{dg}{dx} + g(x) \frac{dp}{dx} \\ &= p(x) \frac{d}{dx} (q^{-1}) + \frac{1}{q(x)} \frac{dp}{dx} \\ &= -p(x)q^{-2} \frac{dq}{dx} + \frac{1}{q(x)} \frac{dp}{dx} \\ &= \frac{1}{q^2} \left(q(x) \frac{dp}{dx} - p(x) \frac{dq}{dx} \right).\end{aligned}$$

Mathematical Diversion 12.1. Note that in the above calculation we have used the chain rule, $((f \circ g)(x))' = f'(g(x))g'(x)$, which we did not discuss until after we had introduced the quotient rule. Also, note that if we

were mathematicians we would need to carefully think about when $p(x)/q(x)$ is differentiable, and implementing the product rule does not need that, it just gives that $p(x)(q(x))^{-1}$ is differentiable if $p(x)$ and $(q(x))^{-1}$ is. This is why for mathematicians, they would prove the quotient rule using differentiation from first principles, or using logarithmic differentiation. The interested reader should have a look at the proof of various derivative properties section of [Dawkins, 2025b] to see the proof in this way.

Next, following [Dawkins, 2025b], we prove the chain rule as follows: let $y = f(u)$, $u = g(x)$ then we know that

$$\frac{du}{dx} = \lim_{h \rightarrow 0} \frac{u(x+h) - u(x)}{h},$$

and that

$$\lim_{h \rightarrow 0} \left(\frac{u(x+h) - u(x)}{h} - \frac{du}{dx} \right) = \lim_{h \rightarrow 0} \frac{u(x+h) - u(x)}{h} - \lim_{h \rightarrow 0} \frac{du}{dx} = \frac{du}{dx} - \frac{du}{dx} = 0.,$$

Now we can define

$$v(h) = \begin{cases} \frac{u(x+h) - u(x)}{h} - \frac{du}{dx} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

which is continuous at $h = 0$ since $\lim_{h \rightarrow 0} v(h) = 0 = v(0)$

Chapter 13

Standard Derivatives and Integrals

There are several functions that it is worth knowing the derivatives and integrals of. When we first introduced their derivatives we either derived them from first principles or used techniques like the product rule, chain rule, or integration by parts, to derive them from already known results. However, this takes time. You do not need to memorise the following list of standard derivatives and integrals, but you may find the list to be a useful resource when going through the tutorial problems or when revising for the exam.

Remember, you should still be able to derive these results if you need to, the list is just intended as an aid.

13.1 Derivatives

13.2 Integrals

Table 13.1: Table of standard derivatives

$y = f(x)$	$\frac{dy}{dx} = f'(x)$
n constant	0
x	1
x^n , n constant	nx^{n-1}
e^{kx} , k constant	ke^{kx}
$\ln(x)$	$\frac{1}{x}$
$\sin(kx)$	$k \cos(kx)$
$\cos(kx)$	$-k \sin(kx)$
$\tan(kx)$	$k \sec^2(kx)$
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$
$\arccos(x)$	$-\frac{1}{\sqrt{1-x^2}}$
$\sinh(x)$	$\cosh(x)$
$\cosh(x)$	$\sinh(x)$
$\tanh(x)$	$\operatorname{sech}^2(x)$
$f(x)g(x)$	$f'(x)g(x) + f(x)g'(x)$
$\frac{f(x)}{g(x)}$	$\frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$

Table 13.2: Table of standard integrals

$y = f(x)$	$\int f(x)dx$
constant k	$kx + c$
x^n	$\frac{x^{n+1}}{n+1} + c$
$\frac{1}{x}$	$\ln(x) + c$
e^{kx} , k constant	$\frac{e^{kx}}{k} + c$
x^{-n}	$\frac{x^{-n+1}}{-n+1} + c$, $n \neq 1$
$\cos(x)$	$\sin(x) + c$
$\sin(x)$	$-\cos(x) + c$
$f(x)g'(x)$	$f(x)g(x) - \int f'(x)g(x)dx$

Bibliography

- Collins. asymptotically. In *Collinsdictionary.com*. URL <https://www.collinsdictionary.com/dictionary/english/asymptotically>. Accessed 24th July 2025.
- P Dawkins. Paul's Online Notes: Algebra, 2025a. URL <https://tutorial.math.lamar.edu/Classes/Alg/Alg.aspx>. Last accessed 6th October 2025.
- P Dawkins. Paul's Online Notes: Calculus I, 2025b. URL <https://tutorial.math.lamar.edu/Classes/CalcI/CalcI.aspx>. Last accessed 5th June 2025.
- O.E. Fernandez. *Everyday Calculus: Discovering the Hidden Math All around Us*. Princeton University Press, 2017. ISBN 9780691175751.
- S. Hart. *Once Upon a Prime: The Wondrous Connections Between Mathematics and Literature*. Flatiron Books, 2023. ISBN 9781250850898.
- D. Jordan and P. Smith. *Mathematical Techniques: An Introduction for the Engineering, Physical, and Mathematical Sciences*. OUP Oxford, 2008. ISBN 9780199282012.
- R. Lissamen and E. West. *Mei Numerical Methods*. MEI Structured Mathematics (a+AS Level) Series. Hodder Education Group, 2004. ISBN 9780340814611.
- P.J. Nahin. *Inside Interesting Integrals: A Collection of Sneaky Tricks, Sly Substitutions, and Numerous Other Stupendously Clever, Awesomely Wicked, and Devilishly Seductive Maneuvers for Computing Hundreds of Perplexing Definite Integrals From Physics, Engineering, and Mathematics (Plus Numerous Challenge Problems with Complete, Detailed Solutions)*. Undergraduate Lecture Notes in Physics. Springer International Publishing, 2020. ISBN 9783030437886.
- Z.H. Nitecki. *Calculus Deconstructed: A Second Course in First-Year Calculus*. AMS/MAA Textbooks. Mathematical Association of America, 2022. ISBN 9781470466756.
- M. Parker. *Love Triangle: How Trigonometry Shapes the World*. Penguin Publishing Group, 2025. ISBN 9780593418116.
- Rod Pierce. Maths is Fun: Calculus, 2024. URL <https://www.mathsisfun.com/calculus/>. Last accessed 12th November 2025.
- K. F. Riley, M. P. Hobson, and S. J. Bence. *Mathematical methods for physics and engineering*. Cambridge University Press, third edition, 2006. ISBN 9780521679718.