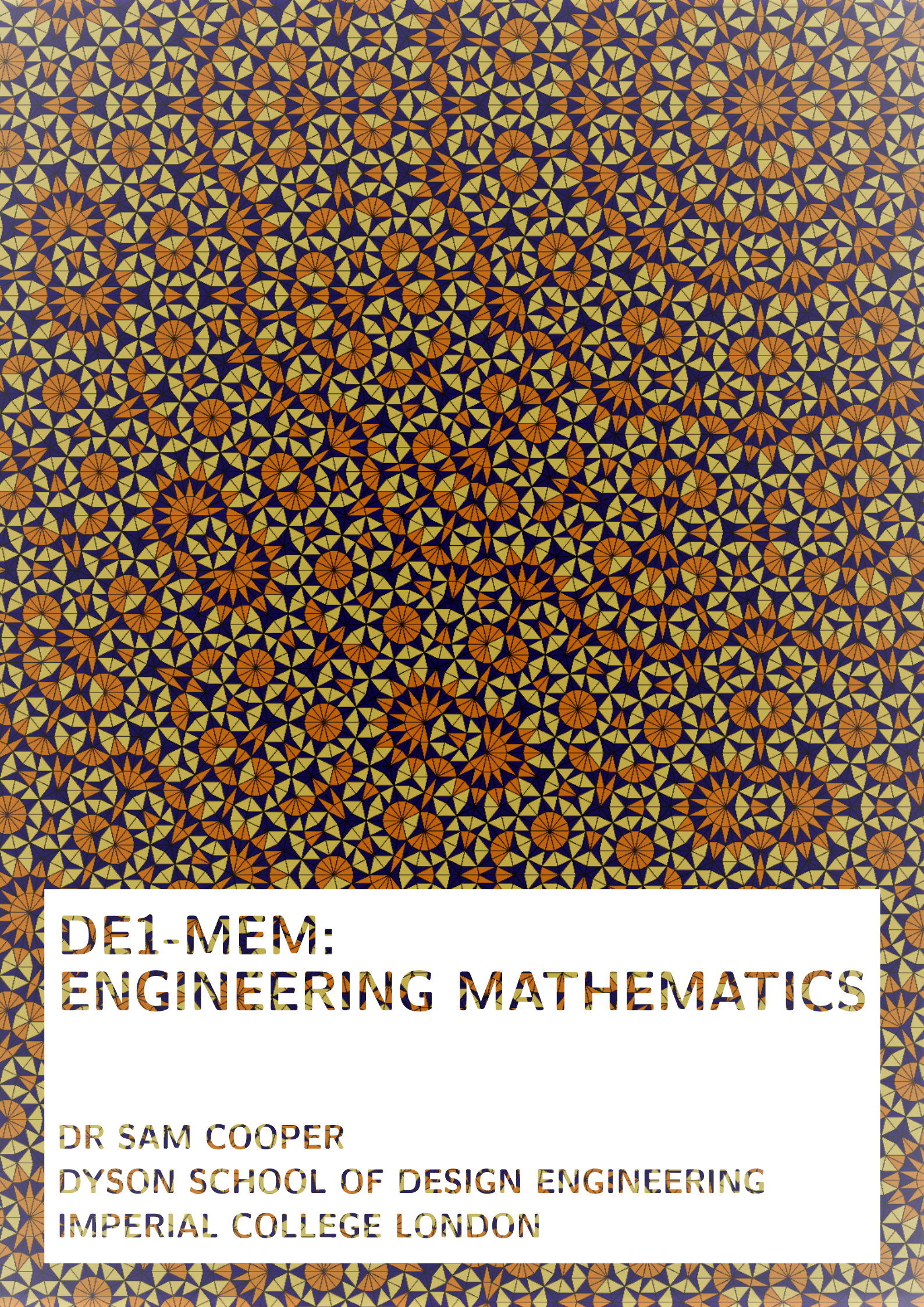
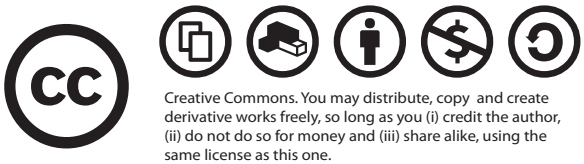


DE1-MEM: ENGINEERING MATHEMATICS

DR SAM COOPER
DYSON SCHOOL OF DESIGN ENGINEERING
IMPERIAL COLLEGE LONDON



These notes were written by Dr Sam Cooper and Dr Freddie Page of the Dyson School of Design Engineering, Imperial College London - corrections to samuel.cooper@imperial.ac.uk.

Thanks to Dr Khalil Rhazaoui for bravely pulling together the first draft of the notes for this course, which were helpful as a basis for creating later versions. Thanks also to Prof. Dave Dye who mentored me as a lecturer and whose notes were helpful for some of the later chapters of this book.

Contents

About the Course

This course is a rapid introduction (or reminder for some) to a range of topics that you will find useful during your engineering career. A huge amount of wonderful resources have become freely available online in the past few years, in the form of videos, blogs, forums, wikis etc.. My hope for this course is that you finish with the confidence necessary to look up questions that you don't understand and hopefully re-purpose methods from one area to another.

Some undergraduate courses expect students to memorise a lot of formulae and derivations; however, now that all of mankind's collected knowledge is just a few clicks away, there is no longer much value in this! Instead, we will focus on developing an intuitive understanding of the various topics, which I hope will not only be more useful, but also much more enjoyable and satisfying!

These notes are not intended to be comprehensive (that is what the internet is for), but instead hope to offer a fast paced and engaging description of the concepts, pitched at a level appropriate to DE1. Some the material is based on notes developed by Dr Rhazaoui, who taught the first iteration of this course.

Course Support and Assessment

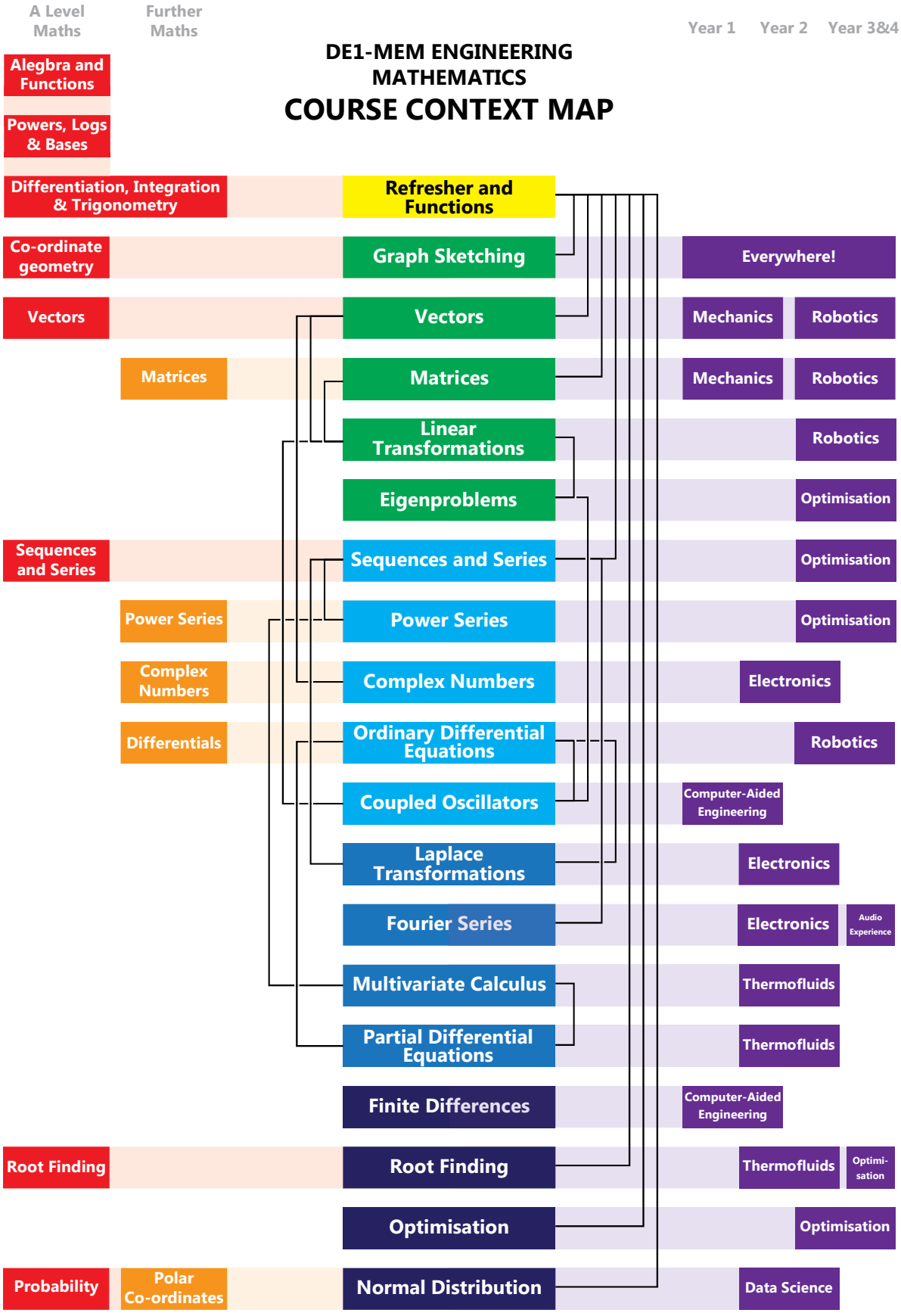
Learning maths is a very personal activity, with each student having their own approach; however, to really understand what's going on, there is no way around putting in the work on your own, occasionally getting stuck and thinking your way out. That said, I really hope the notes, lectures, online videos, tutorial sheets and quizzes help to push you in the right direction and keep you motivated!

Every week, you will take a short non-credit quiz to help me (and you) understand how you're getting on. The course will be assessed through a combination 4 progress tests at half termly intervals, as well as 2 more substantial exams at the beginning of terms two and three. The course is two terms long and each week we will have 2 one hour lectures introducing the material. We will also have weekly tutorial sessions which will be 2 hours in the first term and 1 hour in the second. These sessions are primarily intended for you to ask the tutors questions about the material from the previous weeks and are not ideal for quite study. We will use Learning Catalytics to support the learning process, by running live quizzes.

Further Resources

KL Stroud and DJ Booth, *Engineering Mathematics*, 7th Ed., Macmillan, 2013 (Imperial library 510.246STR), is probably *the* core text for 1st year Maths, although ML Boas, *Mathematical Methods in the Physical Sciences*, 3rd Ed., Wiley, 2006 (Imperial library: 530.15BOA) is a bit less wordy and goes into some more advanced topics as wells.

WolframAlpha is a brilliant mathematical resource and if you are ever stuck with a question, this should be one of your first ports of call. Finally, I would like to recommend several wonderful YouTube series, including WelchLabs, 3Blue1Brown, blackpenredpen and Numberphile as a source of mathematical inspiration and delight.


A Level Maths
Further Maths
Year 1
Year 2
Year 3&4
DE1-MEM ENGINEERING MATHEMATICS
COURSE CONTEXT MAP
Alegbra and Functions
Powers, Logs & Bases
Differentiation, Integration & Trigonometry
Co-ordinate geometry
Vectors
Matrices
Sequences and Series
Power Series
Complex Numbers
Differentials
Root Finding
Probability
Polar Co-ordinates
Refresher and Functions
Graph Sketching
Vectors
Matrices
Linear Transformations
Eigenproblems
Sequences and Series
Power Series
Complex Numbers
Ordinary Differential Equations
Coupled Oscillators
Laplace Transformations
Fourier Series
Multivariate Calculus
Partial Differential Equations
Finite Differences
Root Finding
Optimisation
Normal Distribution
Everywhere!
Mechanics
Robotics
Mechanics
Robotics
Robotics
Optimisation
Optimisation
Optimisation
Electronics
Robotics
Computer-Aided Engineering
Electronics
Electronics
Audio Experience
Thermofluids
Thermofluids
Computer-Aided Engineering
Thermofluids
Optimi-sation
Optimisation
Data Science

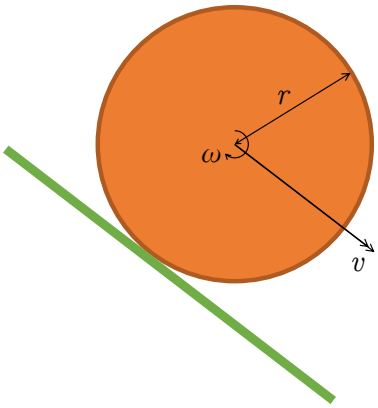
Term 1

Chapter 0

Refresher

This chapter is meant to serve as a high-speed reminder of a few concepts from high school which you will need to progress with this course. Some of you may have forgotten (or even never known) some of this material; however, if you let anyone from the teaching team know this, then I'm sure we can get you up to speed in no time! :-)

0.1 Algebra



You will all be familiar with the core idea of algebra, which is that you can represent numbers and concepts (like functions) with symbols. This is very powerful as it allows you to move away from discussing specific cases and instead describe generalised ideas. For example, if I say that the speed of a cylinder rolling down a hill is v, its rotational speed is ω, and its radius is r; then by considering a bit of geometry, I can write down the relation $v = \omega r$ and this will be true for many different combinations of these three parameters. Fun fact: the word "algebra" comes from the Arabic "al-jabr" which means "the reunion of broken parts"; bonus fact, the word "algorithm" comes from the name of a specific 8^{th} century Persian mathematician called Al-Khwārizmī.

Algebra is at the heart of much of what we do in mathematics and I'm sure you will have done plenty of it at high school. However, for this course I am much more interested in you developing strong mathematical intuition than being masters of grinding through long algebra problems.

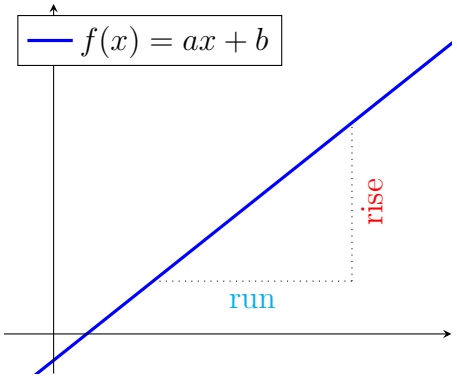
To a great extent, computers have taken over most of the mathematical tasks that engineers were once expected to do; however, you still need to know the key ideas in order to take advantage of the computer's power (and to know what to do when the computer gets stuck). You'll also need to know how to manipulate algebraic expressions, using tools such as "Partial Fractions" (see table), algebraic long division or simultaneous equations.

Fraction $\frac{N(x)}{D(x)}$	Form of denominator, D(x)	Partial Fraction Form (where A, B and C are unknown constants)
$\frac{N(x)}{(ax+b)(cx+d)}$	Linear Factors	$\frac{A}{ax+b} + \frac{B}{cx+d}$
$\frac{N(x)}{(ax+b)^2}$	Repeated Linear Factors	$\frac{A}{ax+b} + \frac{B}{(ax+b)^2}$
$\frac{N(x)}{(ax+b)(cx+d)^2}$	Linear and Repeated Linear Factors	$\frac{A}{ax+b} + \frac{B}{cx+d} + \frac{C}{(cx+d)^2}$
$\frac{N(x)}{(ax+b)(x^2+c^2)}$	Linear and Quadratic (which cannot be factorised) Factors	$\frac{A}{ax+b} + \frac{Bx+C}{x^2+c^2}$

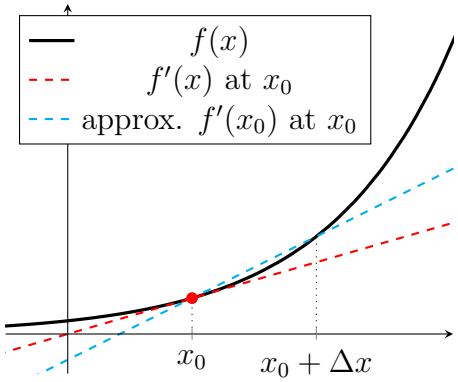
0.2 Calculus

We will be using calculus in almost every topic in this module, so it seems sensible for us to start with a quick refresher. However, I won't be expecting you to remember lots of tricky substitutions and identities because, since the advent of the internet, these have turned out to not be very useful things to know. Instead, what we'll be focusing on is the core understanding behind each topic, which will hopefully give you mathematical confidence, without being too dull.

If you are given the graph of a linear function (*i.e.*, $f(x) = ax + b$ in the adjacent figure), you can calculate the slope by simply drawing a right-angle triangle and then calculating the ratio of the vertical to the horizontal lines (*i.e.*, "rise over run" or "RoR"). You can perform this operation anywhere along the line because this function has a constant gradient, but what do we do if our function's gradient is variable?



For the general function $f(x)$, the "gradient" at a point is the slope of the curve at that point. In the figure below, we're looking for the gradient of the **black** line at point x_0 (the red point). To apply the RoR approach again, we must pick another point further along the x-axis, which we will call $(x_0 + \Delta x)$. However, as you can see from the figure, the line passing through these two points is not a great approximation to the slope at x_0.



What I hope you can also appreciate is that as our two points move closer together (*i.e.*, as Δx gets smaller), then this approximation will improve. We can formalise and extend this rational to write the following expression, which says that "in the limit" as Δx "goes to zero" (*i.e.*, becomes *infinitesimally* small), then our approximation will become exact. So our gradient becomes (three different derivative notations styles also shown):

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} \equiv \frac{\mathrm{d}}{\mathrm{d}x}f(x) \equiv f'(x) = \lim_{\Delta x \to 0}\left(\frac{f(x + \Delta x) - f(x)}{\Delta x}\right)$$

This is still just the same old RoR concept, but taken to an extreme limit. You may not have encountered the "lim" notation before, but what it's asking you to do is find the value of the expression as Δx goes to 0, but not actually at 0 as this would break our fraction (*i.e.*, you can't divide by zero) - we'll cover this concept in more detail in a later chapter.

0.3 Using calculus

Although you will probably have lots of derivatives memorised by now, it's important to remember that you are just using shortcuts to evaluate the "lim(RoR)" equation above.

For example, if $g(x) = 3x^2 - 5$, then we can simply substitute this into the expression above and rearrange to find the derivative. Make sure you are comfortable working through the following example:

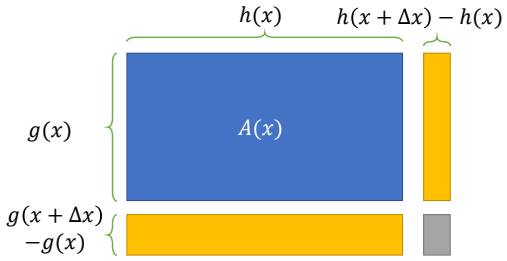
$$\begin{aligned} g'(x) &= \lim_{\Delta x \to 0} \left(\frac{(3(x+\Delta x)^2 - 5) - (3x^2 - 5)}{\Delta x} \right) \\ &= \lim_{\Delta x \to 0} \left(\frac{(3x^2 + 6x\Delta x + 3\Delta x^2 - 5) - (3x^2 - 5)}{\Delta x} \right) \\ &= \lim_{\Delta x \to 0} \left(\frac{6x\Delta x + 3\Delta x^2}{\Delta x} \right) = \lim_{\Delta x \to 0} (6x + 3\Delta x) \quad = \quad 6x \end{aligned}$$

You can see that once you've work through the algebra, it's that final step where the magic of the "lim" happens. We have $6x + 3\Delta x$, but as Δx becomes very small we can just ignore it.

Now that we've understood the core concept, we can start to build a list of time saving rules so that we don't have to use lim(RoR) equation each time. As we've already seen in our first example, polynomials can be dealt with using two rules. The **Power Rule** tells us that differentiation of a simple power can be efficiently calculated by multiplying the original power to the front and then reducing the power by 1 (*i.e.*, $f(x) = ax^b \Rightarrow f'(x) = abx^{b-1}$). Also in the above example is the **Sum Rule**, which says that the derivative of the sums is the sum of the derivatives (*i.e.*, $f(x) = g(x) + h(x) \Rightarrow f'(x) = g'(x) + h'(x)$).

The **Product Rule**, which tells us how to differentiate the product of two functions, so if $f(x) = g(x)h(x)$,

$$\frac{\mathrm{d}(f(x))}{\mathrm{d}x} \equiv f'(x) = g'(x)h(x) + g(x)h'(x)$$



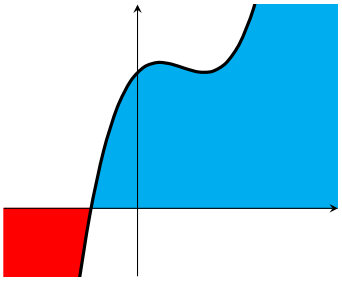
Perhaps the simplest way to think about the product rule is to consider $g(x)$ and $h(x)$ to be the length of two sides of a rectangle of area $A(x)$. This means that to differentiate w.r.t. x (NB "w.r.t." is short for "with respect to") is simply to ask how does $A(x)$ change with x. We can imagine that for a certain function, increasing x by some small Δx will increase $A(x)$ by the amount shown in the two yellow and one grey boxes in the adjacent figure. If you now write down an expression for this increase in area, divide it by Δx and once again take the limit, you'll recover the expression above. When you try this, you'll notice that the area of the grey box ends up being ignored.

The **Chain Rule** is slightly harder to visualise, but a similar logic can be applied. Consider a function of a function $g(h(x))$. We can differentiate this thing by using the following expression.

$$f(x) = g(h(x)) \qquad \Longrightarrow \qquad f'(x) = \frac{\mathrm{d}}{\mathrm{d}x} g(h(x)) = g'(h(x))\ h'(x) \equiv \frac{\mathrm{d}g}{\mathrm{d}h} \times \frac{\mathrm{d}h}{\mathrm{d}x} = \frac{\mathrm{d}g(h(x))}{\mathrm{d}x}$$

For example, if $f(x) = g(h(x))$, where $g(x) = 2x^2 + 3$ and $h(x) = 5x^3 - 1$. As $g'(x) = 4x$ and $h'(x) = 15x^2$, therefore, $f'(x) = g'(h(x))\ h'(x) = 4(5x^3 - 1) \times 15x^2 = 300x^5 - 60x^2$. In this simple example, you could equally have found this result by simply multiplying out $g(h(x))$ and taking the derivative of the resulting expression.

0.3.1 Integration



Integration is just the inverse operation to differentiation, but can also be thought of as finding the area between a function and the axis (NB. If a function is negative, this area counts as negative, for example, the integral of the sine function between 0 and π is 1, but between 0 and 2π it's gone down to 0 again).

The **Power Rule** for integration tells us that the integral of a simple power can be efficiently calculated by increasing the power by 1 and then dividing the coefficient by this new power (*i.e.*, $f(x) = ax^b \Rightarrow \int f(x)\,\mathrm{d}x = \frac{a}{b+1}x^{b+1} + c$). This is clearly the inverse operation to the differential power rule, with the only key difference being the appearance of a new term c. This term is simply a constant and you can see that if you had started by differentiating a function which contained a constant and then integrated again, you wouldn't know what this constant was (*e.g.* $f(x) = 3x^2 + 3 \Rightarrow f'(x) = 6x \Rightarrow \int 6x\mathrm{d}x = 3x^2 + c$).

For the **Sum Rule**, the same rules apply as for differentiation, whereby the integral of the sums is the sum of the integrals (*i.e.*, $f(x) = g(x) + h(x) \Rightarrow \int f(x)\,\mathrm{d}x = \int g(x)\,\mathrm{d}x + \int h(x)\,\mathrm{d}x$).

The key time saving rule that you should be aware of for integration is called **Integration by Parts** or sometimes just **Parts** for short. One way to think about this process is as a rearrangement of the product rule for differentiation. So, for the function $f(x) = g(x)h(x)$:

$$\frac{f(x)}{\mathrm{d}x} = \frac{g(x)}{\mathrm{d}x}h(x) + \frac{h(x)}{\mathrm{d}x}g(x) \quad \xrightarrow{\text{integrate \& rearrange}} \quad \int \left(g(x)\frac{h(x)}{\mathrm{d}x}\right)\mathrm{d}x = f(x) - \int \left(h(x)\frac{g(x)}{\mathrm{d}x}\right)\mathrm{d}x$$

To make this easier to remember, it can be shortened to $\int g\mathrm{d}h = gh - \int h\mathrm{d}g$. Make sure you understand why this is and how to use it, as we'll be using it a lot in the chapter on Fourier series.

In general, integration is tough (tougher than differentiation) and often doesn't give you nice solution in terms of elementary functions.

0.4 Powers, logs & bases

In the simple case where an exponent, n, is a positive integer, it specifies the number of times a variable, b, is multiplied by itself.

$$\begin{aligned} a &= b^n \\ &= \underbrace{b \times \ldots \times b}_{n \text{ times}} \end{aligned}$$

However, as you will have seen, this definition can be expanded to allow for any exponent (or "power") positive or negative, real or complex. The following equations show several representations of the same number, achieved through manipulating and interpreting its exponent. Make sure you understand how to convert between each of these forms.

$$7^{-\frac{2}{3}} = 7^{-\frac{1}{3}} \times 7^{-\frac{1}{3}} = \frac{7^{\frac{1}{3}}}{7} = \frac{7^{-\frac{1}{3}}}{7^{\frac{1}{3}}} = \frac{1}{7^{\frac{2}{3}}} = \frac{1}{\sqrt[3]{(7^2)}} = \left(\frac{1}{\sqrt[3]{7}}\right)^2$$

A logarithm (or "log") is the inverse operation to exponentiation, where b is now referred to as the *base* of the logarithm.

$$\log_b(a) = n \qquad (\text{"Log to the base } b \text{ of } a \text{ equals } n\text{"})$$

When dealing with addition or multiplication, we have a clear picture in our mind of what an equation is asking us to do, but people tend to be less clear with logs, which is perhaps because they cannot turn the mathematical statements into sentences.

$x = 2 + 10$
"*What* is two add ten?"

$x = 2 \times 10$
"*What* is two lots of ten?"

$x = 10^2$
"*What* is ten times itself?"

$x = \log_{10}(100)$
"*What* power of ten makes one hundred?"

It can also be useful to refer to an "easy to recall" example, such as $\mathbf{\log_{10}(100) = 2}$, to help you remember how to convert between logarithms and exponents.

N.B. You will often see the expression "$\ln(x)$", which is simply the logarithm with the Euler's number, $e = 2.718...$ as its base: $\ln(x) = \log_e(x)$. This is commonly referred to as the "natural logarithm". We'll be meeting e again later in the course.

We now need to learn how to manipulate logs:

Rule - Addition

$$\log_b(x) + \log_b(y) = \log_b(xy)$$

Example - We now know that $\mathbf{\log_{10}(100) = 2}$, so clearly $\log_{10}(100) + \log_{10}(100) = 2 + 2 = 4$, but using the addition rule this also implies that $\log_{10}(100 \times 100) = \log_{10}(10000) = 4$, which makes sense, as $10^4 = 10000$.

Rule - Subtraction

$$\log_b(x) - \log_b(y) = \log_b\left(\frac{x}{y}\right)$$

Example - We can see that $\log_2(32) - \log_2(4) = \log_2(2^5) - \log_2(2^2) = 5 - 2 = 3$, but we could have also arrived at this result by using the subtraction rule, as $\log_2(32) - \log_2(4) = \log_2(32/4) = \log_2(8) = \log_2(2^3) = 3$. You should also notice here that the subtraction rule is just a logical extension of the addition rule and follows from our discussion of powers at the beginning of this chapter.

Rule - Powers

$$\log_b(x^p) = p\log_b(x)$$

Example - If you are given the expression $-0.2\log_2(243)$ you can convert this to an alternative form where the coefficient is within the log function as follow: $\log_2(243^{-0.2}) = \log_2(243^{-\frac{1}{5}}) = \log_2(\frac{1}{\sqrt[5]{243}}) = \log_2(\frac{1}{3})$

0.4.1 Change of base

We can re-express $\log_b(x)$ in terms or an arbitrary base c using the following formula.

$$\log_b(x) = \left(\frac{1}{\log_c(b)}\right)\log_c(x)$$

Example - $\log_8(64)$ can be express in base 2 as $\left(\frac{1}{\log_2(8)}\right)\log_2(64)$

This can be useful for expressing all the terms in an equation in the same base, which makes manipulation easier. Using the power rule from the previous section, we can also clearly re-express this in the following manner.

$$\begin{aligned}\log_b(x) &= \log_c\left(x^{1/\log_c(b)}\right)\\ &= \log_c\left(\sqrt[\log_c(b)]{x}\right)\end{aligned}$$

Example - $\log_9(x)$ can be express in base 3 as $\log_3(x)/\log_3(9) = \log_3(x)/2 = \log_3(\sqrt{x})$

0.5 Engineers love

0.5.1 Unit comparisons and dimensional analyses

Rather than looking at equations as just a collection of abstract numbers and symbols, engineers are usually attributing some physical meaning to them. This means that each term *may* have units. Furthermore, each term may be higher dimensional than a simple scalar (e.g. 7) and could be a vector (e.g. [3, 2, 4] or matrix, [1, 2; 5, 3], etc.).

Crucially, only quantities with the same units and dimensions may be added (+), subtracted (-) or compared ($=, <, >$). This is very useful as it allows us to quickly assess whether a problem has been correctly stated and frequently to spot ways to simplify an expression. Also, it means you have less to remember as you can always check the units to see if, for example, your fraction is the right way up.

Think of the following equation of motion, where s is distance, u is speed and t is time.

$$s = ut + 0.5at^2 \qquad \text{units} \rightarrow \qquad [\text{m}] = \left[\frac{\text{m}}{\text{s}}\right][\text{s}] + [??][\text{s}^2]$$

Just by simple comparison of the units, you can tell that the a term must have units of $\left[\frac{\text{m}}{\text{s}^2}\right]$ and therefore be an acceleration term.

0.5.2 Order of magnitude approximations

Another crucial engineering skill is "order of magnitude analysis", which is essentially a method for simplifying equations by ignoring certain terms.

Consider the expression $y = \frac{x+1}{x^2}$. If you were told to evaluate this expression only for values of $x >> 1$ (i.e. x *much* larger than 1), it would be reasonable to forget about the 1 and say $y \approx 1/x$. This kind of approximation is typically what we might call a "back of the envelope" calculation.

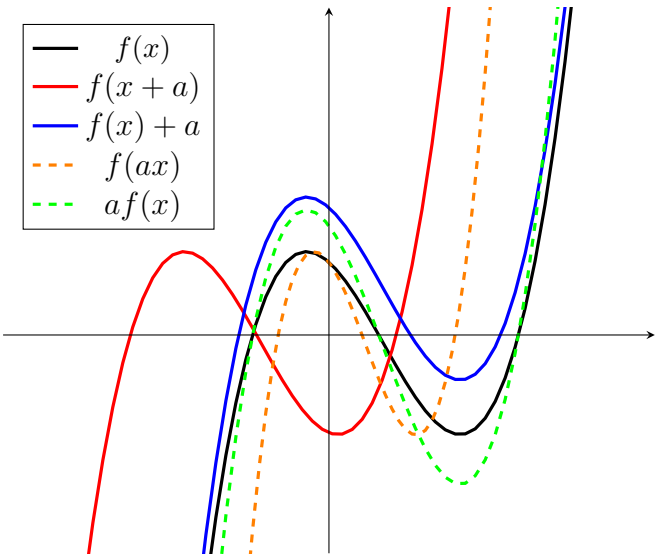
In complicated equations, such as the Navier-Stokes equations which describe viscous fluid flow, often the first step in their analysis is to work through each of the terms and determine which are small enough to be ignored - this is just the kind of thing engineers love and you'll be doing plenty of it on this degree!

We also have some special notation to characterise the "order of magnitude", $\mathcal{O}(x)$, which we'll be putting in to use in the chapter on power series approximations.

0.5.3 Curve sketching

Although the whole next chapter is about curve sketch (as it's such an important topic), you should all have covered various simple transforms that enable you to shift and stretch a curve on the plane.

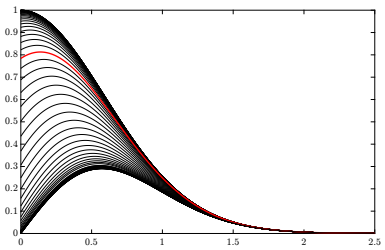
Study the adjacent figure and make sure you can see how the constant a (assuming $a > 0$) transforms the original black curve to the four new curves shown.



Transforms of a cubic function, using a constant factor $a > 0$.

Conclusion

If anything we covered in this chapter was new to you (or old... but still mysterious), then please let someone from the teaching team know and we will do our best to get you up to speed. Alternatively, if you are feeling confident enough to have a look online, then there as so many wonderful resources available to help you - Khan Academy is an excellent place to start with high school topics like these.



Chapter 1

Functions

In many ways this is the most important chapter of the course - if you are able to sketch and manipulate functions with confidence, then all the other methods we will discuss will be much simpler. Ultimately, sketches in general are just diagrams designed to convey some specific bits of information whilst not worrying too much about others - function sketching is no different.

1.1 Curve Sketching

When sketching a curve, there are several key features which need to be considered.

1. General Shape
2. Intercepts ($x = 0$ and $y = 0$)
3. Asymptotes
4. Stationary Points ($\frac{\mathrm{d}y}{\mathrm{d}x} = 0$)
5. Inflection Points ($\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 0$)
6. Domain and Range

1.1.1 General Shape

It is very useful, before you start calculating any of the specific features, to have a picture in your mind of roughly how the curve should look. The following four plots are to help you remember some common functions that you should be familiar with. Put your finger over the colour indicators in the legend and make sure you can pair up the curves with the functions. Apologies for how busy each figure is, but this is the best way to see the patterns!

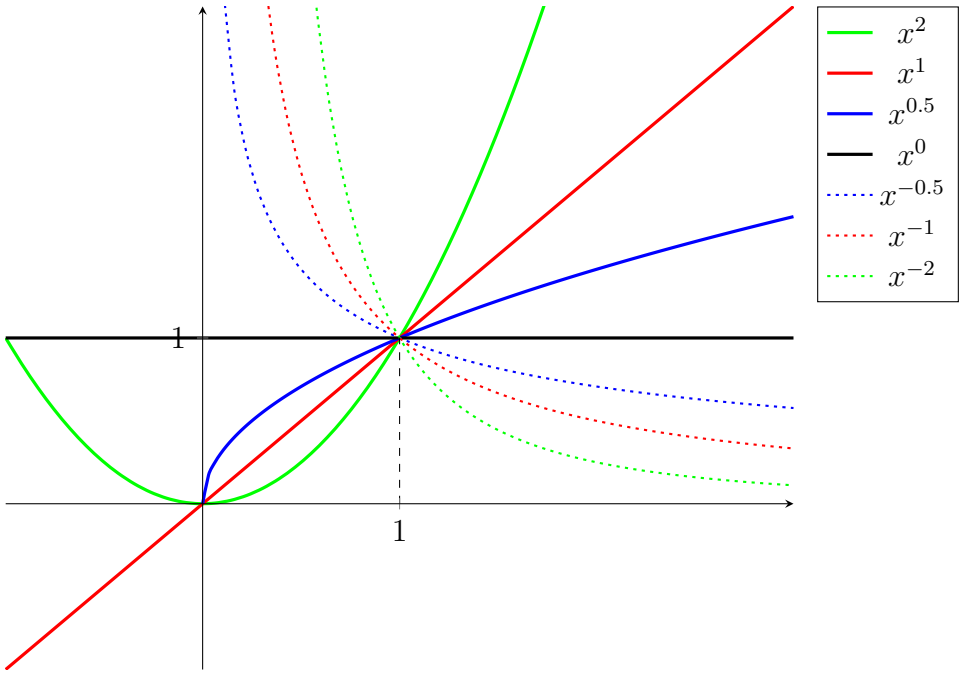
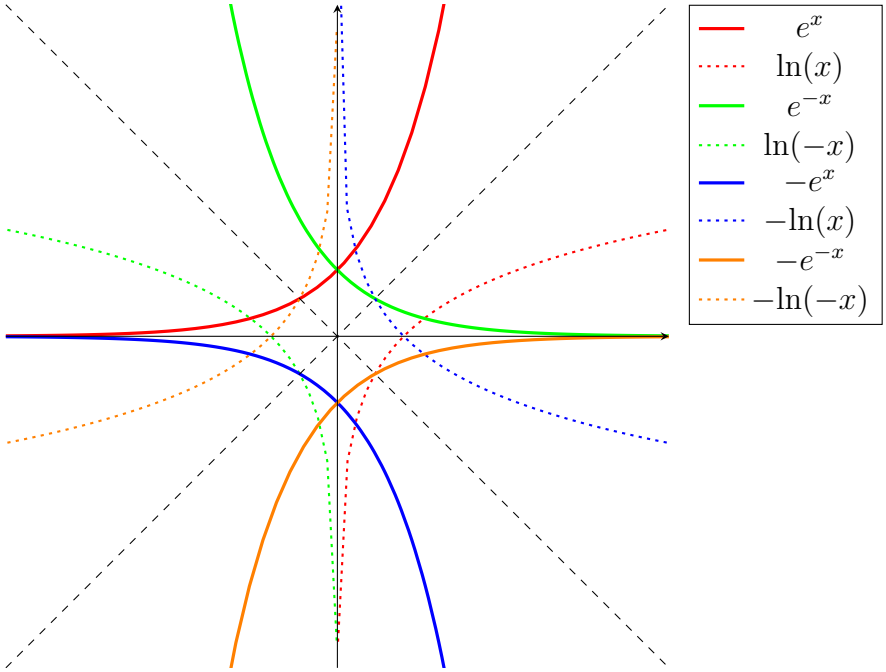
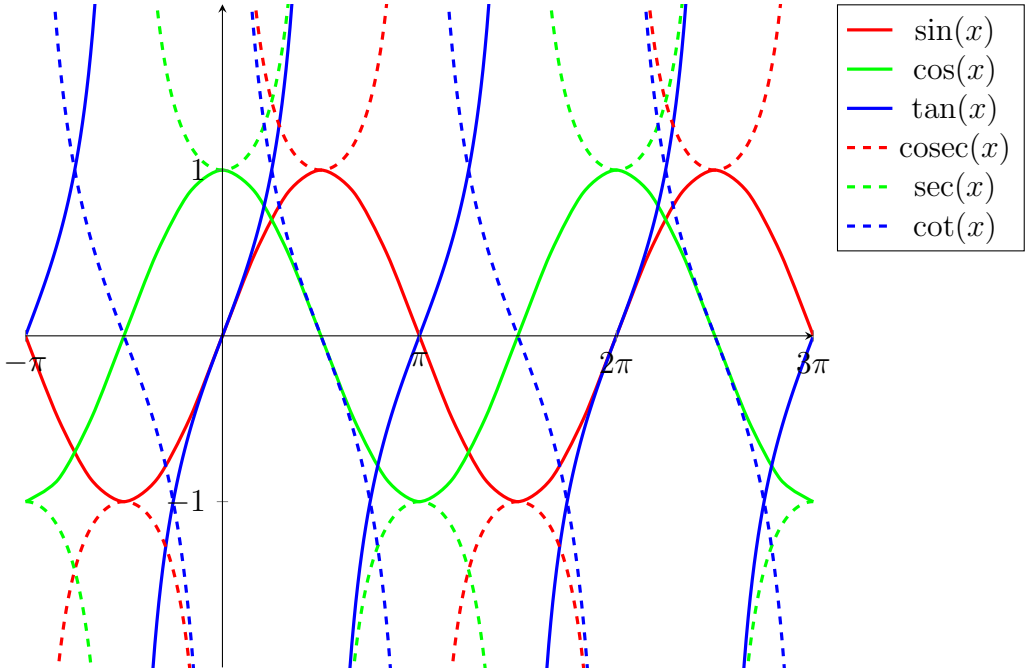


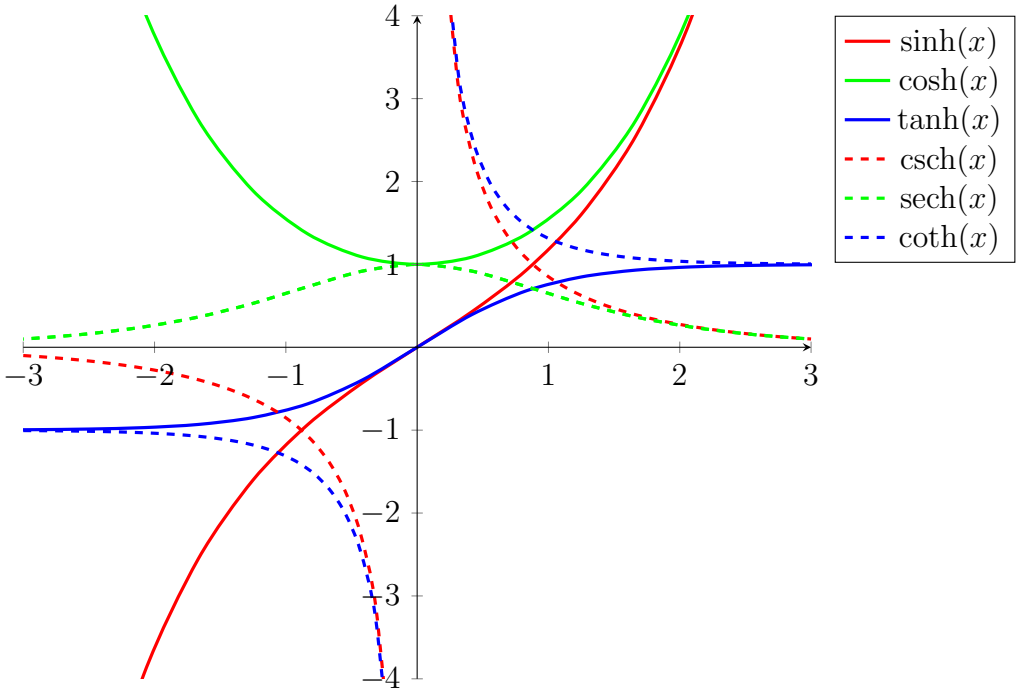
Figure show the effect of varying the index of variable.



Natural functions, illustrating that inverse functions can be constructed with simple reflections across the line $y = x$



Trigonometric functions



Hyperbolic functions

1.1.2 Intercepts

Once the general shape has been established, it is then often useful to be able to label certain points of interest. If you are given an *explicit* equation (*i.e.* in the form $y = f(x)$), then a trivial point to find is the intercept of the vertical axis. This is evaluated by setting the independent variable to zero.

Example - For the curve $y = 3x^3 - 47x + 9$, the y-intercept occurs at $y = 3(0)^3 - 47(0) + 9 = 9$

The points at which the curve crosses the x-axis are called *roots*. For some simple equations, they can be found by inspection.

Example - The root of the curve $y = \frac{x-1}{x^2}$, can be found by considering when the function would equal zero. This will occur only when the numerator of the fraction is also zero; therefore, we need only solve $x - 1 = 0$, giving us $x = 1$.

For some other functions, we must first rearrange the equation to a form that yields the roots.

Example - The roots of the curve $y = x^2 + 4x - 21$, can be found by first factorizing the equation to the form $y = (x - 3)(x + 7)$. In order to solve this equation at $y = 0$, we must find the two values of x that cause each of the bracketed terms to be zero. Therefore, the roots occur at $x = 3$ and $x = -7$.

Furthermore, the roots of all equations of the form $y = ax^2 + bx + c$ can be found using the familiar "quadratic formula".

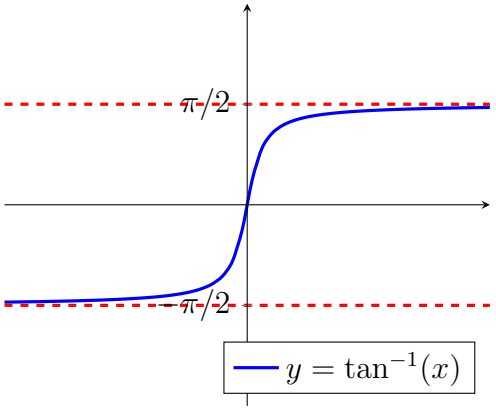
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

However, there remain many equations which cannot be tackled with any of the above methods. Consider, for example, the function $y = x^{3/2} - x + 7$. To find the roots in this case we are forced to employ *numerical methods*, which are discussed in a later chapter.

1.1.3 Asymptotes

An asymptote is a straight line that is continually approached by a given curve, but does not meet it at any finite distance. Asymptotes can be vertical, horizontal or oblique (slanted), as illustrated in the following figure.

If a function can be expressed as a fraction, then a vertical asymptote will occur when the denominator equals zero. Also, if the degree of the numerator is one higher than the denominator, it may also have a slant asymptote.

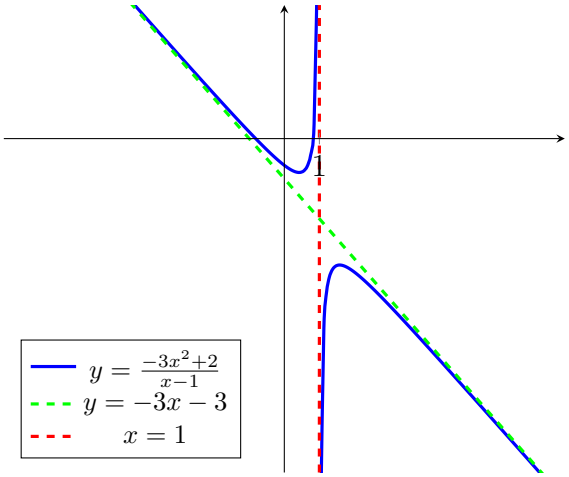


Two horizontal asymptotes

Example - For the equation $y = \frac{-3x^2+2}{x-1}$, there is an easy to spot vertical asymptote when the denominator of the fraction equals zero (*i.e.*, at $x = 1$).

However, notice that the degree of the numerator is higher than that of the denominator, which means there may also be a slant asymptote. The next step is to perform algebraic long division.

$$
\begin{array}{r|l}
 & -3x-3 \\
\hline
x-1 & -3x^2 \quad +2 \\
 & \underline{3x^2-3x} \\
 & -3x+2 \\
 & \underline{3x-3} \\
 & -1
\end{array}
$$



Vertical and slant asymptotes

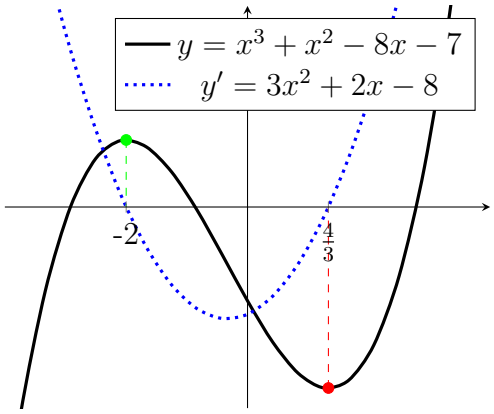
Which tells us to expect an asymptote on the line $y = -3x - 3$. Now that we have our slant asymptotes, we should think about whether our function will be above or below this line. Perhaps the simplest way of doing this is just to sample the pair of points either side of $x = 1$ and see if they are positive or negative.

1.1.4 Stationary Points

Stationary points are where the gradient of a curve is zero. They can be found by differentiating the function and finding the values of x where the differential is zero.

Example - Differentiating the function $y = x^3 + x^2 - 8x - 7$, gives the expression $\frac{dy}{dx} = 3x^2 + 2x - 8 = (3x - 4)(x + 2)$. Stationary points occur at $(3x - 4) = 0$ and $(x + 2) = 0$, yielding $x = 4/3$ and $x = -2$.

If the gradient of the function changes sign at the stationary point, then it is called a "turning point". It is also possible to determine whether a turning point is a local maximum or minimum by differentiating a second time and evaluating the second differentials at each turning point. If the second differential is positive, then the point is a minimum and *vice versa*.



Plot showing a polynomial, its derivative and the stationary points.

Example - Differentiating the function $y = x^3 + x^2 - 8x - 7$ twice yields $\frac{d^2y}{dx^2} = 6x + 2$. Taking the stationary points from the previous example, we find that evaluating the second derivative at the stationary point $x = 4/3$ gives $6(4/3)+2 = 10$, so it is a local minimum, and similarly at the stationary point $x = -2$ gives $6(-2) + 2 = -10$ so it is a local maximum.

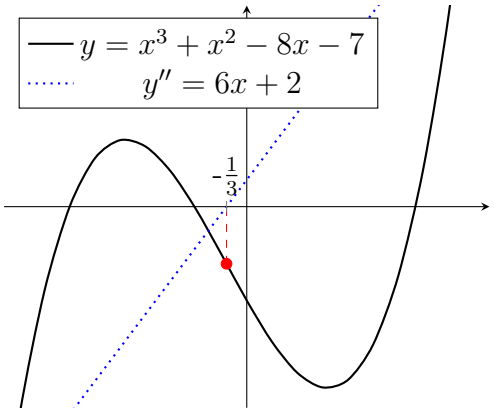
If the gradient of the function does not change sign at the stationary point, then it is called

a point of "horizontal inflection". Inflection points are discussed in the next section, but to visualise a curve with a stationary point that is not a turning point, think of the function $y = x^3$.

Finally, if you'd like to evaluate the y-coordinates of stationary points, simply substitute their x-coordinate back into the original equation (this might sound obvious, but people do forget!).

1.1.5 Inflection Points

An inflection point is a point on a curve at which the sign of the curvature (*i.e.*, the concavity) changes. Inflection points may be stationary points (*e.g.* the function $y = x^3$), but do not have to be and they are not local maxima or local minima. They can be located by finding where the second derivative of a function equals zero.



Plot showing a polynomial, its second derivative and the inflection point.

Example - Differentiating the function $y = x^3 + x^2 - 8x - 7$ twice yields $\frac{d^2y}{dx^2} = 6x + 2$. Setting this to zero, we find that $6x + 2 = 0$, which gives $x = -1/3$.

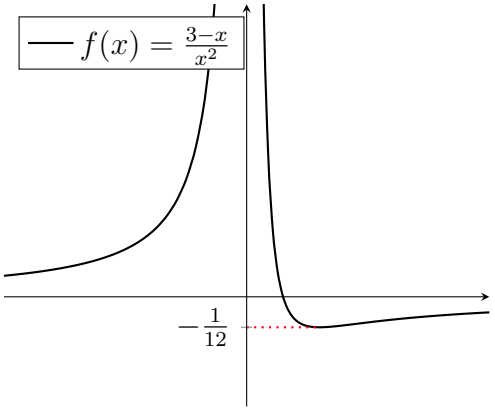
1.1.6 Domain and Range

The *domain* is the set of all x coordinates that have a corresponding y coordinate.

The *range* is the set of all y coordinates that have a corresponding x coordinate.

They can be expressed using set notation, where square brackets "[ ]" signify that the point is included and round brackets "( )" signify that it is excluded. By convention, infinities are consider to be excluded. If our domain has multiple regions, separated by discontinuities, then we can express this concept using the union symbol "∪".

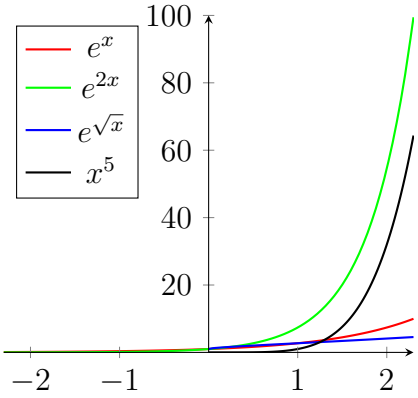
Example - For the function $y = \frac{3-x}{x^2}$, as shown in fig. 1.9, there is an asymptote at $x = 0$ and the global minimum (*i.e.*, the lowest point) occurs at the coordinate $(6, -\frac{1}{12})$. We can therefore express the domain as $(-\infty, 0) \cup (0, \infty)$ and the range as $[\frac{-1}{12}, \infty)$



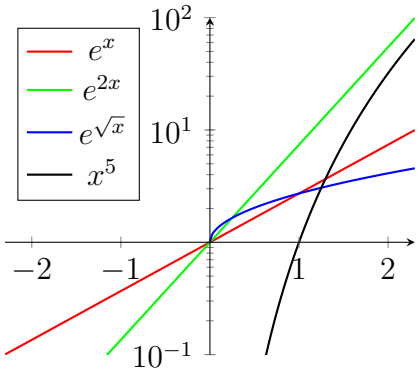
Plot of asymptotic example function

1.1.7 Log axes

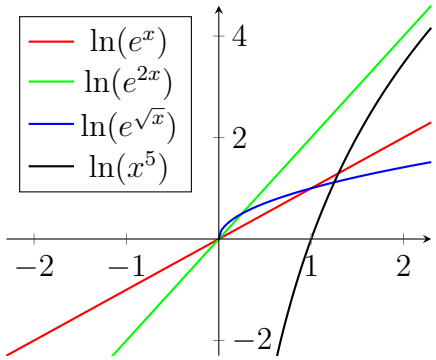
One application of logs that you will encounter frequently as an engineer is plotting graphs where one ("log-linear") or both ("log-log") of the axes use a log scale. For example, a log-linear plot might be required when the independent variable causes the dependant variable to range over multiple scales; whereas, the log-log plot can be used to extract "power law" relationships (such as growth). The following figures plot the same three functions in each of the three graphs.



Linear Cartesian axes



Semi-log axes in base 10



Linear axes plotting natural log of the functions

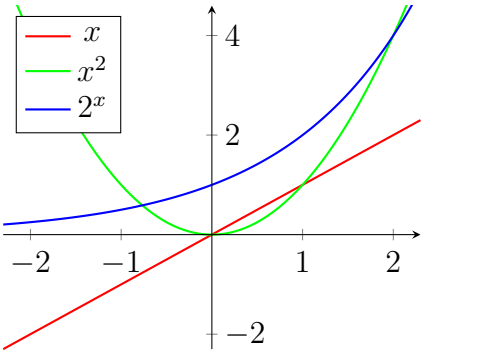
The left figure shows the functions on linear axes with which you are familiar. The middle figure, plotted with a semi-log (base 10) y-axis, makes the first two functions into straight lines. Finally, in figure on the right, by plotting the log of the functions in the appropriate base (in this case base e), allows the coefficient of the power to be directly measured from the graph as the gradient for the first two functions, but not for the last.

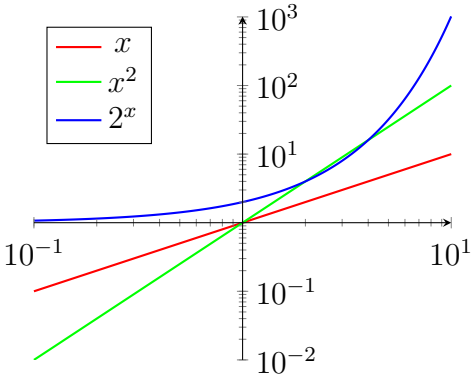
Log-log axes

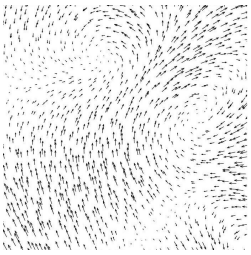
In the case of log-log axis, perhaps the easiest way to see whether a function will be a straight line is to take the log of both sides of you expression and then make the following substitution.

$$X = \log(x) \qquad \& \qquad Y = \log(y)$$

And then check if this substituted function is itself linear. For example, considering the function $y = 7x^2$ and then taking logs (any base is fine) of both sides, we get $\log(y) = \log(7x^2) = \log(7) + \log(x^2) = \log(7) + 2\log(x)$. Now, making the above substitution, we get $Y = \log(7) + 2X$. Remembering that $\log(7)$ is just a number, we see that our new expression matches the form $y = mx + c$ and so must be a straight line on a log-log scale.







Chapter 2

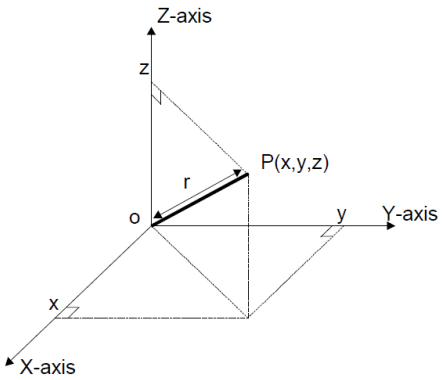
Vectors

When you first start learning maths at school, you immediately encounter "scalars", which can be formally described as a number with a "magnitude, but no direction"... or less formally as *just a number*. However, as the formal definition suggests, we have ways of expressing multiple associated concepts in a single object and the most simple of these is the vector. Perhaps the most common example of the difference between a vector and a scalar is that between speed and velocity. If we say that a car is travelling at 20 km h^{-1}, this is a scalar, but if we say it is going 20 km h^{-1} due North West, we have a vector. Vectors can also be thought of as a list, in which the order of the contents matters.

In this section of the course we will introduce some formal mathematical notation and rules for the manipulation (addition, multiplication, *etc.*) of these vector quantities. It is often very useful to represent vectors and their associated processes as lines on a 2D plane, to reinforce the underlying theory with a physical intuition. However, when the problem is in 3 (or more) dimensions, illustrating these concepts can become very difficult and you will have to rely on the rules you learned in 2D. Several different notational styles can be used with vectors, beyond the explicit coordinate representation (v_1, v_2, v_3), such as

Underlined : $\underline{v}$	Bold lower case : $\mathbf{v}$	Column : $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$
Arrow over points : $\overrightarrow{AB}$	Vector arrow : $\vec{v}$	Unit vectors : $v_1\hat{\mathbf{i}} + v_2\hat{\mathbf{j}} + v_3\hat{\mathbf{k}}$

2.1 Co-ordinate geometry



The theory of co-ordinate geometry is very closely associated with vectors, so let's start by discussing the adjacent 3D illustration in x, y, z-space. Here you can see a vector from the origin to point $P(x, y, z)$. You can calculate the length (or *magnitude*) of the line $\overrightarrow{OP}$ using Pythagoras theorem.

$$|\overrightarrow{OP}| = r = \sqrt{x^2 + y^2 + z^2}$$

You can also use trigonometry to work out all it associated angles.

$$\angle POx = \arccos(x/r)$$
$$\angle POy = \arccos(y/r)$$
$$\angle POz = \arccos(z/r)$$

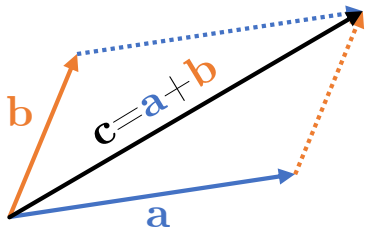
Importantly, if two vectors have the same length and the same direction, they are identical.

2.1.1 Unit vectors

A vector whose length/magnitude/modulus is 1, (*i.e.*, $|\underline{v}| = 1$) is called a "unit vector" and is often written with a little hat, $\hat{v}$.

It is sometimes convenient to describe vectors in terms of these unit vectors and corresponding coefficients. The unit vectors oriented along the three orthogonal Cartesian axis are

$$\hat{\mathbf{i}} = (1, 0, 0) \qquad \hat{\mathbf{j}} = (0, 1, 0) \qquad \hat{\mathbf{k}} = (0, 0, 1)$$



You can also find the unit vector of any arbitrary vector by dividing it by its own magnitude (*e.g.* $\hat{\mathbf{a}} = \frac{\mathbf{a}}{|\mathbf{a}|}$).

This concept makes vector **addition** straightforward, as we simply add each direction separately. For example, if we wanted to find the "resultant force", $\mathbf{c}$, when the two forces $\mathbf{a} = 3\hat{\mathbf{i}} + 2\hat{\mathbf{j}} - 1\hat{\mathbf{k}}$ and $\mathbf{b} = -1\hat{\mathbf{i}} - 2\hat{\mathbf{j}} + 4\hat{\mathbf{k}}$ are both applied to the same point, we just add each component to get $\mathbf{c} = 2\hat{\mathbf{i}} + 3\hat{\mathbf{k}}$.

Another classic real world example is to imagine a boat trying to cross a river from West to East. Its motor allows it to travel at a speed of $|\mathbf{b}| = 4$ m s^{-1} *relative to the water* and the river is flowing from North to South at $|\mathbf{r}| = 3$ m s^{-1} *relative to the land* . By first converting these two pieces of information into vectors, we can then add them together and find the velocity of the boat relative to the bank. By taking East to be in the direction of $\hat{\mathbf{i}}$ and North to be in the direction of $\hat{\mathbf{j}}$, we can rewrite the problem as $\mathbf{b} = 4\hat{\mathbf{i}}$ and $\mathbf{r} = -3\hat{\mathbf{j}}$. So, the velocity of our boat, $\mathbf{v}$, must be $\mathbf{v} = \mathbf{b} + \mathbf{r} = 4\hat{\mathbf{i}} - 3\hat{\mathbf{j}}$, which can also be described as a speed of 5 m s^{-1} at a bearing of 127°, relative to the land.

2.1.2 Basis vectors

This is a big topic in its own right, but it's worth just mentioning it as a follow on from the previous section. As we saw above, you can think of the vector (a, b) in terms of the scalar a multiplied by the unit vector $\hat{\mathbf{i}}$ and the scalar b multiplied by the unit vector $\hat{\mathbf{j}}$. What's interesting to consider is that we tend to implicitly assume the use of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ as our basis vectors in 2D, but we don't have to. We could, for example, choose the vectors $\vec{v} = \hat{\mathbf{i}} + 3\hat{\mathbf{j}}$ and $\vec{w} = 2\hat{\mathbf{i}} - \hat{\mathbf{j}}$ as our basis and still be able to reach every point on the 2D plane using linear combinations of these two vectors. For example, in our $\vec{v}$, $\vec{w}$ basis, the vector $(3, -1)$ is equivalent to $3\vec{v} - \vec{w} = \hat{\mathbf{i}} + 10\hat{\mathbf{j}}$.

However, if we mistakenly choose a pair of basis vectors that were parallel, we would no longer be able to access the whole 2D plane, but just a line instead (no better than just using a single

vector). The region accessible by the linear combination of vectors is called the "span". If two vectors are parallel, it means they are pointing in the same direction, but their lengths can be different. This means you can write the expression $\underline{a} = \lambda \underline{b}$ and find a value for lambda. If $\lambda > 0$ they are parallel, however if $\lambda < 0$, they are "anti-parallel", which means they are pointing in exactly opposite directions.

2.2 Vector multiplication

There are three methods of vector multiplication that we will cover in this course, which we will consider by applying them to the vectors $\mathbf{a} = (1, 2, 3)$ and $\mathbf{b} = (4, 5, 6)$. They are stated concisely below (without engineering context), just to have them all in one place.

$\mathbf{a} \circ \mathbf{b}$ - **Entrywise product** or **Hadamard product** - This is where you multiply each pair of terms in the two vectors to yield a new vector: $\mathbf{a} \circ \mathbf{b} = (a_1 b_1, a_2 b_2, a_3 b_3) = (4, 10, 18)$ (NB. Can only be performed on vectors of the same size and returns a vector of the same size).

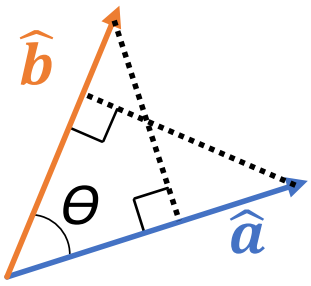
$\mathbf{a} \cdot \mathbf{b}$ - **Dot product** or **Scalar product** or **Inner product** - This is where you first perform the entrywise product and then add all the terms in the resulting vector together: $\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3 = 4 + 10 + 18 = 32$ (NB. Returns a scalar).

$\mathbf{a} \times \mathbf{b}$ - **Cross product** or **Vector product** - This is where you find the difference of the products of the cross matched terms either side of the current index (explained again below!): $\mathbf{a} \times \mathbf{b} = (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1) = (12 - 15, 12 - 6, 5 - 8) = (-3, 6, -3)$. (NB. Returns same size vector and only possible with 3D vectors (or 7D, but we won't be using these!)).

You can now blindly apply these three definitions without much difficulty, but if we take a closer look at the dot product and cross product, we can start to understand what they mean.

2.2.1 Dot product

The dot product can be thought of as a kind of directional multiplication, where for a pair of vectors, the products of their components in each dimension are found and then added together. Perhaps the most intuitive way to understand the dot product is through the concept of *projection*, where we relate any two vectors by the shadow they cast on each other if a light was shone orthogonally to the vector being shadowed (NB. $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$).



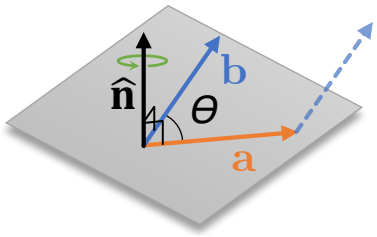
Thinking about this in terms of trigonometry, as each shadow's path is perpendicular to a vector, they must form a right angled triangle.This means we can now write down the standard definition of the dot product in terms of the angle, θ between the two vectors; however, I think the interpretation of this concept becomes clearer when it is rearranged slightly

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos(\theta) \quad \xrightarrow{\text{rearrange}} \quad \frac{\mathbf{a}}{|\mathbf{a}|} \cdot \frac{\mathbf{b}}{|\mathbf{b}|} = \hat{\mathbf{a}} \cdot \hat{\mathbf{b}} = \cos(\theta)$$

such that, after cancelling out the magnitude of the two vectors, you are just comparing the two corresponding unit vectors, giving you the angle between them. One of the most useful features of the dot product is as a convenient test for orthogonality (and therefore linear independence), which can simply be represented as when $\theta = 90°$. It follows that because $\cos(90) = 0$, then $\hat{\mathbf{a}} \cdot \hat{\mathbf{b}}$ must also equal zero if $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are at right angles to each other.

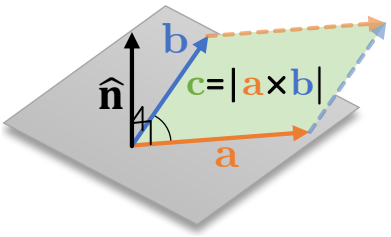
2.2.2 Cross product

There are two key applications of the cross product that you are likely to encounter as an engineer. Firstly, as a method for calculating rotational effects (such as moments) and secondly for calculating parallelogram areas. NB, unlike the dot product, the result of a cross product is a vector.



When dealing with moments, imagine a lever, $\mathbf{a}$, connected to the origin, being acted on at its tip by a force vector, $\mathbf{b}$. When you look at the diagram, this can seem a bit confusing because both vectors are coming out of the origin, as this is how they will be typically represented; however, the diagram also shows you a dashed line of what this physical interpretation is implying. So, because $\mathbf{a}$ and $\mathbf{b}$ are clearly not orthogonal (*i.e.*, $\theta \neq 90°$), only some of the force will be converted into a rotational moment. This moment is represented as a vector normal to the plane described by $\mathbf{a}$ and $\mathbf{b}$, with the direction implying the direction of rotation according to the right hand rule. There are two things to notice here, firstly that the order now matters (*i.e.*, $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$) and secondly if $\theta = 0$, then the cross product is zero, which makes sense as there will be no rotational moment. (NB. $\hat{\mathbf{i}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}}$).

For area calculation, you simply imagine a parallelogram contained by two pairs of the two vectors, all connected up! NB, to help visualise, we show the arrows connected as if adding, but they may have different units, so the sum is *meaningless*.

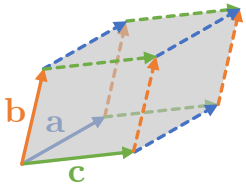


Once again, a more general interpretation of the concept is shown through the same simple rearrangement we used for the dot product. The cross product of two unit vectors gives you a new vector in the direction normal to their plane with a length equal to the sine of their angle.

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}| \sin(\theta)\hat{\mathbf{n}} \quad \xrightarrow{\text{rearrange}} \quad \frac{\mathbf{a}}{|\mathbf{a}|} \times \frac{\mathbf{b}}{|\mathbf{b}|} = \hat{\mathbf{a}} \times \hat{\mathbf{b}} = \sin(\theta)\hat{\mathbf{n}}$$

2.2.3 Triple scalar product

The last case is simply a combination of the dot and cross product, typically referred to as the triple scalar product. For the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, the triple product will give you the volume of the parallelepiped mapped by the three sets of parallel edges made from the three vectors. Clearly, if $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are not all linearly independent, then the volume will be zero. Notice the order of the dot product doesn't matter, but

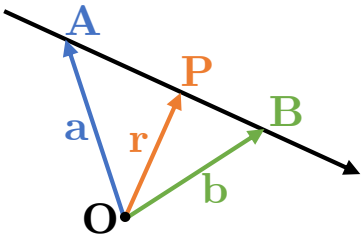


the cross product does!

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\equiv\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})\equiv-(\mathbf{b}\times\mathbf{a})\cdot\mathbf{c}\equiv\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})\equiv(\mathbf{c}\times\mathbf{a})\cdot\mathbf{b}$$

2.3 Vector equation of a line

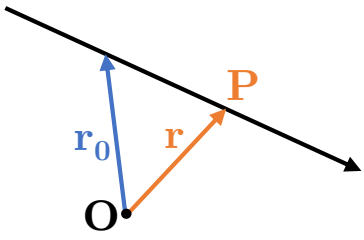
If we are going to represent the world with vectors, we're sometimes going to need to define lines or planes that do not pass through the origin. Imagine, for example, that we define the origin as the location of a telescope and want to model the motion of a passing asteroid... hopefully one that doesn't pass through the origin!



Consider the line in the adjacent figure that goes through points A and B. We can describe the direction of this vector by noticing that $\overrightarrow{AB}=\overrightarrow{OB}-\overrightarrow{OA}$. This allows us to write an expression for the family of vectors which take us to every point on this line

$$P=\overrightarrow{OA}+\lambda\overrightarrow{AB}\qquad\xrightarrow{\text{alternatively}}\qquad P=\overrightarrow{OB}+\kappa\overrightarrow{AB}$$

where λ and κ are just scalar parameters. It's important to realise that these two equations both describe the same black line, and whats more we could have picked any point that touches this line to construct an expression. Notice, however, that for the first formulation, when $\lambda=0$ we're at point A, whereas when $\kappa=0$ we're at point B. Similarly, at $\lambda=1$ we're at point B, but when $\kappa=1$ we're somewhere on the black line to the right of B. To get to A in the second formulation, we'd need $\kappa=-1$.

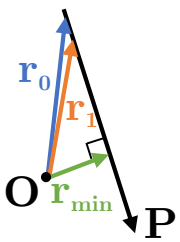


If we wish to interpret our line as the sequence of points traced out by an asteroid, then we can call our initial position $\mathbf{r}_0$ and our velocity $\mathbf{v}$. For a constant velocity model, we can simply add the vector of the starting location, to the product of the velocity vector and the time.

$$\mathbf{r}=\mathbf{r}_0+t\mathbf{v}$$

Now that we have the engineering interpretation, we can use some of the methods that we learned above to solve engineering problems. A radar station is tracking a high speed vehicle. When it's first spotted, it is at position (32, 45) km relative to the station. One minute later, the vehicle is at position (29, 41) km from the station. What's the vehicle's speed?

The vector representing the distance between the two observations is $(29,41)-(32,45)=(-3,-4)$ km, so the magnitude of this vector is the scalar distance travelled, $|(-3,-4)|=\sqrt{(-3)^2+(-4)^2}=5$ km. If this distance took one minute, then the speed of the vehicle must be $5\times60=300$ km h^{-1}. We can now also write a vector expression for the vehicle's location.



$$\mathbf{r}(t)=\begin{bmatrix}32\\45\end{bmatrix}+t\begin{bmatrix}-180\\-240\end{bmatrix}$$

where the position vector $\mathbf{r}$ is measured in kilometres and the time t is a scalar measured in hours. Assuming that the jet maintains a constant velocity, we can also find the minimum expected distance between the vehicle and the station. This will occur when the vehicle's location vector relative to the radar station, $\mathbf{r}$, is orthogonal to its velocity vector (see the green and black lines in the figure above). One way to think about this is that the vehicle is at its closest when it's drive has to look out her side window to see the station.

We can construct an expression for this using the dot product, which should equal zero when $\theta = 90°$.

$$\mathbf{r}_{\text{min}} \cdot \begin{bmatrix} -180 \\ -240 \end{bmatrix} = \begin{bmatrix} 32 - 180t \\ 45 - 240t \end{bmatrix} \cdot \begin{bmatrix} -180 \\ -240 \end{bmatrix} = -5760 + 32400t - 10800 + 57600t$$
$$= -16560 + 90000t = 0$$

So, $16560 = 90000t$ meaning that the vehicle will be closest to the station at $t = 0.184$ hours after its initial sighting

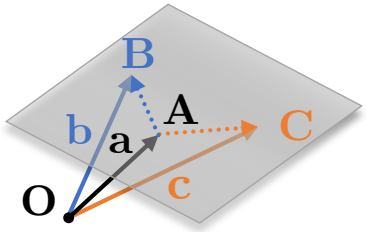
$$\mathbf{r} = \begin{bmatrix} 32 \\ 45 \end{bmatrix} + 0.184 \begin{bmatrix} -180 \\ -240 \end{bmatrix} = \begin{bmatrix} -1.12 \\ 0.84 \end{bmatrix} \quad \Rightarrow \quad |\mathbf{r}_{\text{min}}| = 1.4 \text{ km}$$

2.3.1 Equations of planes

As we saw in our section on the cross product, we can define the orientation of a plane using just **one vector** (the normal vector). However, following on from our discussion on basis vectors we can build an alternative description that allows us to access each point on this plane more directly using **two vectors**, as long as this plane passes through the origin. Furthermore, as with the equation of a line, if we'd like to describe a plane that doesn't pass through the origin, we are going to need **three vectors**.

Perhaps the simplest way to think of an expression for a plane is by first writing the equation of a line and then adding another term which is just a second line with its own scalar parameter μ (NB. This is only a plane if $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are not parallel).

$$P = \overrightarrow{OA} + \lambda\overrightarrow{AB} + \mu\overrightarrow{AC}$$



You can think of this as starting at point A and travelling λ steps along $\overrightarrow{AB}$ and μ steps along $\overrightarrow{AC}$.

Alternatively, you can think of the equation of a plane (more standard definition) by considering the fact that an arbitrary line on the plane, $\overrightarrow{AB}$=(**b**-**a**), must have a dot product of zero with the normal to the plane, $\mathbf{n}$, (*i.e.*, they must be orthogonal).

$$(\mathbf{b} - \mathbf{a}) \cdot \mathbf{n} = 0 \quad \xrightarrow{\text{rearrange}} \quad \mathbf{b} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n}$$

If we assume that the coordinates of **a** are fixed, but we let **b** move around the plane to the allowed values of x, y and z, then we can re write the expression in Cartesian form as

$$\alpha x + \beta y + \gamma z = p$$

where p is the dot product of our fixed point, **a**, and the normal, **n**.

For example, let's now find the equation of a plane that passes through the points A=(3,2,0), B=(1,3,-1) and C=(0,-2,3).

Clearly, the lines from $\overrightarrow{\text{AB}}$=(-2,1,-1) and $\overrightarrow{\text{AC}}$=(-3,-4,3) must both be parallel to the plane. So, using the cross product, we can calculate a normal vector to the plane

$$\begin{aligned}\mathbf{n} &= \overrightarrow{AB} \times \overrightarrow{AC} = (-2,1,-1) \times (-3,-4,3) \\ &= \det\left(\begin{bmatrix} \hat{\mathbf{i}} & -2 & -3 \\ \hat{\mathbf{j}} & 1 & -4 \\ \hat{\mathbf{k}} & -1 & 3 \end{bmatrix}\right) = (3-4)\hat{\mathbf{i}} + (3+6)\hat{\mathbf{j}} + (8+3)\hat{\mathbf{k}} = (-1,9,11)\end{aligned}$$

So, our plane equation can now be written in the $\mathbf{b} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n}$ form.

$$\mathbf{b} \cdot (-1,9,11) = (3,2,0) \cdot (-1,9,11) = -3 + 18 + 0 = 15$$

Therefore, $(x,y,z) \cdot (-1,9,11) = 15$, which can be written $-x + 9y + 11z = 15$

You can check that this is correct by substituting the original points and making sure that they all satisfy the equation.

This is about as far as our discussion of vectors can go without introducing matrices (I already sneaked the determinant in above... if you don't know what this is then check out the following chapter!).

Chapter 3

Matrices

What is a Matrix?

A *matrix* is a rectangular array of *elements*. These elements could be anything, but they are usually numbers when we discuss them as engineers. The core idea of matrices is *ordering*, by which we mean that the *location* of each element in our matrix tells us something about it. Compare a shopping list, which is an unordered vector where the location of each item on it doesn't really matter; to a digital image, which is a matrix of numbers representing colours, where clearly the location matters a lot!

$$\begin{pmatrix} 1 & 3 & 0 \\ -2 & 8 & 2 \\ 4 & 0 & -1 \\ \frac{1}{2} & 0 & 117 \end{pmatrix}$$

The above matrix is a 4×3 matrix, *i.e.*, it has four rows and three columns, so 12 elements in total.

We use matrices in mathematics and engineering because often we need to deal with several variables at once - *e.g.* coordinate vectors, as we saw in the previous chapter are a type of matrix which only has either a single column or a single row. In the following chapter on "linear transformations" we'll see a particular interpretation of the structure of a matrix relating to operations that transform vectors, but in this chapter we're simply going to lay out the language and start building up a toolbox.

It turns out that many operations that are needed to be performed on coordinates of points are **linear operations** and so can be organised in terms of rectangular arrays of numbers (matrices). Then we find that matrices themselves can, under certain conditions, be added, subtracted and multiplied so that there arises a whole new set of algebraic rules for their manipulation

In general, a matrix 'A' with dimensions of $(n \times m)$ looks like:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,m-1} & a_{1,m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2,m-1} & a_{2,m} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3,m-1} & a_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & a_{n-1,3} & \dots & a_{n-1,m-1} & a_{n-1,m} \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & a_{n,m-1} & a_{n,m} \end{pmatrix}$$

It is convention to denote entries within a matrix a_{ij}, where i denotes the row and j denotes the column, also a capital letter is typically used to define the matrix itself.

The following sections describe methods for calculating various matrix operations; **however**, since the advent of modern computing, no one does this by hand any more... So you might wonder why we're going over it! The answer is partly so that you know what's going on in the machine and partly to make you appreciate how incredible computing power is!

3.1 Matrix Operations

3.1.1 Addition

It is possible to add two matrices together, but *only if they have the same dimensions.* We simply add the corresponding entries to form a new matrix of the same size:

$$\begin{pmatrix} 1 & 3 & 0 \\ -2 & 8 & 2 \\ 4 & 0 & -1 \\ \frac{1}{2} & 0 & 117 \end{pmatrix} + \begin{pmatrix} 4 & 0 & 1 \\ 0 & -8 & -3 \\ 5 & 1 & -2 \\ \frac{1}{2} & 1 & -50 \end{pmatrix} = \begin{pmatrix} 5 & 3 & 1 \\ -2 & 0 & -1 \\ 9 & 1 & -3 \\ 1 & 1 & 67 \end{pmatrix}$$

If two matrices do not have the same dimensions they cannot be added, or we say the sum is 'not defined'.

3.1.2 Multiplication or "Rows times cols"

When multiplying matrices, keep the following in mind: if the number of columns of the first matrix equals the number of rows of the second, then you can proceed. The process for multiplying is as follows: to find entry (n, m) of the resulting matrix, take row n of the first matrix and column m of the second matrix and then find the sum of the product of each pair of entries (*i.e.*, the dot product). For example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 & 7 \\ 8 & 9 & 0 \end{pmatrix} = \begin{pmatrix} (1 \times 5) + (2 \times 8) & (1 \times 6) + (2 \times 9) & (1 \times 7) + (2 \times 0) \\ (3 \times 5) + (4 \times 8) & (3 \times 6) + (4 \times 9) & (3 \times 7) + (4 \times 0) \end{pmatrix}$$

$$= \begin{pmatrix} 21 & 24 & 7 \\ 47 & 54 & 21 \end{pmatrix}$$

Symbolically, if we have the matrices A and B:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2,m} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,1} & a_{n-1,2} & a_{n-1,3} & \dots & a_{n-1,m} \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & a_{n,m} \end{pmatrix} \quad \& \quad B = \begin{pmatrix} b_1 & b_2 & b_3 & \dots & b_{1,q} \\ b_{21} & b_{22} & b_{23} & \dots & b_{2,q} \\ b_{31} & b_{32} & b_{33} & \dots & b_{3,q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{p-1,1} & b_{p-1,2} & b_{p-1,3} & \dots & b_{p-1,q} \\ b_{p,1} & b_{p,2} & b_{p,3} & \dots & b_{p,q} \end{pmatrix}$$

Then the product AB is given by:

$$AB = \begin{pmatrix} \sum_{i=1}^{m} a_{1i}b_{i1} & \sum_{i=1}^{m} a_{1i}b_{i2} & \dots & \sum_{i=1}^{m} a_{1i}b_{iq} \\ \sum_{i=1}^{m} a_{2i}b_{i1} & \sum_{i=1}^{m} a_{2i}b_{i2} & \dots & \sum_{i=1}^{m} a_{2i}b_{iq} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} a_{ni}b_{i1} & \sum_{i=1}^{m} a_{ni}b_{i2} & \dots & \sum_{i=1}^{m} a_{ni}b_{iq} \end{pmatrix}$$

Where $\sum_{i=1}^{m} a_{1i}b_{i1}$ stands for $a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} + \cdots + a_{1n}b_{n1}$, etc. Note that we must have $m = p$ such that the number of columns in the first matrix must equal the number of rows in the second; otherwise, we say the product is *undefined*.

The quick way to check whether a sequence of operations is allowed for matrices of different sizes is the following. Simply write down their dimensions as "rows×cols" and check that, wherever two adjacent matrices are multiplied, the adjacent dimensions are the same. Consider:

$$A = (2 \times \underline{3)(3} \times \underline{5)(5} \times \underline{1)(1} \times 7) \qquad \& \qquad B = (2 \times \underline{\underline{4)(3}} \times \underline{5)(5} \times \underline{1)(1} \times 7)$$

The operation for A would work and the dimensions of the resulting matrix are those at the outer most of the operation, *i.e.*, (2×7), whereas the operation for B is undefined.

3.1.3 Scalar Multiplication

Another form of matrix multiplication is called *scalar multiplication*. This involves simply multiplying each entry of the matrix:

$$3\begin{pmatrix} -1 & 2 & 0 \\ 4 & 1 & -2 \end{pmatrix} = \begin{pmatrix} -3 & 6 & 0 \\ 12 & 3 & -6 \end{pmatrix}$$

3.2 Rules of Addition and Multiplication

There are rules which matrix addition and multiplication obeys:

Associative Addition	$(A + B) + C = A + (B + C)$
Associative Multiplication	(AB)C = A(BC)
Commutative Addition	$A + B = B + A$
Non-Commutative Multiplication	$AB \neq BA$
Distributive	$A(B + C) = AB + AC$
Moving Constants	$A(\lambda B) = \lambda(AB)$

Example - Consider the following matrices:

$$A = \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 4 \\ 1 & -2 \end{pmatrix}, \quad C = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Non-commutative behaviour can clearly be shown by comparing AB to BA:

$$AB = \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} -1 & 4 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} (1 \times -1) + (0 \times 1) & (1 \times 4) + (0 \times -2) \\ (3 \times -1) + (2 \times 1) & (3 \times 4) + (2 \times -2) \end{pmatrix} = \begin{pmatrix} -1 & 4 \\ -1 & 8 \end{pmatrix}$$

$$BA = \begin{pmatrix} -1 & 4 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} (-1 \times 1) + (4 \times 3) & (-1 \times 0) + (0 \times -2) \\ (1 \times 1) + (-2 \times 3) & (1 \times 0) + (-2 \times 2) \end{pmatrix} = \begin{pmatrix} 11 & 8 \\ -5 & -4 \end{pmatrix}$$

And *distributive* behaviour (i.e. $A(B + C) = AB = AC$) may be shown with the following:

$$\begin{aligned} AC &= \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 12 \end{pmatrix} \\ BC &= \begin{pmatrix} -1 & 4 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 10 \\ -4 \end{pmatrix} \\ AC + BC &= \begin{pmatrix} 2 \\ 12 \end{pmatrix} + \begin{pmatrix} 10 \\ -4 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} \\ A + B &= \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix} + \begin{pmatrix} -1 & 4 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix} \\ (A + B)C &= \begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} \end{aligned}$$

Noting that $(A + B)C = AC + BC$.

3.3 Transpose

Another operation on matrices is the *transpose*. This reverses the rows and columns, or equivalently, reflects the matrix along the leading diagonal. The transpose of A is normally written A^t, thus:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,m-1} & a_{1,m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2,m-1} & a_{2,m} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3,m-1} & a_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & a_{n-1,3} & \dots & a_{n-1,m-1} & a_{n-1,m} \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & a_{n,m-1} & a_{n,m} \end{pmatrix}$$

$$A^t = \begin{pmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{m-1,1} & a_{m,1} \\ a_{12} & a_{22} & a_{32} & \dots & a_{m-1,2} & a_{m,2} \\ a_{13} & a_{23} & a_{33} & \dots & a_{m-1,3} & a_{m,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{1,n-1} & a_{2,n-1} & a_{3,n-1} & \dots & a_{m-1,n-1} & a_{m,n-1} \\ a_{1,n} & a_{2,n} & a_{3,n} & \dots & a_{m-1,n} & a_{m,n} \end{pmatrix}$$

Note that the transpose of a $(n \times m)$ matrix is a $(m \times n)$ matrix.

Example

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

3.4 Square matrices

3.4.1 Identity Matrix

A square matrix is a matrix with the same number of rows as columns, *i.e.*, $(n \times n)$. There are a number of special square matrices, however, a particularly important one is the 'Identity Matrix'. This matrix fulfils a similar role to the number '1' in calculations such that multiplying a matrix by the idenity matrix results in no change to the matrix:

$$AI_{n\times n} = A \qquad \text{and} \qquad I_{n\times n}A = A$$

The identity matrix is a $(n \times n)$ matrix with the number one across the leading diagonal and zeros in every other position:

$$I_{2\times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_{3\times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad I_{4\times 4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad etc....$$

Whereby the size $(n \times n)$ of the identity matrix is generally inherited from the other matrices involved in the operation unless otherwise stated.

Example - Let:

$$A = \begin{pmatrix} 1 & 7 \\ 3 & 2 \end{pmatrix}, \; I = I_{2\times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$AI = \begin{pmatrix} 1 & 7 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} (1 \times 1) + (7 \times 0) & (1 \times 0) + (7 \times 1) \\ (3 \times 1) + (2 \times 0) & (3 \times 0) + (2 \times 1) \end{pmatrix} = \begin{pmatrix} 1 & 7 \\ 3 & 2 \end{pmatrix}$$

$$IA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 7 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} (1 \times 1) + (0 \times 3) & (1 \times 7) + (0 \times 2) \\ (0 \times 1) + (1 \times 3) & (0 \times 7) + (1 \times 2) \end{pmatrix} = \begin{pmatrix} 1 & 7 \\ 3 & 2 \end{pmatrix}$$

Showing that matrix A has not been altered.

3.4.2 Determinants

The determinant is a property of a square matrix and is a scalar number defined by the entries within the matrix. It is very useful for calculating 'how much larger' (or smaller) a linear transformation has changed the original value (we'll be talking about this in the next chapter, don't worry for now!). The determinant of a matrix A is denoted as either $\det(A)$ or $|A|$.

Calculating the determinant of a (2×2)-matrix

Calculating the determinant of a (2×2) matrix is a trivial process. For a matrix A:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

The determinant is simply the multiplication of $a \times d$ followed by the subtraction of $b \times c$:

$$\det(A) = |A| = ad - bc$$

Example - Consider the following:

$$\det \begin{pmatrix} 2 & 3 \\ 1 & 5 \end{pmatrix} = (2 \times 5) - (3 \times 1) = 7$$

$$\det \begin{pmatrix} -1 & 2 \\ 3 & -6 \end{pmatrix} = (-1 \times -6) - (2 \times 3) = 0$$

Minors and Cofactors

Before calculating the determinant of larger matrices the concept of *minors* and *cofactors* must be introduced.

Minors - Let A be a $(n \times n)$-matrix:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,n-1} & a_{1,n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2,n-1} & a_{2,n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3,n-1} & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & a_{n-1,3} & \dots & a_{n-1,n-1} & a_{n-1,n} \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & a_{n,n-1} & a_{n,n} \end{pmatrix}$$

Then, the minor m_{ij}, for each i and j is the determinant of the $(n-1 \times n-1)$-matrix obtained by deleting the i^{th} row and the j^{th} column. If you are left with a (2×2)-matrix the determinant is then taken. For example, in this notation:

$$m_{11} = \begin{pmatrix} a_{22} & a_{23} & \dots & a_{2,m-1} & a_{2,m} \\ a_{32} & a_{33} & \dots & a_{3,m-1} & a_{3,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,2} & a_{n-1,3} & \dots & a_{n-1,m-1} & a_{n-1,m} \\ a_{n,2} & a_{n,3} & \dots & a_{n,m-1} & a_{n,m} \end{pmatrix}$$

$$m_{21} = \begin{pmatrix} a_{12} & a_{13} & \dots & a_{1,m-1} & a_{1,m} \\ a_{32} & a_{33} & \dots & a_{3,m-1} & a_{3,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,2} & a_{n-1,3} & \dots & a_{n-1,m-1} & a_{n-1,m} \\ a_{n,2} & a_{n,3} & \dots & a_{n,m-1} & a_{n,m} \end{pmatrix}$$

Example - The minors of the matrix A:

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

$$m_{11} = \begin{vmatrix} 4 & 3 \\ 0 & -2 \end{vmatrix} = -8 \qquad m_{12} = \begin{vmatrix} 0 & 3 \\ -5 & -2 \end{vmatrix} = 15 \qquad m_{13} = \begin{vmatrix} 0 & 4 \\ -5 & 0 \end{vmatrix} = 20$$

$$m_{21} = \begin{vmatrix} 1 & -1 \\ 0 & -2 \end{vmatrix} = -2 \qquad m_{22} = \begin{vmatrix} 2 & -1 \\ -5 & -2 \end{vmatrix} = -9 \qquad m_{23} = \begin{vmatrix} 2 & 1 \\ -5 & 0 \end{vmatrix} = 5$$

$$m_{31} = \begin{vmatrix} 1 & -1 \\ 4 & 3 \end{vmatrix} = 7 \qquad m_{32} = \begin{vmatrix} 2 & -1 \\ 0 & 3 \end{vmatrix} = 6 \qquad m_{33} = \begin{vmatrix} 2 & 1 \\ 0 & 4 \end{vmatrix} = 8$$

Cofactors

The numbers called 'cofactors' are almost the same as minors, except some have a minus sign in accordance with the following pattern:

$$\begin{pmatrix} + & - & + & - & \dots \\ - & + & - & + & \dots \\ + & - & + & - & \dots \\ - & + & - & + & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The best way to remember this is as an 'alternating' pattern of positive and negative signs. Combining the minors from the previous example with this grid we get the cofactors:

$$c_{11} = m_{11} = -8 \qquad c_{12} = -m_{12} = -15 \qquad c_{13} = m_{13} = 20$$
$$c_{21} = -m_{21} = 2 \qquad c_{22} = m_{22} = -9 \qquad c_{23} = -m_{23} = -5$$
$$c_{31} = m_{31} = 7 \qquad c_{32} = -m_{32} = -6 \qquad c_{33} = m_{33} = 8$$

Calcluating the determinant of a (3×3)-matrix

In order to calculate the determinant of a (3×3)-matrix, choose *any* row or column. Then, multiply each entry by its corresponding cofactor, and add the three products. This gives the determinant.

Example - to show that *any* row or column may be used take matrix A:

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

Using the top row:

$$a_{11} = 2 \quad m_{11} = \begin{pmatrix} 2 & 1 & -1 \\ 0 & 4 & 3 \\ -5 & 0 & -2 \end{pmatrix} \quad \therefore \quad c_{11} = \begin{pmatrix} 2 & -1 & -1 \\ 0 & 4 & -3 \\ -5 & 0 & -2 \end{pmatrix}$$

$$a_{11}c_{11} = 2 \begin{vmatrix} 4 & -3 \\ 0 & -2 \end{vmatrix} = (2 \times -8) = -16$$

$$a_{12} = 1 \quad m_{12} = \begin{pmatrix} 2 & 1 & -1 \\ 0 & 4 & 3 \\ -5 & 0 & -2 \end{pmatrix} \quad \therefore \quad c_{12} = \begin{pmatrix} 2 & -1 & -1 \\ 0 & 4 & -3 \\ -5 & 0 & -2 \end{pmatrix}$$

$$a_{12}c_{12} = 1 \begin{vmatrix} 0 & -3 \\ -5 & -2 \end{vmatrix} = (1 \times -15) = -15$$

$$a_{13} = -1 \quad m_{13} = \begin{pmatrix} 2 & 1 & -1 \\ 0 & 4 & 3 \\ -5 & 0 & -2 \end{pmatrix} \quad \therefore \quad c_{13} = \begin{pmatrix} 2 & -1 & -1 \\ 0 & 4 & -3 \\ -5 & 0 & -2 \end{pmatrix}$$

$$a_{13}c_{13} = -1 \begin{vmatrix} 0 & 4 \\ -5 & 0 \end{vmatrix} = (-1 \times 20) = -20$$

Or more succinctly:

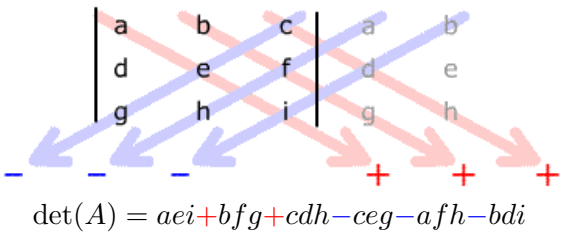
$$\det A = a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} = (2 \times -8) + (1 \times -15) + (-1 \times 20) = -51$$

Using the second column we can see the outcome is the same:

$$\det A = a_{12}c_{12} + a_{22}c_{22} + a_{32}c_{32} = (1 \times -15) + (4 \times -9) + (0 \times -6) = -51$$

Although it doesn't matter which is chosen, it is common for the top row to be chosen. Note that it is not necessary to work out all the minors (or cofactors), just the three necessary!

The entire process above is neatly summarised in the adjacent diagram for a 3×3 matrix. It's massively faster, so please make sure you understand what it's asking; however, it's still important that you understand the above so that you can tackle larger matrices.



$$\det(A) = aei + bfg + cdh - ceg - afh - bdi$$

Finding the determinant of an $(n \times n)$ matrix
The procedure for large matrices is exactly the same as for a (3×3) matrix: choose a row or column, multiply the entry by the corresponding cofactor and add them up. But of course each minor is itself the determinant of a $(n-1 \times n-1)$-matrix so for example, in a (4×4) determinant, it is necessary to do four (3×3) determinants - quite a lot of work... but lucky you will never have to do this, thanks to computers!

3.5 Inverses

Let A be an $(n \times n)$-matrix, and let I be the $(n \times n)$ identity matrix. Sometimes, there exists a matrix A^{-1} (called the *inverse* of A) with the property:

$$AA^{-1} = I = A^{-1}A$$

3.5.1 Inverse of a (2×2) Matrix

From the above definition, we can simply write down the general case for a 2×2 matrix leaving the inverse matrix as four unknowns.

$$AA^{-1} = I \qquad \text{then} \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

And now we're simply looking to find the 4 values of x in terms of $a, b, c, \& \ d$.

First we can write down the four simultaneous equations generated by applying the matrix multiplication.

$$\begin{aligned} ax_{11} + bx_{21} &= 1 \\ ax_{12} + bx_{22} &= 0 \\ cx_{11} + dx_{21} &= 0 \\ cx_{12} + dx_{22} &= 1 \end{aligned}$$

So by rearranging the first equation we can say $x_{11} = \frac{1 - bx_{21}}{a}$, which can then be substituted into the third equation to give us

$$c\left(\frac{1-bx_{21}}{a}\right)+dx_{21}=0 \qquad \xrightarrow{\text{rearrange}} \qquad x_{21}=\frac{-c}{ad-bc}$$

We can follow the same process for the other three unknown x term. which will yield the following expression

$$A^{-1}=\begin{pmatrix}\frac{d}{ad-bc} & \frac{-b}{ad-bc}\\ \frac{-c}{ad-bc} & \frac{a}{ad-bc}\end{pmatrix} \qquad \xrightarrow{\text{re-express}} \qquad A^{-1}=\frac{1}{\det(A)}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}$$

To check that our results is right, we can just multiply it back through by the original matrix:

$$\begin{aligned}A^{-1}\times A &= \frac{1}{ad-bc}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}\begin{pmatrix}a & b\\ c & d\end{pmatrix}\\ &= \frac{1}{ad-bc}\begin{pmatrix}da-bc & db-bd\\ -ca+ac & -cb+ad\end{pmatrix}\\ &= \frac{1}{ad-bc}\begin{pmatrix}ad-bc & 0\\ 0 & ad-bc\end{pmatrix}\\ &= \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}=I\end{aligned}$$

3.5.2 Inverse of (3×3) matrices (or higher)

We can follow the same process of simultaneous equation solving for 3×3 matrices and generate an explicit equation for the inverse; however the equation is complex, so it's typically easier to remember an algorithm for its calculation rather than finding it directly and this algorithm scales to an arbitrary $n\times n$ matrix.

Recall the definition of a *minor*: given an $(n\times n)$-matrix, A, the minor m_{ij} is the determinant of the $(n-1\times n-1)$-matrix by omitting the i^{th} row and j^{th} column. Then the cofactor is then the minor multiplied by the 'alternating' positive and negative patterns.

Example

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

The minors will therefore be:

$$\begin{aligned}m_{11}&=\begin{vmatrix}1 & 0\\ 2 & 1\end{vmatrix}=1 & m_{12}&=\begin{vmatrix}-2 & 0\\ 3 & 1\end{vmatrix}=-2 & m_{13}&=\begin{vmatrix}-2 & 1\\ 3 & 2\end{vmatrix}=-7\\ m_{21}&=\begin{vmatrix}0 & 4\\ 2 & 1\end{vmatrix}=-8 & m_{22}&=\begin{vmatrix}1 & 4\\ 3 & 1\end{vmatrix}=-11 & m_{23}&=\begin{vmatrix}1 & 0\\ 3 & 2\end{vmatrix}=2\\ m_{31}&=\begin{vmatrix}0 & 4\\ 1 & 0\end{vmatrix}=-4 & m_{32}&=\begin{vmatrix}1 & 4\\ -2 & 0\end{vmatrix}=8 & m_{33}&=\begin{vmatrix}1 & 0\\ -2 & 1\end{vmatrix}=1\end{aligned}$$

Meaning that the cofactor matrix will be:

$$\begin{pmatrix} 1 & 2 & -7 \\ 8 & -11 & -2 \\ -4 & -8 & 1 \end{pmatrix}$$

The next step is to take the transpose:

$$\begin{pmatrix} 1 & 8 & -4 \\ 2 & -11 & -8 \\ -7 & -2 & 1 \end{pmatrix}$$

Finally we can divide by the determinant (which is -27 in this case) to provide the inverse matrix:

$$A^{-1} = \frac{1}{-27}\begin{pmatrix} 1 & 8 & -4 \\ 2 & -11 & -8 \\ -7 & -2 & 1 \end{pmatrix}$$

Which if we check:

$$A^{-1} \times A = \frac{1}{-27}\begin{pmatrix} 1 & 8 & -4 \\ 2 & -11 & -8 \\ -7 & -2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 4 \\ -2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} = \frac{1}{-27}\begin{pmatrix} (1-16-12) & (0+8-8) & (4+0-4) \\ (2+22-24) & (0-11-16) & (8+0-8) \\ (-7+4+3) & (0-2+2) & (-28+0+1) \end{pmatrix}$$

$$= \frac{1}{-27}\begin{pmatrix} -27 & 0 & 0 \\ 0 & -27 & 0 \\ 0 & 0 & -27 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The same procedure works for $(n \times n)$ matrices:

1. Work out minors
2. Put in the $-$ signs to form the cofactors
3. Take the transpose
4. Divide by the determinant

Furthermore, an $(n \times n)$ matrix has an inverse if and only if the determinant is not zero. So it's a good idea to calculate the determinant *first*, just to see if the rest of the procedure is necessary.

3.6 Linear Systems

The method of solving simultaneous equations most school students are aware of involves re-arranging one equation such that it may be 'inserted' into the other. Try to solve the two equations:

$$2x + y = 3 \qquad \text{and} \qquad 5x + 3y = 7$$

By rearranging we can calculate the values of x and y:

$$\begin{array}{ccccc} y = 3 - 2x & \implies & 5x + 3(3 - 2x) = 7 \\ \therefore & x = 2 & \text{and} & y = -1 \end{array}$$

However, we can use matrices to provide a more systematic approach to solving simultaneous equations (despite this particular example being solved very simply without using matrices!). To do so we can re-write the equations in a slightly different way:

$$\begin{pmatrix} 2x & y \\ 5x & 3y \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

Now we can check that the first matrix is equal to the product:

$$\begin{pmatrix} 2x + y \\ 5x + 3y \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and so altogether we have a matrix equation:

$$\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

The next stage is to use the inverse of the (2×2)-matrix, so let's calculate that now.

$$\begin{aligned} \text{Let} \quad A &= \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \\ A^{-1} &= \frac{1}{(2 \times 3) - (1 \times 5)} \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} \end{aligned}$$

Now, we take the matrix equation above, and multiply by A^{-1}:

$$\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

$$\begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

Then, doing the multiplication:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (3 \times 3) + (-1 \times 7) \\ (-5 \times 3) + (2 \times 7) \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

and so $x = 2$ and $y = -1$ as required. So, provided we can work out the inverse of the matrix of coefficients, we can solve simultaneous equations. Finding efficient algorithms for inverting matrices is at the heart of computer science!

3.6.1 Larger Systems

The same thing works with 3 equations and x, y and z. Suppose we have:

$$x + 2y + 2z = -1$$
$$3y - 2z = 2$$
$$2x - y + 8z = 7$$

Then, the matrix form is, with a row for each equation and a column for each variable:

$$\begin{pmatrix} 1 & 2 & 2 \\ 0 & 3 & -2 \\ 2 & -1 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 7 \end{pmatrix}$$

Now, we denote the (3×3)-matrix by A and calculate the inverse of A. The minors are as follows:

$$m_{11} = \begin{vmatrix} 3 & -2 \\ -1 & 8 \end{vmatrix} = 22 \qquad m_{12} = \begin{vmatrix} 0 & -2 \\ 2 & 8 \end{vmatrix} = 4 \qquad m_{13} = \begin{vmatrix} 0 & 3 \\ 2 & -1 \end{vmatrix} = -6$$
$$m_{21} = \begin{vmatrix} 2 & 2 \\ -1 & 8 \end{vmatrix} = 18 \qquad m_{22} = \begin{vmatrix} 1 & 2 \\ 2 & 8 \end{vmatrix} = 4 \qquad m_{23} = \begin{vmatrix} 1 & 2 \\ 2 & -1 \end{vmatrix} = -5$$
$$m_{31} = \begin{vmatrix} 2 & 2 \\ 3 & -2 \end{vmatrix} = -10 \qquad m_{32} = \begin{vmatrix} 1 & 2 \\ 0 & -2 \end{vmatrix} = -2 \qquad m_{33} = \begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = 3$$

So we get the following matrix of cofactors:

$$\begin{pmatrix} 22 & -4 & -6 \\ -18 & 4 & 5 \\ -10 & 2 & 3 \end{pmatrix}$$

We can then calculate the determinant (taking the top row):

$$\det A = (1 \times 22) + (2 \times -4) + (2 \times -6) = 22 - 8 - 12 = 2$$

Then calculate the inverse (by taking the transpose and dividing by the determinant):

$$A^{-1} = \frac{1}{2} \begin{pmatrix} 22 & -18 & -10 \\ -4 & 4 & 2 \\ -6 & 5 & 3 \end{pmatrix} = \begin{pmatrix} 11 & -9 & -5 \\ -2 & 2 & 1 \\ -3 & \frac{5}{2} & \frac{3}{2} \end{pmatrix}$$

Now, we return to solving the simultaneous equations, we can multiply both sides by A^{-1}:

$$\begin{pmatrix} 1 & 2 & 2 \\ 0 & 3 & -2 \\ 2 & -1 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 7 \end{pmatrix}$$

$$\begin{pmatrix} 11 & -9 & -5 \\ -2 & 2 & 1 \\ -3 & \frac{5}{2} & \frac{3}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 0 & 3 & -2 \\ 2 & -1 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 11 & -9 & -5 \\ -2 & 2 & 1 \\ -3 & \frac{5}{2} & \frac{3}{2} \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 7 \end{pmatrix}$$

Given that $A^{-1}A = I$ and $IA = A$, we can show that:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (11 \times -1) + (-9 \times 2) + (-5 \times 7) \\ (-2 \times -1) + (2 \times 2) + (1 \times 7) \\ (-3 \times -1) + (\frac{5}{2} \times 2) + \frac{3}{2} \times 7 \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -64 \\ 13 \\ \frac{37}{2} \end{pmatrix}$$

Checking this with the original equations we can see that:

$$\begin{aligned} x + 2y + 2z = -64 + 2(13) + 2(\frac{37}{2}) = -64 + 26 + 37 &= -1 \\ 3y - 2z = 3(13) - 2(\frac{37}{2}) = 39 - 37 &= 2 \\ 2x - y + 8z = 2(-64) - 13 + 8(\frac{37}{2}) = -128 - 13 + 148 &= 7 \end{aligned}$$

3.7 Labels

The **trace** of a square matrix is the sum of the terms along its leading diagonal. The trace of a matrix is the same as the trace of it's transpose.

$$\text{tr}(A) = \text{tr} \begin{pmatrix} 2 & 1 & -1 \\ 0 & 4 & 3 \\ -5 & 0 & -3 \end{pmatrix} = 2 + 4 - 3 = 3$$

Matrices are often labelled in terms of the distributions of the elements as this can be convenient for giving you an idea of how to deal with it.

A **symmetric** matrix is a square matrix that is equal to its transpose; that is, it satisfies the condition $A = A^T$.

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

A **skew-symmetric** (or anti-symmetric) matrix is a square matrix whose transpose equals its negative; that is, it satisfies the condition $A^T = -A$.

$$\begin{pmatrix} 0 & 2 & -1 \\ -2 & 0 & -4 \\ 1 & 4 & 0 \end{pmatrix}$$

An **orthogonal** matrix has the identity $A^T A = AA^T = I$, which can be re-expressed as $A^T = A^{-1}$.

$$\frac{1}{3} \begin{pmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{pmatrix}$$

A **triangular** matrix is a square matrix that is either lower triangular or upper triangular. In a lower triangular matrix, all the entries above the main diagonal are zero. In a upper triangular matrix, all the entries below the main diagonal are zero.

$$LT = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 0 & 0 \\ 1 & 4 & 0 \end{pmatrix} \qquad \& \qquad UT = \begin{pmatrix} 0 & 5 & 9 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

A **diagonal** matrix is a square matrix that is both upper and lower triangular; that is all entries outside the main diagonal are zero.

$$\begin{pmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

A **singular** matrix is a square matrix that is not invertible. A square matrix is singular if and only if its determinant is 0.

$$\det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3 & 2 & 1 \end{pmatrix} = 5 + 36 + 24 - 12 - 8 - 45 = 0$$

Conclusions

This is the longest chapter in the DE1-MEM course, but in many ways I hope you can agree that it's the most straightforward as it's just about a set of tools and conventions used for manipulating matrices. Handy phrases like "rows times cols" will help you remember whether a multiplication operation is defined, but please rest assured that in the exam (and in your career!) you will never be asked to simply invert monstrous 5×5 matrices by hand, as this would only test your calculator skills.

Chapter 4

Linear Transformations

This chapter occasionally uses some slightly intimidating/complicated language to describe a relatively straightforward concept; however, it's important that you are exposed to the formal way to talk about these ideas so that you know what to ask the internet when you get stuck!

4.1 Demystifying linear transformations

"Vectors spaces" are simply spaces in which vectors can exist... almost too simple to bother explaining, but we're going to refer to them a lot in this chapter, so it's important that we're all starting from the same place! For example, we can talk about a 1D (one dimensional) vector space existing "in $\mathbb{R}$" (*i.e.*, in the real numbers), a 2D space in $\mathbb{R}^2$ or an nD space existing in $\mathbb{R}^n$. The adjacent figures show 1D, 2D and 3D vectors space, each containing three random vectors.

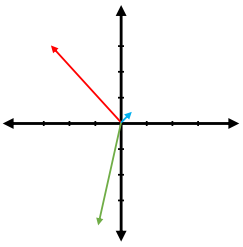
Definition - Linear transforms obey the following rules: Let V and W be vector spaces. A linear transformation (or "mapping" or "map") from V to W is a function $T : V \to W$ such that, for vectors v and w and scalars λ:

$$T(v + w) = T(v) + T(w)$$

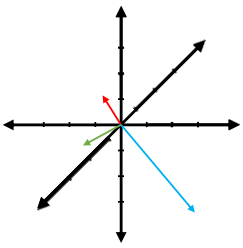
(*i.e.*, transform of the sum is the sum of the transforms)

$$T(\lambda v) = \lambda T(v)$$

(*i.e.*, transform of scaled vector equals the scale of the transformed vector)



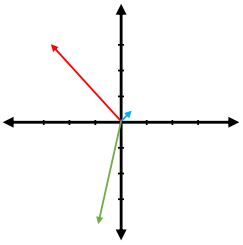
4.2 One dimension



Let's start by talking about transformations in $\mathbb{R}$ space. Look at the three vectors in the top figure on the side of this page. They are actually just defined by a single number, so really they are just scalars in disguise, but bear with me! Think about how you might transform the blue vector into the red vector? You'd simply multiply it by a scalar, which in this case would be the number 3. Similar, to get to the green vector, we could have multiplied the blue by -2. Essentially "scaling" is the only thing you can do in 1D, not very exciting!

4.3 Two dimensions

Things are significantly more interesting in 2 dimensions. As with the 1D case, we can still apply simple scalings and if that scaling is the same in all directions, we can just think about it as a scalar. So, we can transform the vector $\vec{a} = (3, 2)$ by a scaling factor of two to $2\vec{a} = (6, 4)$.



However, we can also represent this process using a (2×2) matrix by multiplying the identity matrix by the scaling parameter;

$$2I_{2\times 2}\begin{pmatrix}3\\2\end{pmatrix} = 2\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}3\\2\end{pmatrix} = \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}\begin{pmatrix}3\\2\end{pmatrix} = \begin{pmatrix}6\\4\end{pmatrix}$$

We refer to the matrix that we apply to our vector as our *transformation matrix*.

But what if we wanted to increase the width of our vector by a factor of 2, but the height by a factor of 5? This is still just a scaling, but more complicated than our first example.

$$\begin{pmatrix}2 & 0\\0 & 5\end{pmatrix}\begin{pmatrix}3\\2\end{pmatrix} = \begin{pmatrix}6\\10\end{pmatrix}$$

The answer to the question is shown to the right and we can check by simple matrix multiplication that it is correct. Shown below this is the inverse operation, which will hopefully seem fairly obvious to you, but also help you understand matrix inversion that we discussed last chapter.

$$\begin{pmatrix}3\\2\end{pmatrix} = \begin{pmatrix}1/2 & 0\\0 & 1/5\end{pmatrix}\begin{pmatrix}6\\10\end{pmatrix}$$

Even if this case seems fairly simple, what about more complicated examples? What if we had a vector and we wanted to rotate it around the origin or shear it horizontally, or both? To answer this question and essentially explain everything else about transforms, we need to talk about basis vectors. However, before we do this, it's worth pausing to mention that the concept "inverse" matrices that we met last chapter is easy to explain here.

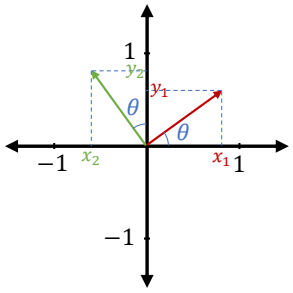
4.3.1 Basis vectors

In 2D, the standard basis vectors are $\hat{\mathbf{i}} = \begin{pmatrix}1\\0\end{pmatrix}$ and $\hat{\mathbf{j}} = \begin{pmatrix}0\\1\end{pmatrix}$. So writing $\vec{v} = \begin{pmatrix}3\\2\end{pmatrix}$ can also be thought of as $\vec{v} = 3\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$. What we haven't talked about before is that when you apply a matrix transformation, the columns of the matrix can be interpreted as the new basis for our transformed system. For example, in 2D, we've already seen from the scaling example above that the transformation $\begin{pmatrix}2 & 0\\0 & 5\end{pmatrix}$, moves $\hat{\mathbf{i}}$ from (1,0) to (2,0) and $\hat{\mathbf{j}}$ from (0,1) to (0,5).

If we wanted to build a transformation matrix which rotated all vectors 90 degrees anti-clockwise around the origin, we just need to think about what would happen to our basis vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. Picture it in you head... rotating $\hat{\mathbf{i}}$ by 90° anti-clockwise would make it point vertically upwards to (0,1) and rotating $\hat{\mathbf{j}}$ by 90° anti-clockwise would make it point in the negative direction on the horizontal axis to (-1,0)... so to build the 90° anti-clockwise transformation matrix, we simply write these two new basis vectors as the columns: of our transformation matrix $\begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}$.

4.3.2 Rotation

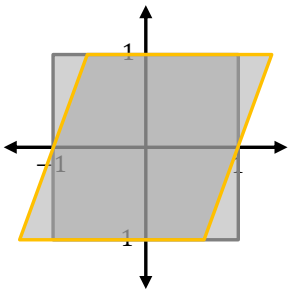
We can generalise the rotation in 2D to any angle we wish, by thinking about polar coordinates. In the adjacent figure, we can imagine that our basis vectors, $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$, have been transformed to the red and green vectors respectively. Based on our diagram and knowledge of trigonometry, the transformation matrix is:



$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

4.3.3 Shear

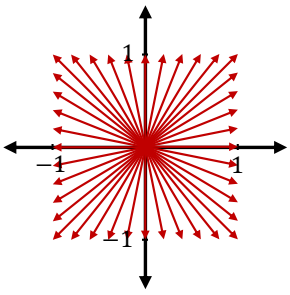
The adjacent figure illustrates the effect of shear on a square in a 2D space. The initial square has a dark grey border, but the square with a golden border shows it after shearing parallel to the horizontal axis. The shear matrices parallel to the horizontal and vertical axes are:



$$S_x = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \qquad \& \qquad S_y = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$$

Notice that if we are performing either of the above "pure" shear operations, then one of the basis vectors remains unchanged. For example, in our figure, the $\hat{\mathbf{i}}$ basis vector has not moved from it's original position. This is an important observation, which will be discussed in the next chapter.

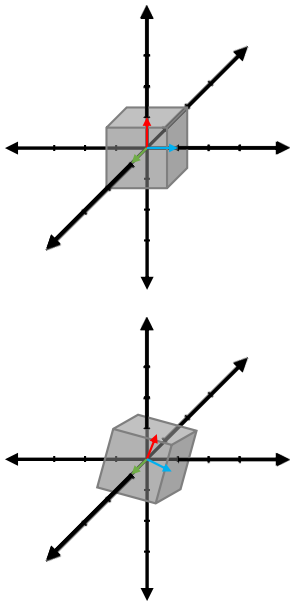
4.3.4 Visualising transformations



When thinking about basis change, it's is often more informative to imagine applying the effect of the transformation to a region of the same dimensionality as the space considered (*i.e.*, on an area in 2D or a volume in 3D). This is analogous to thinking about the effect on many vectors all at the same time, as in the top figure on the left, but drawing these vectors can be tedious, so instead we can just draw a simple shape, such as the square shown below it (any other shape would be fine, but a square is very convenient). We can now apply a transformation to this shape instead of to a particular vector and see the effect it has. Below are examples of scaling, rotation and shear.

4.4 Three dimensions

Nothing much new happens in 3D compared to 2D, except that the matrices are now larger and the figures are harder to draw. Scaling, rotation and shear, along with all the various combinations of these transformations, are still possible, although there are now more directions in which we can perform each of these actions.



The adjacent figures show a cube, as well as it's three basis vectors. The second figure the effect of a rotation around the axis aligned with the green vector. The structure of the rotation matrix now depends on the axis which we are rotating around. Although this is not something you'd be asked to calculate by hand. The only 3D transformation you should remember is the pure scaling example:

$$T = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

4.5 Determinant and Inverse

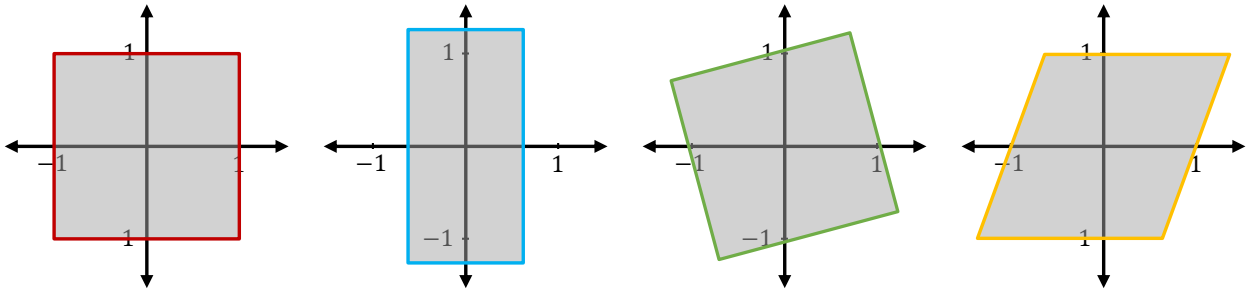
We met both the determinant and inverse in the previous chapter, but we didn't really talk much about what they did. In the geometrical interpretation of matrices, they turn out to be really quite obvious!

The inverse matrix is simple the transformation required to "undo" our initial transformation. So, if our transformation turns a square into a diamond, then the inverse will turn that diamond back into a square.

The determinant is also very straightforward as it simply describes the change in size (*i.e.*, area in 2D or volume in 3D) that is caused by our transformation. Consider the examples of a 90° rotation and a factor of 3 vertical scaling below.

$$\det \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = 1 \qquad \& \qquad \det \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} = 3$$

Clearly, a 90° rotation will not change the area of the square, so it has a determinant of 1; furthermore, all simple rotations will leave the area unchanged, which you can see by finding the determinant of the generalise rotation matrix.



For scaling, the change in area is going to be the product of the change in each dimension, which is exactly what the determinant would give us.

$$\det \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = \cos^2(\theta) + \sin^2(\theta) = 1$$

4.5.1 Nullspace (or Kernel)

So, now that we have a strong geometrical interpretation of matrices, let's return to the linear algebra perspective. Consider the linear system (expressed first as a system of equations, then in matrix form):

$$\left.\begin{array}{ll} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_n & = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_n & = b_2 \\ \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = b_m \end{array}\right\} \quad Ax = b$$

Various methods exist for finding the solution to large systems similar to the one above *by hand* and this is very commonly taught on undergraduate engineering courses. However, the method for solving this offers you little insight into what the solution means, it's simply a tool and in the computing age it's now a tool you will almost certainly never use. So, we won't be covering this in DE1-MEM.

What is worth answering are the following questions: What b can we solve this for? Is the solution unique?

To investigate this, we write what is called the "homogeneous equation", which is where we simply set $b = 0$, giving:

$$Ax = 0$$

Conveniently, if the homogeneous equation has just one solution, then so does $Ax = b$; if the homogeneous equation has many solutions, then so does $Ax = b$. Clearly, homogeneous systems *always* have at least one solution, because if you set $x = 0$, then it doesn't matter what A is.

The solution of $Ax = 0$ is called the nullspace or kernel of A and it gives you all the vectors by which you can multiply A by and get zero.

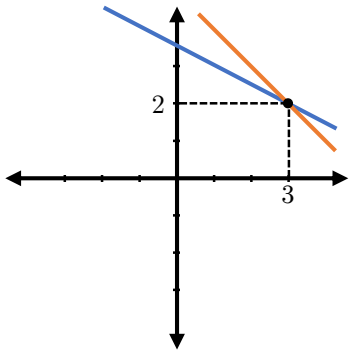
Example - Consider the linear system

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$$

We can then find the combinations of x & y which solve the homogeneous equation (*i.e.*, setting the right hand side to a zero vector) .

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \qquad \Rightarrow \qquad x + 2y = 0 \quad \& \quad x + y = 0$$

Solving this simple case as a pair of simultaneous equations, we can see that $x = 0$ & $y = 0$, so we know that $Ax = b$ will have a single unique solution. To solve the $Ax = b$ case, we can either frame it once again as a system of simultaneous equations or, based on of discussion of linear systems in the previous chapter, we can find the inverse of A.

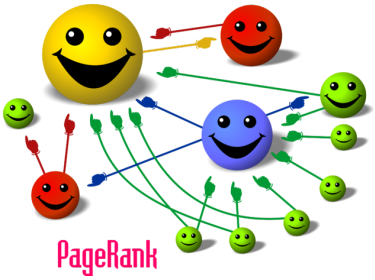


$$A^{-1} = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix} \quad \Rightarrow \quad A^{-1}b = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 7 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

In this case, the solution is $x = 3$ & $y = 2$.

Before we move on, we should briefly think again about the geometrical interpretation of the above. Consider for a pair of simultaneous equations given in the above example, each equation gives you a line on a 2D plane. The solution to the system exists if those lines cross at a point... but the only way that two lines could not cross is if they were parallel.

Similarly, in 3D, each row of our system would give us an equation of the form $a_1x+a_2y+a_3z = b_n$, which is the equation of a flat 2D surface in our 3D space. Our equation will only have a unique solution if those three surfaces intersect at a single point... but once again, they are guaranteed to do this somewhere unless some of them are parallel.



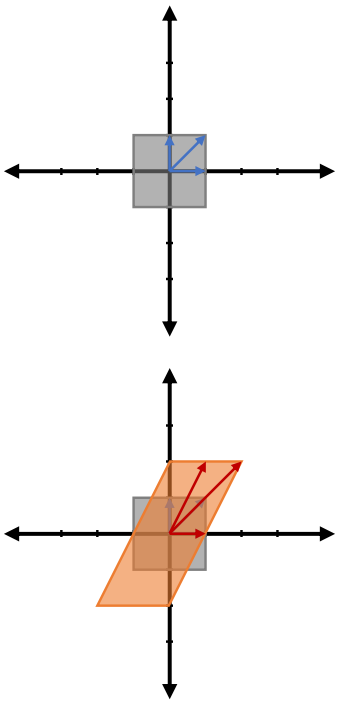
Chapter 5

Eigenproblems

5.1 Definitions

In the previous chapter we talked about the geometrical interpretation of linear transformations, as well as some of properties of transformation matrices. In this chapter we're just going to introduce a further concept for analysing matrices. This topic is often considered to be a fairly challenging and abstract part of undergraduate engineering, but I hope to show you in this introductory page that eigenproblems have a very clear physical interpretation.

The word "eigen" is perhaps most usefully translated from German as meaning "characteristic". So, when I say that we will be looking for *eigenvalues* and *eigenvectors*, this suggests that these values and vectors are in some sense characteristic of a particular matrix.



In the previous chapter we saw lots of examples of applying linear transformations to vectors spaces (primarily in $\mathbb{R}^2$). In the adjacent figure you can see our initial square in grey, as well as three vectors in blue. Beneath this you can see a new shape overlayed on top, which is the result of applying the transformation $T = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ (you should be able to look at a 2D transformation and write down the applied matrix).

Notice that of the three highlighted vectors, something different has happened to each one. The initially vertical $\hat{\mathbf{j}}$ vector has not only been stretched longer, but has also had it's direction changed; the initially diagonal (1,1) vector has doubled in length, but still points in the same direction; and the initially horizontal $\hat{\mathbf{i}}$ vector is still horizontal and its length is unchanged.

Eigenvectors are simply the vectors which, after applying a transformation, still lie on the same span (*i.e.*, have not change direction). Each eigenvector has a corresponding eigenvalue, which is just the amount that the vector has been stretched along its span by the transformation.

So, without doing any calculation, we can now say that the transformation T has two eigenvectors, v, with corresponding eigenvalues, λ, which are:

$$v_1 = (1, 0) \text{ with } \lambda_1 = 1 \qquad \text{and} \qquad v_2 = (1, 1) \text{ with } \lambda_2 = 2$$

5.2 Calculating Eigensolutions

On the previous page we found our eigenvectors and values by inspection, but this was only possible because I chose a convenient transformation whose eigensolutions are easy to see. However, we didn't check whether there were more than two solutions (there are not) and we also didn't build a proper method for less obvious cases (or cases in higher dimensions).

We can build a more formal definition by considering what we saw on the previous page, where we said that vectors would be considered to be eigenvectors if, after applying a transformation, they stayed on the same span, although their length was allowed to change. So, for an eigenvector, experiencing the transformation is no different from experiencing a simple scaling.

This means that we can write,

$$A\underline{x} = \lambda\underline{x}$$

where A is an $n \times n$ transformation, $\underline{x}$ is a vector and λ is a scalar parameter. The solutions to this equation are all the vectors, $\underline{x}$, which when transformed by A, it would be just the same as if they were just stretched by a factor of λ. So, $\underline{x}$ must be our eigenvectors and λ our corresponding eigenvalues.

For example, consider the following transformation matrix $A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}$

You can easily verify that

$$\begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

Hence, 3 is an eigenvalue of A. Vector $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ is an eigenvector of A corresponding to the scalar eigenvalue 3.

5.3 Finding All Eigenvalues

By recalling that multiplying a vector by a scalar is the equivalent to multiplying it by the identity matrix, I, times that scalar, we can re-express our eigenproblem to:

$$A\underline{x} = \lambda\underline{x} \quad \xrightarrow{\text{re-express}} \quad A\underline{x} = \lambda I\underline{x} \quad \xrightarrow{\text{rearrange}} \quad A\underline{x} - \lambda I\underline{x} = 0$$

which we can then factorise to

$$(A - \lambda I)\underline{x} = 0$$

We are looking for values of $\underline{x}$ and λ for which the above equation is true. However, there is a trivial solution when $\underline{x}$ itself is just the zero vector (*e.g.* (0, 0) in $\mathbb{R}^2$), but this solution is not very interesting (and also doesn't count as an eigenvector by definition). So, if we're not allowing $\underline{x}$ to be zero, the solutions must occur when the action of $A - \lambda I$ on $\underline{x}$ results in a zero vector.

Remembering that a matrix can always be thought of as a transformation, and the determinant of that matrix is the scaling factor applied to the size of the transformed space. So, if a matrix

has a determinant of zero, it means that all the vectors it is applied to will be crushed down to a dimension lower than previously! So, solutions to the above equation exist only when $\det(A - \lambda I) = 0$.

Let's look again at the 2D example we calculated above:

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = \det\left(\begin{bmatrix} 1-\lambda & -1 \\ 2 & 4-\lambda \end{bmatrix}\right)$$

Hence:

$$\begin{aligned} \det(A - \lambda I) &= (1-\lambda)(4-\lambda) + 2 = 0 \\ &= \lambda^2 - 5\lambda + 6 = 0 \end{aligned}$$

In general, $\det(A - \lambda I)$ is a polynomial function of λ, which we refer to as the *characteristic polynomial* of A. Setting this characteristic polynomial equal to zero is referred to as the *characteristic equation*.

To make $\det(A - \lambda I) = 0$, we can set λ to $\lambda_1 = 3$ and $\lambda_2 = 2$. These are all the eigenvalues of A and, in fact, all $n \times n$ matrices have n eigenvalues; however, some might be repeated. It's is also possible that eigenvalues might be complex numbers, so questions (in exams!) often ask you to find all the *real* numbered eigenvalues.

Notice that the sum of the eigenvalues is equal to the trace of the matrix ($\lambda_1 + \lambda_2 = 2 + 3 = 5 = \text{tr}(A) = 1 + 4$) and that the product of the eigenvalues is equal to the determinant of the matrix ($\lambda_1 \times \lambda_2 = 2 \times 3 = 6 = \det(A) = 1 \times 4 - -1 \times 2$). In fact, this is always true.

5.4 Finding All Eigenvectors

5.4.1 For a 2×2 matrix

Continuing with our example, let λ be a value satisfying the characteristic equation, namely, λ is an eigenvalue of A. As will be shown, $\underline{x}$ always constitutes a vector space (*e.g.* $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$), which we denote as EigenSpace(λ), such that the eigenvectors of A corresponding to λ are exactly the non-zero vectors in EigenSpace(λ).

Consider again matrix A. Given that we know that $\lambda_1 = 3$ and $\lambda_2 = 2$, we can now find both eigenvectors. Firstly, for $\lambda_1 = 3$:

$$(A - \lambda_1 I)\underline{x} = (A - 3I)\underline{x} = \left(\begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}\right)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Hence, any $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ satisfying $-2x_1 - x_2 = 0$ is a solution to the above system. The set of such vectors can be represented in a parametric form: $x_1 = t$ and $x_2 =$-2t for any $t \in \mathbb{R}$. Note that

this is a vector space - which we denote as EigenSpace(λ_1) - of dimension 1. Every non-zero vector in EigenSpace(λ_1) is an eigenvector corresponding to λ_1.

Similarly for $\lambda_2 = 2$:

$$(A - \lambda_2 I)\underline{x} = (A - 2I)\underline{x} = \left(\begin{bmatrix}1 & -1\\ 2 & 4\end{bmatrix} - \begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix}\right)\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \begin{bmatrix}-1 & -1\\ 2 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

Hence, any $\begin{bmatrix}x_1\\ x_2\end{bmatrix}$ satisfying $-x_1 - x_2 = 0$ is a solution to the above system. The set of such vectors can be represented in a parametric form: $x_1 = t$ and $x_2 =$-t for any $t \in \mathbb{R}$. This is a vector space, EigenSpace(λ_2) of dimension 1. Every non-zero vector in EigenSpace(λ_2) is an eigenvector corresponding to λ_2.

5.4.2 For a 3×3 matrix

In a similar vein, consider a new matrix A:

$$A = \begin{bmatrix}4 & 6 & 0\\ -3 & -5 & 0\\ -3 & -6 & 1\end{bmatrix}$$

Its characteristic equation is:

$$\det(A - \lambda I) = 0 \qquad \Rightarrow \qquad \begin{vmatrix}4-\lambda & 6 & 0\\ -3 & -5-\lambda & 0\\ -3 & -6 & 1-\lambda\end{vmatrix} = 0$$

Notice that the two zeros in the 3rd column mean that 4 of the 6 terms of the determinant will be zero.

$$(1-\lambda)\begin{vmatrix}4-\lambda & 6\\ -3 & -5-\lambda\end{vmatrix} = 0$$

$$\begin{aligned}(1-\lambda)((4-\lambda)(-5-\lambda)+18) &= 18\\ (\lambda-1)^2(\lambda+2) &= 0\end{aligned}$$

Hence, A has two eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = -2$.

Solving $(A - \lambda I)\underline{x} = 0$ for $\lambda_1 = 1$:

$$(A - \lambda_1 I)\underline{x} = 0$$

$$\begin{bmatrix}4-1 & 6 & 0\\ -3 & -5-1 & 0\\ -3 & -6 & 1-1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix}3 & 6 & 0\\ -3 & -6 & 0\\ -3 & -6 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix}3 & 6 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = 0$$

Hence, any $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ satisfying $3x_1 + 6x_2 = 0$ is a solution to the above system. The set of such vectors can be represented in a parametric form: $x_1 = 2u$, $x_2 = -u$ and $x_3 = v$ for any $(u, v) \in \mathbb{R}$. This is a vector space, EigenSpace(λ_1) of dimension 2. Every non-zero vector in EigenSpace(λ_1) is an eigenvector corresponding to λ_1.

Solving $(A - \lambda I)\underline{x} = 0$ for $\lambda_2 = -2$:

$$\begin{bmatrix} 4--2 & 6 & 0 \\ -3 & -5--2 & 0 \\ -3 & -6 & 1--2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 & 6 & 0 \\ -3 & -3 & 0 \\ -3 & -6 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

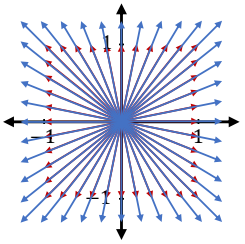
Hence, any $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ satisfying:

$$x_1 + x_2 = 0 \qquad \& \qquad x_2 - x_3 = 0$$

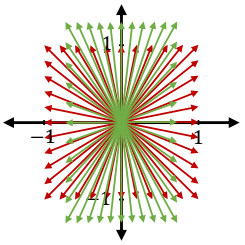
is a solution to the above system. The set of such vectors can be represented in a parametric form: $x_1 = -t$, $x_2 = t$ and $x_3 = t$ for any $t \in \mathbb{R}$. This is a vector space, EigenSpace(λ_2) of dimension 1. Every non-zero vector in EigenSpace(λ_2) is an eigenvector corresponding to λ_2.

5.5 Interpretation of eigensolutions

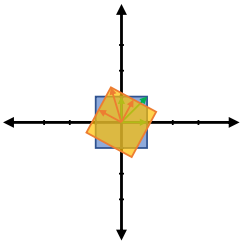
At the beginning of this chapter we observed the effect of a particularly elegant transformation that was a combination of a horizontal shear and vertical stretch, which yielded the easy to spot eigenvectors $(0, 1)$ and $(1, 1)$. There are other systems for which the eigensolutions can be trivially derived. Firstly, for a isotropic (*i.e.*, the same in all directions) scaling, *all vectors* will be eigenvectors (see blue vectors on red adjacent figure); however, if the scaling is anisotropic in magnitude, then only the basis vectors will be eigenvectors (see green vectors on red adjacent figure).



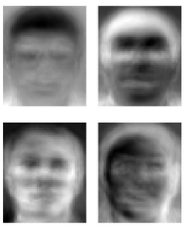
In the case of pure rotational transformations, as you can see from the orange on blue squares in the adjacent figure, all vectors will no longer align with their original span... **except** in the case of a 180° rotation, where *all* vectors will be eigenvectors. If this seems odd, just consider the fact that a 180° rotation will have the same effect as scaling by a factor on -1 in all directions and when you substitute $\theta = 180°$ into the 2D rotation matrix, you can see that all the non-zero terms are on the leading diagonal, just like a scaling.



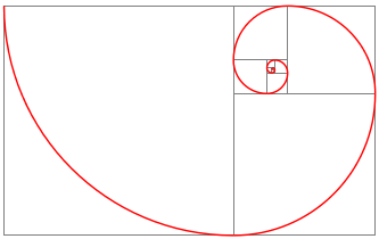
if you imagine a cube in $\mathbb{R}^2$ experiencing pure rotation around one of its axes. Hopefully it should fit with what you've learned so far that this system will only have one real eigenvector, which will be its axis of rotation.



5.5.1 Applications



Before we finish this chapter its worth briefly mentioning three engineering applications of eigenproblems that you can look up. Firstly, Google's famous PageRank algorithm, which uses a matrix to describe the links connecting all pages on the web and then ranks them using an eigenvector. Secondly, data compression can be performed using eigenanalysis to decompose groups of similar data (*e.g.* images of faces, see adjacent for some eigenfaces) into their principal components. Finally, mechanical systems defined springs and masses, it is possible to use eigen analysis to calculate the vibrational modes of the objects.



Chapter 6

Sequences and Series

In mathematics, a **sequence** is a numbered list of terms that may contain repeats and **series** can be thought of as the sum of the terms of an infinite sequence.

Sequence	$a_1, a_2, a_3, a_4, ...$
Series	$a_1 + a_2 + a_3 + a_4 + ...$

6.1 Sequences

Sequences can be defined using formulae, such as in the following three examples where n is a natural number ($\mathbb{N}^+ = \{1, 2, 3, 4, ...\}$) and is called the *index*.

$$\begin{aligned} a_n &= 2n + 3 \\ b_n &= \cos(n\pi) \\ c_n &= (-1)^n \\ d_n &= 7\sqrt[n]{x} \end{aligned}$$

It is common notation to write "(a_n)" to express the sequence $a_1, a_2, a_3, a_4...$ and the two sequence (b_n) and (c_n) are considered to be equal if $b_n = c_n$ for all values of n, which happens to be the case in the examples given above (don't let the different expression of the rules fool you, (b_n) and (c_n) are identical sequences!).

Sequences may also be defined *recursively*, such that a term is defined as a function of previous terms, as in $l_n = 2l_{n-3} - l_{n-1}$.

It may be possible to determine the function defining a sequence based on a few example terms; however, more often, additional information is required, such as what **family** the sequence is from.

For example, for the sequence (h_n)

$$h_1 = 41, \quad h_2 = 43, \quad h_3 = 47 \quad \text{and} \quad h_4 = 53$$

It could be that this sequence was defined by the rule, "the prime numbers in order, starting at 41" and so the next term would be $h_5 = 59$; however, with only the information available to us, the function $h_n = n^2 - n + 41$ could equally be correct, meaning that the next term would be $h_5 = 61$. The human brain is very good at finding patterns, as well as jumping to conclusions!

Arithmetic sequences - This first family of sequences all follow the rule that each term differs from the next by the same fixed amount and are often written

$$a_n = a_1 + d(n-1)$$

where a_1 is the first term and d is called the *common difference*. For example, if we are told that

$$b_1 = 7, \quad b_2 = 12, \quad b_3 = 17, \quad b_4 = 22, \quad ...$$

is part of an arithmetic sequence, then it can be fully describe by the rule

$$b_n = 7 + 5(n-1)$$

Geometric sequences - This family of sequences all follow the rule that each term differs from the next by the same fixed ratio and are often written

$$a_n = a_1 r^{n-1}$$

where a_1 is the first term and r is called the *common ratio*. For example, if we are told that

$$b_1 = \frac{4}{5}, \quad b_2 = \frac{12}{5}, \quad b_3 = \frac{36}{5}, \quad b_4 = \frac{108}{5} \quad ...$$

is part of a geometric sequence, then it can be fully describe by the rule

$$a_n = \frac{4}{5} \times 3^{n-1}$$

Other famous examples of sequences include:

The prime numbers	$2, 3, 5, 7, 11, 13, ...$
The Fibonacci numbers	$0, 1, 1, 2, 3, 5, 8, ...$
The triangle numbers	$1, 3, 6, 10, 15, 21, ...$

6.2 Series

A series is the sum of all the terms of an infinite sequence, so it can be written

$$\sum_{n=1}^{\infty}(a_n) = a_1 + a_2 + a_3 + a_4 + ...$$

However, if only a finite number of terms are summed, this is referred to as a *truncated* series.

The same families that are used to describe sequences also apply to series and for some of these, useful identities can be found that help us understand and manipulate them.

Arithmetic series - The sum of the first m terms of the arithmetic sequence (a_n) constitutes the truncated series S_m, such that

$$S_m = \sum_{n=1}^{m}(a_n) = a_1 + a_2 + a_3 + ... + a_{m-1} + a_m$$

We can derive an expression for S_m by first re-expressing each term of the series using only the first term and the common difference.

$$S_m = a_1 + (a_1 + d) + (a_1 + 2d) + ... + (a_1 + (m-2)d) + (a_1 + (m-1)d)$$

Next we re-express the series again, but this time only using the last term and the common difference.

$$S_m = (a_m - (m-1)d) + (a_m - (m-2)d) + (a_m - (m-3)d) + ... + (a_m - d) + a_m$$

Adding these two forms together yields

$$2S_m = (a_1 + a_m) + (a_1 + a_m) + (a_1 + a_m) + ... + (a_1 + a_m) + (a_1 + a_m)$$

which rearranges to our explicit equation

$$\begin{aligned} S_m =& \frac{m}{2}(a_1 + a_m) \\ S_m =& \frac{m}{2}(a_1 + (a_1 + (m-1)d)) \\ S_m =& \frac{m}{2}(2a_1 + (m-1)d) \end{aligned}$$

Geometric series - The sum of the first m terms of the geometric sequence (a_n) constitutes the truncated series S_m, such that

$$S_m = \sum_{n=1}^{m}(a_n) = a_1 + a_2 + a_3 + ... + a_{m-1} + a_m$$

We can derive an expression for S_m by first re-expressing each term of the series using only the first term and the common ratio.

$$S_m = a_1 + (a_1 r) + (a_1 r^2) + ... + (a_1 r^{m-2}) + (a_1 r^{m-1})$$

Multiplying both sides of the expression above by r

$$S_m r = a_1 r + (a_1 r^2) + (a_1 r^3) + ... + (a_1 r^{m-1}) + (a_1 r^m)$$

The difference between the two equations above is

$$S_m - S_m r = a_1 - a_1 r^m$$

which factorises and rearranges to our explicit equation

$$\begin{aligned} S_m(1-r) =& a_1(1 - r^m) \\ S_m =& a_1 \frac{1 - r^m}{1 - r} \end{aligned}$$

Notice that based on the expression above, even if there are infinitely many terms in the sequence, the series may still be finite if $r < 1$. We will discuss this concept further in the following section. Besides these two families, there are many others which also have handy tricks to help you make use of them.

6.3 Limits and Convergence

A limit is the value that a function or sequence "approaches" as the input or index approaches a specified value. The expression

$$\lim_{x \to n} f(x) = L$$

is read "the limit of f of x, as x goes to n, equals L".

It is often useful to know whether or not a series *converges*, and if so, what it converges to. A variety of convergence tests exist, which allow you to methodically examine the convergence of a series.

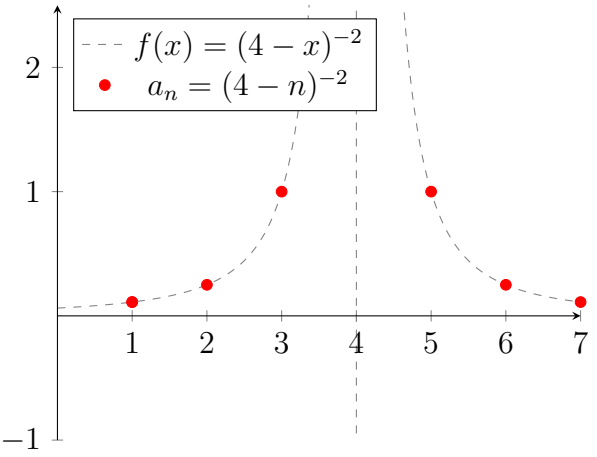
The n$^{\text{th}}$ term test - calculate what the last term of a sequence would be, as it can give us a clue as to whether the corresponding series converges. In fact, we can say that it is "necessary, but not sufficient" (*i.e.*, needed, but not enough on its own), that as the index n goes to infinity, the terms must go to zero in order for a series to converge.

Formally,

$$\text{If} \quad \sum_{n=0}^{\infty} a_n \quad \text{converges, then} \quad \lim_{n \to \infty} a_n = 0$$

and by similar reasoning, if the limit does not go to zero, as the index goes to infinity, the series must diverge. (Although, in the case of the adjacent figure, $f(x)$ does go to infinity, but it still diverges!)

$$\text{If} \quad \lim_{n \to \infty} a_n \neq 0 \quad \text{then} \quad \sum_{n=0}^{\infty} a_n \quad \text{diverges}$$

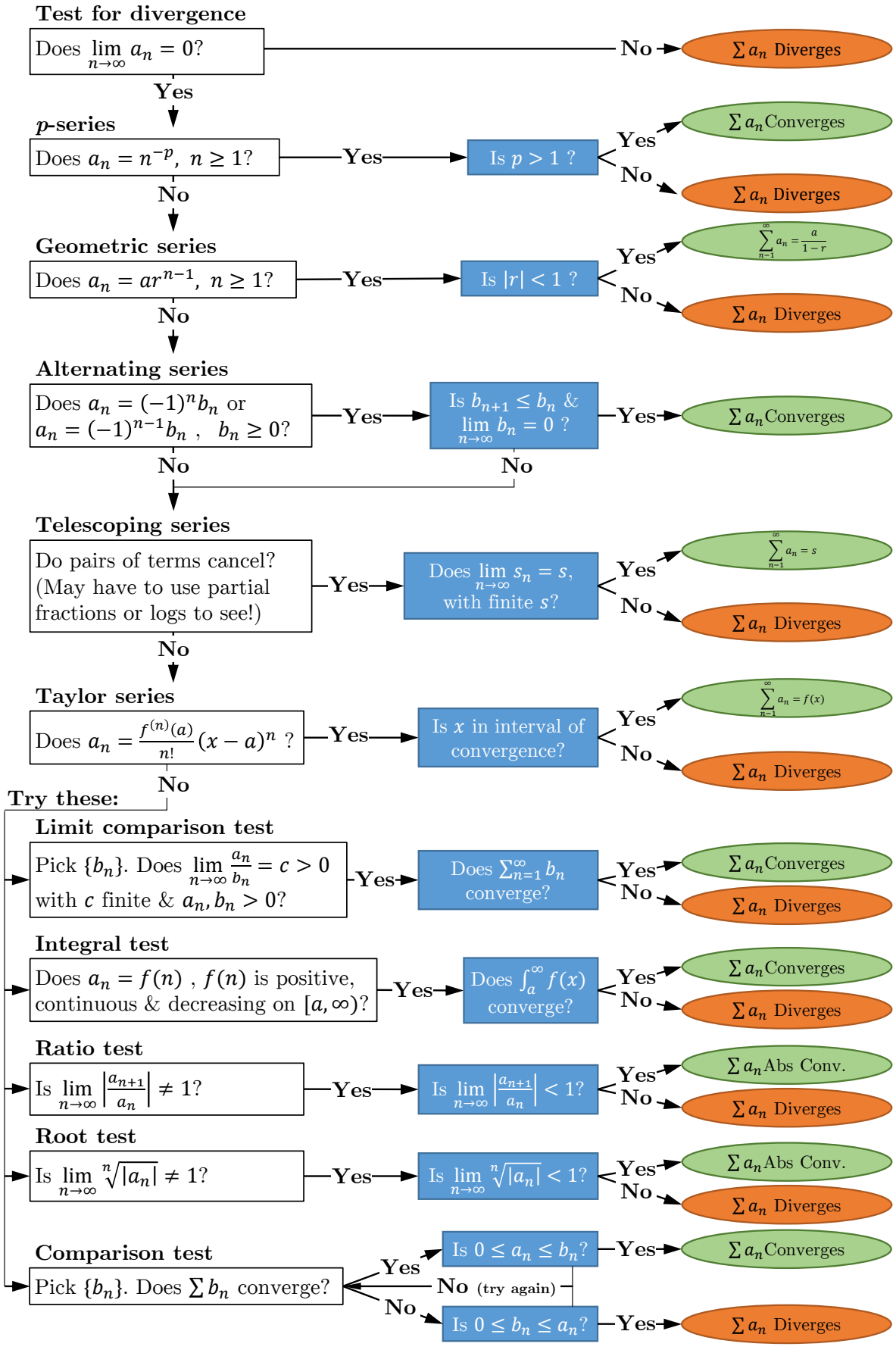


Plot of the terms in a divergent series where the limit at infinity is zero.

To understand why having a limit at infinity of zero is not enough to establish series convergence, consider the function $a_n = (4-n)^{-2}$. The figure shows the terms in this series; however, the point at $n = 4$ is missing as there is a discontinuity in the function. So, although a sequence may eventually "settle down" to zero, if it contains terms that are infinite or undefined along the way, then the corresponding series will not converge.

p-series test - Consider a sequence of the form $a_n = 1/n^p$ and remember that numbers less than 1 get smaller when they are put to powers greater than 1 (*e.g.* 0.5^2=0.25). So the test for convergence here is simply to observe whether the exponent, p, is greater than 1.

There are many other tests that can be applied to determine the convergence of a series. There is also a "sensible" order in which they should be tried that is based in part on how conclusive their results are, but also on how easy it is to apply them, which will hopefully make this process efficient. There is no need for you to learn all of these as they are easy to look up online; however, what is important is that you understand what they mean and how to use them.


Test for divergence
Does $\lim_{n\to\infty} a_n = 0$?
No
$\sum a_n$ Diverges
Yes
p-series
Does $a_n = n^{-p},\ n \geq 1$?
Yes
Is $p > 1$?
Yes
$\sum a_n$ Converges
No
$\sum a_n$ Diverges
No
Geometric series
Does $a_n = ar^{n-1},\ n \geq 1$?
Yes
Is $|r| < 1$?
Yes
$\sum_{n=1}^{\infty} a_n = \frac{a}{1-r}$
No
$\sum a_n$ Diverges
No
Alternating series
Does $a_n = (-1)^n b_n$ or $a_n = (-1)^{n-1} b_n$, $b_n \geq 0$?
Yes
Is $b_{n+1} \leq b_n$ & $\lim_{n\to\infty} b_n = 0$?
Yes
$\sum a_n$ Converges
No
No
Telescoping series
Do pairs of terms cancel? (May have to use partial fractions or logs to see!)
Yes
Does $\lim_{n\to\infty} s_n = s$, with finite s?
Yes
$\sum_{n=1}^{\infty} a_n = s$
No
$\sum a_n$ Diverges
No
Taylor series
Does $a_n = \frac{f^{(n)}(a)}{n!}(x-a)^n$?
Yes
Is x in interval of convergence?
Yes
$\sum_{n=1}^{\infty} a_n = f(x)$
No
$\sum a_n$ Diverges
No
Try these:
Limit comparison test
Pick $\{b_n\}$. Does $\lim_{n\to\infty} \frac{a_n}{b_n} = c > 0$ with c finite & $a_n, b_n > 0$?
Yes
Does $\sum_{n=1}^{\infty} b_n$ converge?
Yes
$\sum a_n$ Converges
No
$\sum a_n$ Diverges
Integral test
Does $a_n = f(n)$, $f(n)$ is positive, continuous & decreasing on $[a, \infty)$?
Yes
Does $\int_a^{\infty} f(x)$ converge?
Yes
$\sum a_n$ Converges
No
$\sum a_n$ Diverges
Ratio test
Is $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| \neq 1$?
Yes
Is $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$?
Yes
$\sum a_n$ Abs Conv.
No
$\sum a_n$ Diverges
Root test
Is $\lim_{n\to\infty} \sqrt[n]{|a_n|} \neq 1$?
Yes
Is $\lim_{n\to\infty} \sqrt[n]{|a_n|} < 1$?
Yes
$\sum a_n$ Abs Conv.
No
$\sum a_n$ Diverges
Comparison test
Pick $\{b_n\}$. Does $\sum b_n$ converge?
Yes
Is $0 \leq a_n \leq b_n$?
Yes
$\sum a_n$ Converges
No (try again)
No
Is $0 \leq b_n \leq a_n$?
Yes
$\sum a_n$ Diverges

6.4 Truncated sum of 1, n, n^2 and n^3

If you were asked to sum an infinite series of ones (*i.e.*, $S = 1 + 1 + 1 + 1 + 1 + ...$), this would clearly never stop growing and would therefore explode to infinity (a very slow and unexciting explosion). However, a truncated series of N ones would of course be finite and add up to N.

$$\sum_{n=1}^{n=N} 1 = N$$

You could even think of this series an arithmetic series where the common difference is zero (so $S_N = \frac{N}{2}(2 \times 1 + (N-1)0) = N$); however, the geometric series type analysis with a common ratio of 1 would not work.

Now imagine a series which simply added up the natural numbers (*i.e.*, $S = 1 + 2 + 3 + ...$). Once again, this series clearly doesn't stop growing, but any truncation would result in a finite answer. Applying the arithmetic series analysis again here (with $d = 1$) leads to the result

$$\sum_{n=1}^{n=N} n = \frac{N}{2}(2 \times 1 + (N-1) \times 1) = \frac{N(N+1)}{2}$$

An alternative approach to evaluating this series is by expanding the binomial expression $(n-1)^2$....

$$(n-1)^2 = n^2 - 2n + 1$$

and then rearranging the result to

$$n^2 - (n-1)^2 = 2n - 1$$

At this point, it's necessary to know that summation is a linear operation, so it has the following two useful properties:

$$\sum (a_n + b_n) = \sum (a_n) + \sum (b_n) \qquad \& \qquad \sum (3a_n) = 3 \sum (a_n)$$

which, combined with our rearranged binomial above, means that we can say

$$\sum_{n=1}^{n=N} (n^2 - (n-1)^2) = \sum_{n=1}^{n=N} (2n-1) = 2 \sum_{n=1}^{n=N} (n) - \sum_{n=1}^{n=N} (1)$$

This may initially not look very helpful, but the left hand side expression is what we refer to as a telescoping series. This means that pairs of terms within the series will cancel with each other. In this case, each term contains a n^2 component and a $(n-1)^2$ component, which is simply the value of n^2 from the previous term in the sequence. Remembering that the series starts from $n = 1$ and that $0^2 = 0$, it becomes clear that only the final n^2 term is not cancelled, so $\sum_{n=1}^{n=N} (n^2 - (n-1)^2) = N^2$. So our binomial expression becomes

$$2 \sum_{n=1}^{n=N} (n) - \sum_{n=1}^{n=N} (1) = N^2$$

Substituting in our knowledge that the sum of N ones is just N, then rearranging, leads to the same result we found earlier

$$\sum_{n=1}^{n=N} n = \frac{N^2 + N}{2}$$

This on its own is not hugely exciting, but it turns out we can use this same approach to find an expression to evaluate any truncated series of the form $\sum n^b$, where b is a positive integer.

6.4.1 Truncated sum of n^b

So, to find an expression for $\sum n^2$, we start by expanding the cubic expression $(n-1)^3 = (n^3 - 3n^2 + 3n - 1)$ and then rearranging this equation to

$$n^3 - (n-1)^3 = 3n^2 - 3n + 1 \qquad \text{So} \qquad \sum_{n=1}^{n=N} (n^3 - (n-1)^3) = 3\sum_{n=1}^{n=N} (n^2) - 3\sum_{n=1}^{n=N} (n) + \sum_{n=1}^{n=N} (1)$$

Once again, we have have a telescoping series, so clearly $\sum_{n=1}^{n=N} (n^3 - (n-1)^3) = N^3$. Combining this with our knowledge of $\sum n$ and $\sum 1$, we can rearrange the expression to make it explicit for the only thing we don't know.

$$\sum_{n=1}^{n=N} n^2 = \frac{1}{3}\left(N^3 + \frac{3}{2}(N^2 + N) - N\right) = \frac{N(N+1)(2N+1)}{6}$$

This continues to work for all positive integer values of b, although the algebra does become tiresome by hand. So, lastly, to find the truncated sum of n^3, you would start from $(n-1)^4$ and follow the same process, leading to the result

$$\sum_{n=1}^{n=N} n^3 = \frac{N^2(N+1)^2}{4}$$

6.4.2 Example of a truncated sum

To evaluate the expression

$$\sum_{n=1}^{n=7} (2n+1)(n-1)(n+3)$$

one strategy is to first expand this expression to the from $(2n^3 + 5n^2 - 4n - 3)$ and then chop it up into it's various components using the linearity properties

$$\sum_{n=1}^{n=7} (2n+1)(n-1)(n+3) = 2\sum_{n=1}^{n=7} (n^3) + 5\sum_{n=1}^{n=7} (n^2) - 4\sum_{n=1}^{n=7} (n) - 3\sum_{n=1}^{n=7} (1)$$

As we have already derived expressions for each of these truncated summations, solving becomes a matter of simple substitution

$$\sum_{n=1}^{n=7} (2n+1)(n-1)(n+3) = 2\left(\frac{7^2(7+1)^2}{4}\right) + 5\left(\frac{7(7+1)(2\times 7+1)}{6}\right) - 4\left(\frac{7^2+7}{2}\right) - 3(7) = 2135$$

6.5 Mind blown

There is a meaningful sense in which the following statement is true and it can be shown in a variety of ways.

$$S_\infty = 1 + 2 + 3 + 4 + 5 + ... = -\frac{1}{12}$$

Seriously though... head to the Numberphile YouTube channel for an explanation. The general idea is that "infinity" isn't really anything like just a "very large number".

Chapter 7

Power Series

It is often useful to re-express a function as a power series. Before we go any further, its important that you understand what this means, so consider the example below:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + ...$$

Try this on your calculator at $x = 1.2$ so that you are convinced that the function on the LHS (left-hand-side) can be approximated with the power series on the RHS.

7.1 Maclaurin Series

For many smooth, continuously differentiable functions (or what are sometimes called "well-behaved" functions), if you know everything about the function at just one place (at $x = 0$ for Maclaurin series), it is possible to use this information to reconstruct the entire curve.

When we say "know everything" at a point, we mean the value of the function and all of its derivatives $(f(0), f'(0), f''(0), f^{(3)}(0), ...)$.

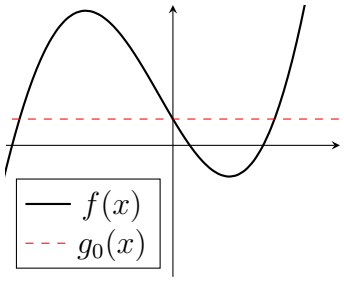
We can use this information to construct a sequence of approximations to our function.

Let's suppose we have some function $f(x)$ and we know everything about it at the point $x = 0$, but nothing about it anywhere else except that it is well-behaved.

We can use the value of the function at $x = 0$ to construct a (pretty bad) approximate function $g_0(x)$, which we shall call our "zeroth order" guess. Clearly, as this function only has one piece of information to work with, it will just be a horizontal line that goes through the y-axis at the same place as the real function, $f(0)$.

$$g_0(x) = f(0)$$

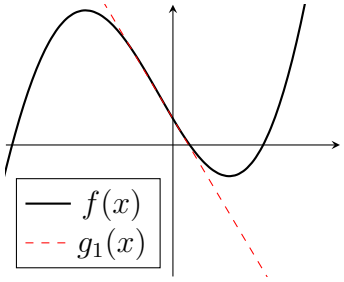
We can see, in the first figure on the right, that this guess function is not great...



0^{th} order approximation

A better approximation can be made by using the next piece of information available to us, which is the value of the function's first derivative $f'(0)$.

Our "first order" guess, $g_1(x)$ uses both pieces of information to construct a straight line (of the form $y = mx + c$) where both the y-intercept and the gradient are the same as the function $f(x)$ at $x = 0$. This is shown in the second figure and is clearly a significant improvement.



1^{st} order approximation

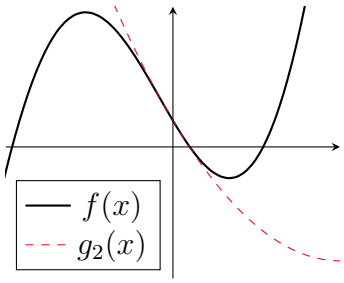
$$g_1(x) = f(0) + f'(0)x$$

Repeating this process, we will now also use the second derivative of our function to help improve our guess function (third figure).

We see that a factor of $1/2$ is required before the x^2 term. To understand why, try differentiating $g_2(x)$ twice to convince yourself that it equals $f''(0)$ as it should.

$$g_2(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2$$

Finding the third order approximation requires using the third derivative, as shown in the last figure. Again, to understand where the $(3 \times 2)^{-1}$ term comes from, try differentiating $g_3(x)$ three times.

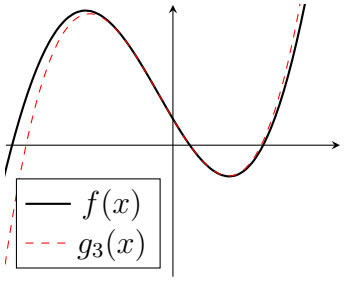


2^{nd} order approximation

Hopefully, you will now see the pattern and be confident that the next term (for the fourth order approximation) will have a factor of $(4 \times 3 \times 2)^{-1}$ before it.

$$g_3(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f^{(3)}(0)}{3 \times 2}x^3$$

With this pattern in mind, we can now write the general equation for the n^{th} order term (Don't forget that $n!$ is defined as $n \times (n-1)!$, so $1! = 1 \times 0!$ and therefore $0! = 1$).



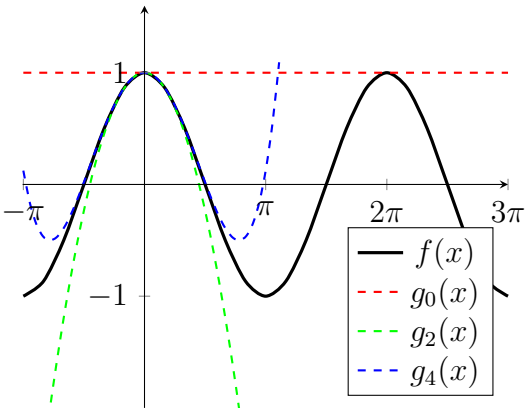
3^{rd} order approximation

$$\begin{aligned} g_n(x) &= \frac{f(0)}{0!} + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \ldots + \frac{f^{(n)}(0)}{n!}x^n \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n \end{aligned}$$

7.1.1 Maclaurin Examples

Example - Consider the function $f(x) = \cos(x)$ (a well-behaved function). It's first four derivatives evaluated at $x = 0$ are:

$$\begin{aligned} f(0) &= \cos(0) = 1 \\ f'(0) &= -\sin(0) = 0 \\ f''(0) &= -\cos(0) = -1 \\ f^{(3)}(0) &= \sin(0) = 0 \\ f^{(4)}(0) &= \cos(0) = 1 \end{aligned}$$



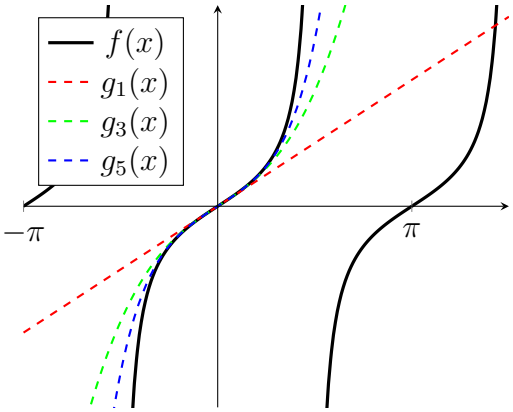
Maclaurin expansion of $\cos(x)$.

Therefore the Maclaurin series expansion is

$$\begin{aligned} g(x) &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \ldots \\ &= \sum_{n=0}^{\infty}(-1)^n \frac{1}{(2n)!}x^{2n} \end{aligned}$$

Example - Consider the function $f(x) = \tan(x)$ (not a well-behaved function, due to its asymptotes). It's first four derivatives evaluated at $x = 0$ are:

$$\begin{aligned} f(0) &= \tan(0) = 0 \\ f'(0) &= \sec^2(0) = 1 \\ f''(0) &= 2\tan(0)\sec^2(0) = 0 \\ f^{(3)}(0) &= -2(\cos(2(0)) - 2)\sec^4(0) = 2 \\ f^{(4)}(0) &= -4(\cos(2(0)) - 5)\tan(0)\sec^4(0) = 0 \end{aligned}$$



Maclaurin expansion of $\tan(x)$.

$$g(x) = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \ldots$$

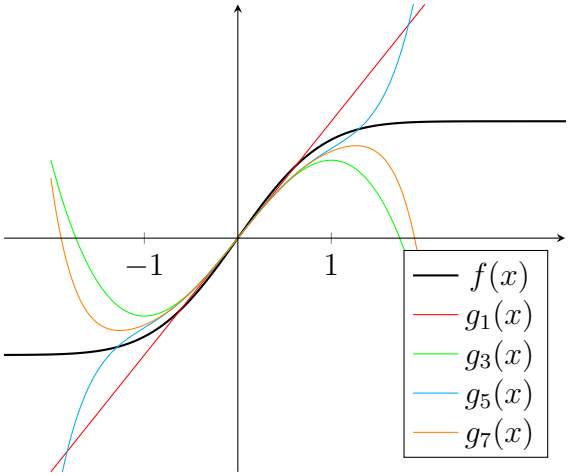
Looking at fig. 7.6 we can see that although our approximation is definitely improving locally (*i.e.*, close to $x = 0$), it is not able to model the function at all after $x = \frac{\pi}{2}$. This is because $\tan(x)$ contains asymptotes and is therefore not considered a well-behaved function.

In the next example, we look again at the error function (from the Normal Distribution Chapter). The adjacent figure shows the error function as well as Maclaurin approximations up to the 7^{th} order, but although the function is quickly represented well near $x = 0$, the improvements seem stop beyond around $x = 1$. However, the Taylor polynomials do continue to improve (slowly) and the error function can be exactly expressed as a Taylor series.

$$
\begin{aligned}
f(0) &= \operatorname{erf}(0) = 0 \\
f'(0) &= \frac{2}{\sqrt{\pi}} e^{-x^2} = \frac{2}{\sqrt{\pi}} \\
f''(0) &= -\frac{4}{\sqrt{\pi}} x e^{-x^2} = 0 \\
f^{(3)}(0) &= -\frac{4}{\sqrt{\pi}} (2x^2 - 1) e^{-x^2} = -\frac{4}{\sqrt{\pi}} \\
f^{(4)}(0) &= -\frac{8}{\sqrt{\pi}} x (2x^2 - 3) e^{-x^2} = 0
\end{aligned}
$$

which yields (once we've added a few more terms),

$$
\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216} - \ldots \right)
$$

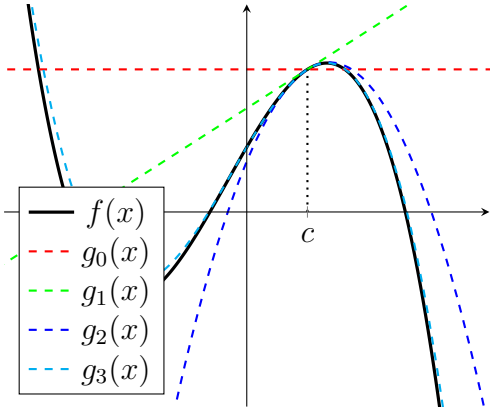


Plot of the error function as well as the Maclaurin series expansions up to 7^{th} order. *N.B.* The "true" curve (black) may itself also have been calculated with a (high order) power series.

7.2 Taylor Series

The Taylor series simply extends the Maclaurin series concept by saying that we can reconstruct well-behaved functions if we know everything about *any* point (*i.e.*, not just at the point $x = 0$ like Maclaurin). The expression can be derived in the same way as the Maclaurin series, but requires some minor rearrangement to find each successive approximation.

$$
g_n(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!} (x - c)^n
$$



Taylor series expansions of an arbitrary function around the point c

Example - Consider the function $f(x) = e^x$. It's first three derivatives evaluated at $x = 1$ are:

.

$$f(1) = e^1 = e$$
$$f'(1) = e^1 = e$$
$$f''(1) = e^1 = e$$
$$f^{(3)}(1) = e^1 = e$$

Notice how this is different from the expression given at the beginning of this chapter for e^x. This is because the first series was expanded around the point $x = 0$ (Maclaurin series), whereas here we have expanded around $x = 1$.

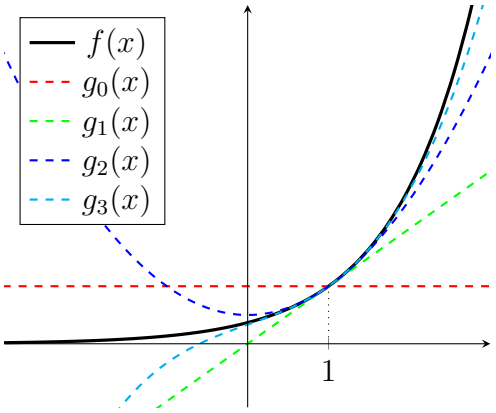
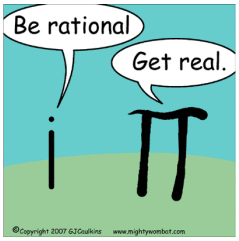


Figure 7.9: Taylor series expansions of the function $f(x) = e^x$ around the point $x = 1$.

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

It is important to select a sensible point from which to expand a series. This is especially true if you are trying to use only a few terms of the series (truncation) to approximate a complicated function, as the further away from your expansion point, the less accurate (on average) your approximations become.

Mind slightly blown - Of course, another way to think about the two different power series approximations of the function $f(x) = e^x$, is that seemingly moving our approximation from $x = 0$ to $x = 1$ caused us to multiply every term by e. But when you consider that $e^{x+1} = ee^x$ it should suddenly make a lot of sense!



Chapter 8

Complex Numbers

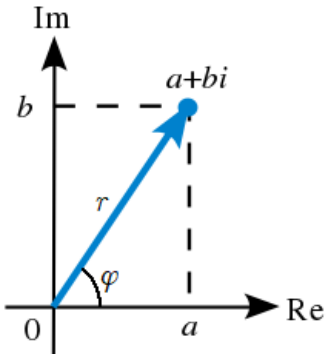
You will most likely have met complex numbers before, but if you haven't **don't panic** because this chapter starts from scratch and aims to make sure that you really *get* them as they're going to turn out to be useful later on!

Essentially, complex numbers are those that can be expressed in the form $a + bi$, where i is the "imaginary unit". A typical definition of i might be "a solution to the equation $x^2 + 1 = 0$". Notice I said *a* solution and not *the* solution, because of course this equation has two solutions ($x = \sqrt{-1} = i$ and $x = -\sqrt{-1} = -i$)[†]. The definition can also be simply stated directly as:

$$i = \sqrt{-1}$$

C.F. Gauss 1777-1855

As we shall see, imaginary numbers turn up a lot in real applications, so their name is perhaps a bit misleading and some of the early pioneers of this field weren't happy about this. In fact, the famous mathematician Carl Friedrich Gauss wished that they had been called "Lateral numbers", which should make a lot of sense to you once you've seen them plotted on a 2D plane. These plots are called **Argand** diagrams and the complex number $a + ib$ corresponds to a point at the Cartesian coordinate (a, b).



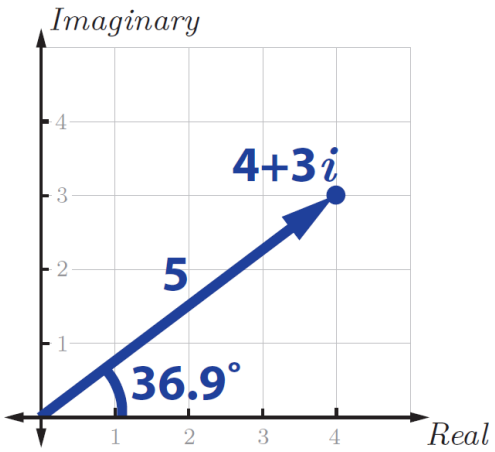
$$\mathrm{Re}(a+bi) = a \qquad \& \qquad \mathrm{Im}(a+bi) = b$$

$$|a+bi| = \sqrt{a^2+b^2} = r \qquad \& \qquad \arg(a+bi) = \arctan\left(\frac{b}{a}\right) = \varphi$$

Equally, we can switch to Polar coordinates, where we now need to know the angle, φ, that the line makes with the positive x-axis (called the *argument* or *phase*), as well as the distance to the point, r, which can be thought of as a radius (referred to as the *magnitude*, *absolute* or *modulus*).

[†]In fact, it is the "Fundamental Theorem of Algebra" which states that a polynomial of order n, must have exactly n roots. Sometimes these are both real (*i.e.*, $x^2 = 1$), sometimes complex (*i.e.*, $x^2 = -1$), sometimes both in the same place (*i.e.*, $x^2 = 0$), and sometimes there's a mix of real and complex. In particular, for polynomials with real coefficients, the roots are always either real or in conjugate pairs.

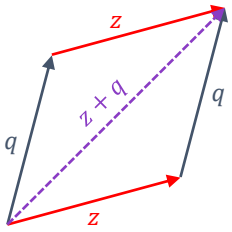
To convert between these two representation, you just need to remember your trigonometry and Pythagoras theorem. The polar description can also be written as, for example, $4 + 3i = 5\angle 36.9°$.



$$z = a + bi = r\cos(\varphi) + r\sin(\varphi)i$$
$$z = re^{i\varphi} = \sqrt{a^2 + b^2}\; e^{i\arctan\left(\frac{b}{a}\right)}$$

where φ is in radians rather than the the degrees shown in the figure. Hopefully its clear from the diagram where the sine and cosine notation came from, but you may be wondering how e got involved. However, you'll have to skip ahead in the notes to see that $e^{ix} = \cos(x) + i\sin(x)$.

8.1 Operations with complex numbers

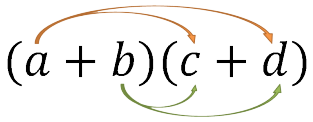


Adding or subtracting complex numbers is simple enough, as you just consider the two parts separately (like with vectors!). For example, consider the two complex numbers $z = a + bi$ and $q = c + di$,

$$z + q = (a + c) + (b + d)i \qquad \& \qquad z - q = (a - c) + (b - d)i$$

Similarly, if I wanted to multiply two complex numbers together, I could just work through the FOIL approach (first, outside, inside, last) and get the answer. For example, consider the two complex numbers $z = a + bi$ and $q = c + di$. However, don't forget to fully simplify at the end, such that any i^2 terms are converted to -1.

$$z \times q = (a + bi)(c + di) = ac + adi + bci + bdi^2$$
$$= ac + adi + bci - bd = (ac - bd) + (ad + bc)i$$



There are several approaches for division, all of which are very tedious using the $(a + bi)$ form. In the method shown below, we first write the expression as a fraction and then perform an operation called "realising the denominator" in which we multiply the top and bottom lines by the **complex conjugate** of the denominator (much like the process of "rationalising the denominator" when surds are involved).

$$q \div z = \frac{(c + di)}{(a + bi)} = \frac{(c + di)}{(a + bi)} \times \frac{(a - bi)}{(a - bi)} = \frac{ac - bci + adi - bdi^2}{a^2 + abi - abi - bi^2} = \frac{(ac + bd) + (ad - bc)i}{a^2 + b^2}$$

The complex conjugate is a very useful concept and it is defined, for $z = a + bi$, as the complex number which has the same real component, a, and an imaginary component of the same size, but opposite sign, $-b$. This, when multiplied with z, has the property of making it into a purely real number and is given the symbol $\bar{z} = a - bi$ (sometime z^* is used instead of $\bar{z}$).

This process gets tedious pretty quickly - imagine if you were asked in a test for $z \times z \times z \div q$...! However, if I represent my complex number in polar form, things get much easier. To multiply z and q, I just multiply their magnitudes and then sum their arguments. The logic behind this becomes clear when the numbers are written in their exponential form $z_1 = r_1 e^{i\varphi_1}$ and $z_2 = r_2 e^{i\varphi_2}$, so therefore $z_1 z_2 = r_1 r_2 e^{i(\varphi_1+\varphi_2)}$. Division is similar.

$$z \times q = (r_z r_q) e^{i(\varphi_z+\varphi_q)}$$
$$z \div q = (r_z / r_q) e^{i(\varphi_z-\varphi_q)}$$

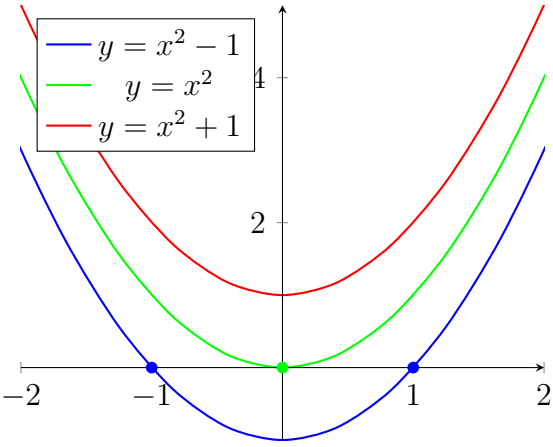
Example - If $z = (3-4i)$ and $q = (12+5i)$, then find z^3/q^2 ...

$$\begin{aligned} z^3/q^2 &= (5^3 \angle(-53.13 \times 3)^\circ)/(13^2 \angle(22.6 \times 2)^\circ) \\ &\xrightarrow{\text{simplify}} = (125\angle -159.4^\circ)/(169\angle 45.2^\circ) = (125/169)\angle(-110.7-45.2)^\circ = 0.74\angle -204.6^\circ \end{aligned}$$

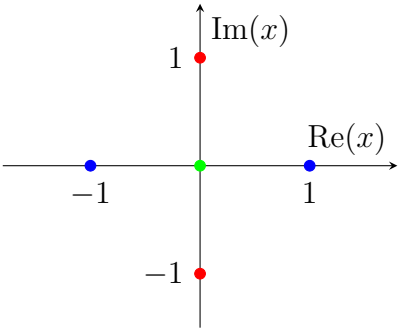
Therefore $z^3/q^2 = -0.6724 + 0.3083i$(expressed in same form as question)

8.2 Finding complex roots

One particularly interesting concept is the process for finding complex roots. Remember that the roots of a function are the points at which $f(x) = 0$, which can usually be thought of as the points at which the function touches the horizontal axis.



In the adjacent figure, you see three functions, for the **blue** function (lowest) you can immediately see its two roots, where as for the **green** (middle) there appears to be only one (although you can think of this as it having one "unique root" as it just has two at the same location). However, the **red** function (top) doesn't appear to have any roots... but we know from the Fundamental Theorem of Algebra that it must have two, so where are they?



Well, as we discussed above, in these fairly simple cases, we can just rearrange our equation to show that the other roots are complex $x = \pm i$. So, this is something that you're going to have to be careful with from a language perspective, because all the functions in the above figure have *two roots* (see figure below), but only one has *two unique, real roots*.

Example - We can apply this same logic to more complicated quadratic expressions. Consider the function $f(x) = x^2 - 2x + 3$. To find the roots of a quadratic, we either factorise (which doesn't work in this case), or put the coefficients into the quadratic formula.

$$x = \frac{2 \pm \sqrt{(-2)^2 - 4 \times 1 \times 3}}{2 \times 1} = \frac{2 \pm \sqrt{4 - 12}}{2} = 1 \pm i\sqrt{2}$$

8.3 De Moivre's Theorem

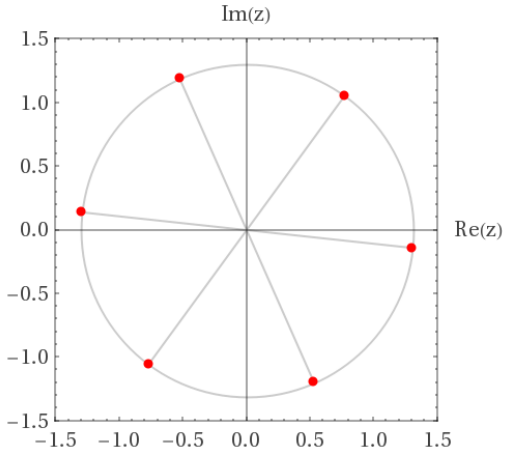
This concept results directly from the exponential polar form of a complex number and primarily gives you a short-cut to finding the roots of a complex number. (NB. e^{ix}=cis$(x) = \cos(x) + i\sin(x)$).

$$z^n = (re^{i\varphi})^n = r^n e^{in\varphi} = r^n(\cos(n\varphi) + i\sin(n\varphi)) = r^n \text{cis}(n\varphi)$$

Example - I want to find z in the expression $z^6 = (4 - 3i)$. First, I would find the magnitude of the right hand side (RHS), which in this case is $\sqrt{4^2 + (-3)^2} = 5$, then, based on the equation above, i can say that the magnitude of our solution must be $r = \sqrt[6]{5}$. Next, find the argument of the RHS using simple trigonometry, $\theta = \arctan(-3/4) = -36.87°$. Now, by comparing this to De Moivre's expression above, we can see that, $\theta = n\varphi = -36.87°$, $n = 6$, so $\varphi = -6.14°$.

We have now found r and φ for one of the roots, but this is actually enough to make finding the other five roots easy (it's a sixth power, so we're expecting six in total) . As we've seen already, when you multiply complex numbers, the magnitude multiplies and the angles sums. So, if we're looking for numbers that each multiply by themselves 6 times to make the same number, then they must all have the same magnitude, $r = \sqrt[6]{5}$.

So, really, we are just looking for different angles which, when multiplied by 6, equal $-36.87°$. We've already found one ($\varphi = -6.14°$), but what about the rest...? You might be thinking, 'but how can there be others?'. At this point we have to remember that in polar coordinates, adding 360° to an angle takes it back to the same place and essentially has no effect. So, we can now say that we are looking for numbers that satisfy the following two criteria $-180° < \varphi \leq 180$ and $6\varphi = (m360 - 36.87)°$, where m is an integer. Rearranging this slightly, gives us $\varphi = (m60 - 6.14)°$, which suggests that, starting from the angle that we already have, the other values of φ are found by simply adding or subtracting 60°, which is another way of saying that these are a set of equally distributed points around the circumference of a circle, radius $r = \sqrt[6]{5}$, as per the adjacent diagram.



So, the roots are $z = \sqrt[6]{5}\angle(-126.14°, -66.14°, -6.14°, 53.86°, 113.86°173.86°)$, which can more succinctly be expressed as $z = \sqrt[6]{5}\angle(m60 - 6.14)°$. Finally, because of the way the question was written, we must now use trig to covert all of these polar representations back to Cartesian form... which you can read off from the figure.

8.3.1 Efficient integration

Currently, if you were asked to evaluate the expression $I = \int e^{ax}\cos(bx)\mathrm{d}x$, you would probably not be very happy with whoever asked and expect a 20 minute session of two stage "integration by parts"... Fortunately, complex number can help here as well!

Recall from earlier in this chapter that $e^{(a+ib)x} = e^{ax}e^{ibx} = e^{ax}(\cos(bx) + i\sin(bx))$, which has a striking similarity to what we've been asked to integrate, except it's now got this addition $i\sin(bx)$ part in it.

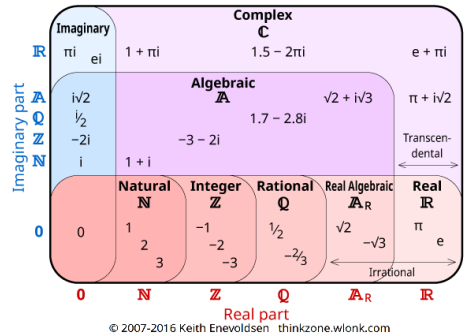
So, the idea is, to take your expression and compress it down into the purely exponential form, then integrate in a single easy step and finally evaluate only the real component of this expression (or only the imaginary part if your question contained a $\sin(bx)$ instead of $\cos(bx)$).

$$\int e^{ax}\cos(bx)dx = \mathrm{Re}\left\{\int e^{(a+ib)x}dx\right\} \qquad \& \qquad \int e^{ax}\sin(bx)dx = \mathrm{Im}\left\{\int e^{(a+ib)x}dx\right\}$$

The way I like to think about this approach is that it takes your initial problem, adds some other bits to make the integration more convenient, but *crucially* it tags this extra stuff with a special sticky label (*i.e.*, the imaginary unit i), which makes it easy to find and ignore later on. Such a simple approach can save you a lot of time and will also make you much less prone to the mistakes that inevitably come from many pages of working.

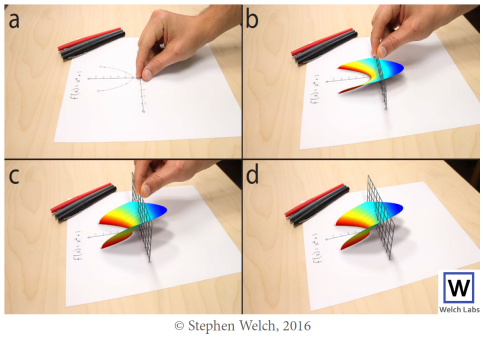
8.4 Imaginary numbers really exist

This is the end of this very brief introduction to imaginary numbers for this course, but we'll be using them later on in other topics. The space of all numbers can be divided up into subsets and the figure to the right illustrates that complex numbers contain all the others that you will have encountered in the past.

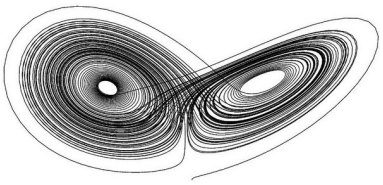


This means that you are now equipped to address a very wide range of maths problems, especially those we'll be seeing in the next 9 weeks!

It you would like to find out more about complex numbers, the phenomenal video series by Welch labs will change your life: http://www.welchlabs.com/



© Stephen Welch, 2016



Chapter 9

Ordinary Differential Equations

You've already met the concept of differentiation in the context of functions and have almost certainly even applied it in the past to distance/speed/acceleration type questions. So, now let's see if we can apply this concept more generally to the class of problems called Ordinary Differential Equations (ODEs). These equations are referred to as "ordinary" because the functions being differentiated are dependant on a single variable. For example, in mechanics so far, both speed and acceleration are a function of time only (*i.e.*, $\frac{dx}{dt}$ & $\frac{d^2x}{dt^2}$), whereas you can imagine, for example, a problem involving the diffusion of heat, in which the local temperature is a function of both time and space (*i.e.*, $T(t,x)$), meaning that you could in principal differentiate T with respect to t or x... or both! We will come back to these in a later chapter.

First off, let's start with some notation. As with so many areas of maths, there are many ways to represent the same concept, so it's important that you don't get stuck always using the same style in class or you might be intimidated when you see something new in practice. However, the general rule is, use whatever style you prefer to use, just *be consistent* throughout any given document, also, if a question is asked in a certain style, be sure to return the final answer in the same style. For the function $y(x)$, here are some equivalent ways of writing the 1st, 2nd and 3rd derivatives:

$$\frac{\mathrm{d}y}{\mathrm{d}x} \equiv y' \equiv \dot{y} \equiv y_x \qquad \& \qquad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} \equiv y'' \equiv \ddot{y} \equiv y_{xx} \qquad \& \qquad \frac{\mathrm{d}^3y}{\mathrm{d}x^3} \equiv y^{(3)}$$

Notice that some of the notation styles stop being appropriate for high order derivatives. It's also important to remember that it's implicit in the above expressions that y is a function of x, and so the first derivative could equally be written $\frac{\mathrm{d}y(x)}{\mathrm{d}x}$... the "$(x)$" is often left out for convenience (which can occasionally cause confusion). I can only apologise for the lack of consistency and ask you to be brave and press on!

9.1 Back to basics

So, let's start from the beginning. Consider the expression

$$y = 1$$

It seems to be stating that some variable y is equal to 1, but just like any other language, often in maths we require some context to understand what is meant. Based on the discussion above,

you may have correctly guessed that what I implicitly mean is that the function $y(x)$ is equal to 1 for any/all values of x. Graphically, this would just be a horizontal line at $y = 1$.

So far so good, however, the next function requires a little bit more thought.

$$y' = 1$$

If you are asked to "find the solution" to this or "solve" it, what this typically means is for you to find (if possible) an expression in the form "$y = ...$" that satisfies the expression. By our previous experience with calculus, we can see that (through integration), the solution must be of the form

$$y = x + c$$

and because this solution contains an unknown variable (in this case c) we would call it a *general* solution. If you tried to sketch it on a graph, you'd have to draw an infinite number of parallel diagonal lines (gradient=1), where each would have a different value of c.

However, if you were given a bit more information, such as that at $x = 0$ then $y = 2$, you can now work out c and say that the solution must be $y = x + 2$. This is called the *particular* solution.

Similarly, for the function

$$y'' = 1$$

the general solution must be of the form

$$y = \frac{1}{2}x^2 + ax + b$$

because for any values of a and b, y'' will always equal 1. Once again, to find a and b, you'd need to be given more pieces of information (two more pieces to be exact).

Nothing complicated here! But now lets see what happens when things get more interesting.... Consider the equation

$$y = y'$$

If you turn this into a slightly more wordy question, you're being asked, "what function exactly equals the derivative of itself?"

Do you know any?

9.2 A function which is its own derivative

Let's take a minute to recall the original "lim(RoR)" (*i.e.*, "limit of rise over run") definition of a derivative:

$$y'(x) = \lim_{\Delta x \to 0} \left(\frac{y(x + \Delta x) - y(x)}{\Delta x} \right)$$

As we're looking for the special case where $y' = y$, let's just substitute this in and see what happens.

$$y'(x) = \lim_{\Delta x \to 0} \left(\frac{y(x + \Delta x) - y(x)}{\Delta x} \right) = y(x)$$

Now we need a good candidate function to investigate. Let's use the function $y = b^x$, where b is some unknown constant. Why this function? Well if we take some test values for b, we can see that when b is large (*i.e.*, $b = 100$), then the gradient of 100^x is always larger than 100^x; whereas, when b is small (*i.e.*, $b = 0.01$) the gradient of 0.01^x is always smaller than 0.01^x... so perhaps there's a sweet spot in the middle!

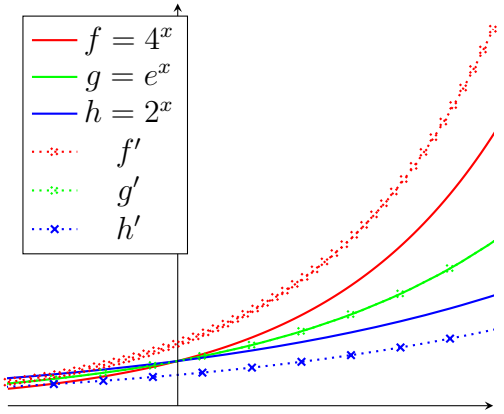
$$b^x = \lim_{\Delta x \to 0} \left(\frac{b^{(x+\Delta x)} - b^x}{\Delta x} \right)$$

This can be rearranged (skipping a few steps here, but make sure you have a go at the algebra on your own if this looks like magic!) to

$$b = \lim_{\Delta x \to 0} \left((1 + \Delta x)^{\frac{1}{\Delta x}} \right)$$

If you try to approximate this on your calculator, you'll probably start to recognise something. At $\Delta x = 0.1$, $b \approx 2.594$; at $\Delta x = 0.01$, $b \approx 2.705$; and at $\Delta x = 0.001$, $b \approx 2.718$... we've just found e, or "Euler's number"! That might seem like a lot of work, just to find something you probably already knew, but Euler's number is at the heart of differential equation analysis, as it's the only function that is its own derivative.... this is going to come in very handy!

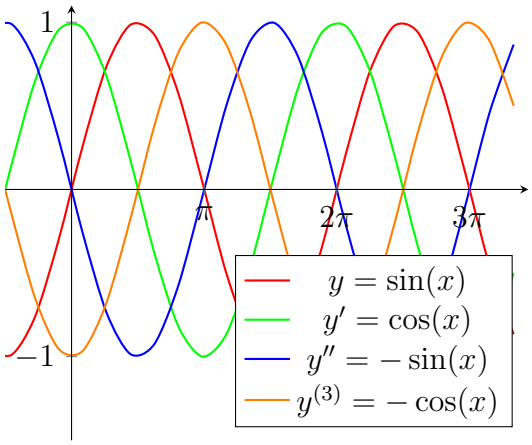
So, if we look again at the ODE $y = y'$, we can now say that the general solution must be of the form $y = ae^x$.



9.2.1 Applying *e* more generally

What about the case of $y = -y'$? In this scenario, the function needs to equal the negative of its derivative, but actually this is no problem, because $y = e^{-x}$ differentiates to $y = -e^{-x}$. Job done. This same logic can also be used to solve $y'' = y$, which is just $y = ae^x + be^{-x}$ (differentiate it twice to check).

However, it's when you consider the case $y = -y''$ that things start to get interesting again. Now we're looking for a function which is the negative of its second derivative, so our e^{-x} trick would not work here because we're differentiating twice and the two negative signs would cancel.



So, can you think of any functions that are the negative of their second derivative... for a hint, see the figure... it's the trigonometric functions! So, now we can write a general solution to $y = -y''$ as $y = A\sin(x) + B\cos(x)$. We have to include both sine and cosine as they both have the negative of themselves as their second derivative. However, if, for example, we were also told that $y = 0$ when $x = 0$, then we'd know that the coefficient B in our general solution must be zero, as

otherwise $B\cos(0)$ would not satisfy this condition. With a second piece of information, such as $y = 3$ when $x = \pi/2$, we can now find our particular solution by simply substituting this into $y = A\sin(x)$, to see that A must equal 3.

You might be wondering why I put so much emphasis on Euler's number if the trigonometric functions are also needed... and you might even be worried that perhaps this means many other functions are needed for special cases... which I might force you to learn for the exam...?

Fear not! Just remember that you can express trig functions in terms of exponentials as follows

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i} \qquad \& \qquad \cos(x) = \frac{e^{ix} + e^{-ix}}{2}$$

So actually, the above $y = -y''$ case was just more Euler's number in disguise. This makes even more sense when you consider that the function $y = e^{ix}$ differentiates to $y' = ie^{ix}$ and then to $y'' = i^2e^{ix} = -e^{ix} = -y$.

9.3 Categories

Various types of ODE exist and recognising which category a specific equation belongs to gives us an indication of how its solution might be found (as well as how difficult the analysis is going to be).

Imagining all possible combinations of variables in an ODE, the vast majority of cases are considered to be ***non-linear***. Non-linear ODEs are typically more difficult to solve and their analysis is beyond the scope of this course. They contain non-linear elements, such as the products of variables/derivatives or terms with exponents, *e.g.*

$$x\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + y^2x - \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = \sin(x) \qquad \text{and} \qquad y\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right) - \sqrt{y} = 0$$

However, the category of ***linear*** ODEs are very useful for describing a variety of physical systems and are also amenable to fairly straightforward analysis. All linear ODEs take the form of a sum of derivatives, each with a coefficient, (example of a second order shown below)

$$\sum_{i=1}^{n} a_i y^{(i)} = f(x) \qquad \xrightarrow{\text{2nd order}} \qquad a\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + b\frac{\mathrm{d}y}{\mathrm{d}x} + cy = f(x)$$

where $f(x)$ is a function of x, which does not need to be linear itself. Furthermore, within the category of linear ODEs, we also have the important sub-category of homogeneous, linear ODEs, which are those expressions for which $f(x) = 0$, *i.e.*,

$$\sum_{i=1}^{n} a_i y^{(i)} = 0 \qquad \xrightarrow{\text{2nd order}} \qquad a\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + b\frac{\mathrm{d}y}{\mathrm{d}x} + cy = 0$$

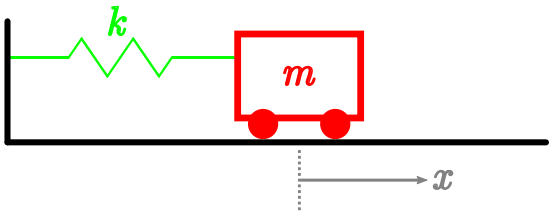
The remainder of this chapter will be dedicated to the analysis of homogeneous, linear, second order ODEs, as they can be used to model several physical systems relevant to engineering.

9.4 ODEs in physical systems

This is the bit where we try to live up to the name "engineering mathematics", by turning the maths into meaning. ODEs can be used to describe a very wide variety of physical systems, but the one that we're going to focus on here is often called the "spring-mass-damper problem".

9.4.1 Spring-mass systems

Imagine a trolley of mass, m, attached to a spring of stiffness, k, that can move on a smooth (frictionless) surface (see figure for colours).



This system can be described by a second-order, linear ODE by summing all of the forces acting on the trolley. Clearly, if the trolley is at rest and the spring is in its neutral position (neither extended or compressed), then there are no forces acting on the trolley. However, if we moved the trolley in the positive x direction, this would extend the spring and the trolley would feel a corresponding force in the negative x direction, pulling it back to the middle. In fact, this force would be negatively proportional to the distance, x, such that $F_{\text{spring}} = -kx$ (where k represents the stiffness of the spring).

We're currently holding the trolley in place, but if we now release it, the force from the spring will accelerate the trolley in the negative x direction (*i.e.*, back toward the middle), from Newton's Second Law, we know that the sum of the forces on an object equals its mass times acceleration, $\Sigma F = ma \equiv m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} \equiv m\ddot{x} \equiv mx''$.

So we can now write an expression for the motion of the trolley, which we can rearrange to see is just a second-order, linear, homogeneous ODE, as discussed above.

$$-kx = ma \xrightarrow{\text{re-express}} m\ddot{x}+kx = 0 \xrightarrow{\text{rearrange}} \ddot{x} = -\frac{k}{m}x \xrightarrow{\text{re-express}} x \propto -\ddot{x}$$

The four expressions above all express the same concept, although the first is perhaps the most descriptive (force balance); the second is the standard for this type of problem; the third statement helps us relate this concept to the discussion in the previous section about derivatives in general; and the final proportionality statement generalises the problem. We are clearly looking for something that is related to the negative of its own second derivative, which as we saw previously is a property of the trigonometric functions.

So, we can say that the general solution for this problem, must be of the form $x = A\sin(\omega_0 t) + B\cos(\omega_0 t)$. Where did the new constant ω_0 come from? Well, compared to our initial discussion, we now have some extra constants (k and m) to account for. Our equation tells us that our second derivative must equal $(-k/m)$ times our solution. Remembering how trig. functions differentiate, we can see that:

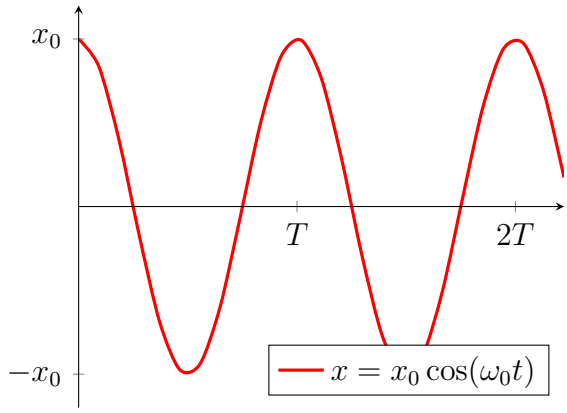
$$
\begin{aligned}
x &= A\sin(\omega_0 t) + B\cos(\omega_0 t),\\
\dot{x} &= A\omega_0\cos(\omega_0 t) - B\omega_0\sin(\omega_0 t),\\
\ddot{x} &= -A\omega_0^2\sin(\omega_0 t) - B\omega_0^2\cos(\omega_0 t) = -\omega_0^2 x
\end{aligned}
$$

Therefore, $\omega_0^2 = \frac{k}{m}$, and so $\omega_0 = \sqrt{\frac{k}{m}}$. Notice that based on this definition ω_0 has the units of 1/time, so we can think of it as a rate or frequency. Referring back to our equations, we can now see that the ω_0 term is a kind of oscillation rate for our trolley system. But how do we find A and B?

9.4.2 Initial conditions

In this idealised system, with no friction/air resistance/energy loss, if we move our trolley away from the middle to an initial starting position x_0 and then let go, its future position as a function of time, $x(t)$, will be described by the general solution $x = A\sin(\omega_0 t) + B\cos(\omega_0 t)$. In fact, now that we know the initial position, we can say that $x = x_0$ when $t = 0$ and substituting this into our general solution gives, $x_0 = A\sin(0) + B\cos(0) = A \times 0 + B \times 1 = B$, so $B = x_0$.

Furthermore, as we know that the trolley initially has zero speed at the instant we release it, we can say that $\dot{x} = 0$ when $t = 0$. Our updated first derivative expression is $\dot{x} = A\omega_0\cos(\omega_0 t) - x_0\omega_0\sin(\omega_0 t)$, so when we substitute our initial speed we get $0 = A\omega_0\cos(0) - x_0\omega_0\sin(0) = A\omega_0$, so $A = 0$.



By considering these two initial conditions, we can now write down our particular solution as $x = x_0\cos(\omega_0 t)$, where $\omega_0 = \sqrt{\frac{k}{m}}$.

Physically, this means that our trolley is oscillating with an amplitude of x_0 and a frequency of $f = \frac{\omega_0}{2\pi} = \frac{\sqrt{k/m}}{2\pi}$, which we refer to as the *resonant* or *characteristic* or *natural* frequency (whereas ω_0 is called the characteristic *angular frequency*). This means that every $T = \frac{2\pi}{\sqrt{k/m}}$ seconds, our trolley returns to the position from which we originally released it (we call t the characteristic time period). As you can see from this expression (and hopefully from the imaginary trolley in your mind), increasing the mass of the trolley or reducing the stiffness of the spring both have the effect of increasing the time period of the oscillation.

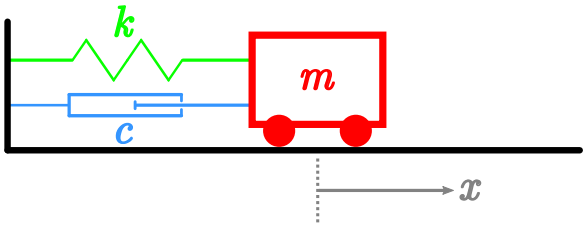
We call these trolley/spring systems "simple harmonic oscillators" and say that they perform **simple harmonic motion** (SHM). Without energy loss, this system must oscillate forever.

Now imagine a slightly different scenario, where we take the same SHM system, but change the initial conditions. This time, rather than giving the trolley an initial displacement, we instead start it in the middle ($x|_{t=0} = 0$), but give it an initial speed, $\dot{x}|_{t=0} \neq 0$ (the vertical bar notation "|" means "such that" or "when", *i.e.*, x when $t = 0$). Working through the analysis

in the same way as the above, but using the two new initial conditions that $x = 0$ and $\dot{x} = v_0$ when $t = 0$, we would now recover the particular solution $x = \frac{v_0}{\omega_0}\sin(\omega_0 t)$. Make sure you can work this through on your own before reading on.

9.4.3 Damping

Now we are going to add another component to the system called a *damper*, which causes a force negatively proportional to the speed, a bit like wind resistance. You can see a simplified representation of a "dashpot damper" in the adjacent figure. The design imagines a sealed tube containing a viscous fluid and a loosely fitting piston. As you move the trolley, the fluid has to squeeze around the piston, which would heat the fluid and dissipate energy (and we can think about c as being a related to the fluid's "viscosity").



Unlike our SHM system from the previous section, we now have a means for the trolley to lose energy. This kind of system is therefore referred to as "damped" and it has the following governing equation, where c is the damping constant:

$$m\ddot{x} + c\dot{x} + kx = 0$$

This will be the first time that we attempt to tackle an ODE with three terms. Fortunately, there is a well defined method for this. We start be re-writing the equation replacing the acceleration/speed/position terms as follows:

$$m\lambda^2 + c\lambda + k = 0$$

We now have what looks like quadratic equation in λ, referred to as the characteristic equation. We can find the roots of this equation by using the quadratic formula as usual.

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad \xrightarrow{\text{substituting}} \qquad \lambda = \frac{-c \pm \sqrt{c^2 - 4mk}}{2m}$$

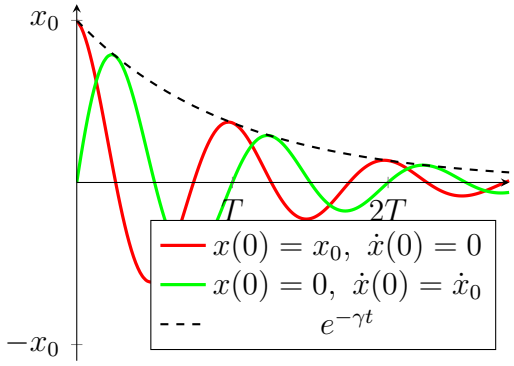
This is where things get interesting. The value of our discriminant ($c^2 - 4mk$) tells us what kind of roots to expect and, assuming m, c and k are all positive (as they would be in nature), there are three possible scenarios.

1. Two complex roots ($c^2 - 4mk < 0$): "Underdamped"

In this scenario, the trolley will still oscillate, but the damper has the effect of gradually draining the kinetic energy from the system. The general solution for this scenario is

$$x = e^{-\gamma t}(A\sin(\omega_d t)) + B\cos(\omega_d t))$$

where $\gamma = \frac{c}{2m}$ is the damping coefficient, and $\omega_d = \sqrt{\omega_0^2 - \gamma^2}$ is the damped natural frequency. The damped natural frequency will always be lower than the undamped natural frequency. As before, in order to find A and B, we need to know the initial conditions of the system (*i.e.*, displacements and velocities).

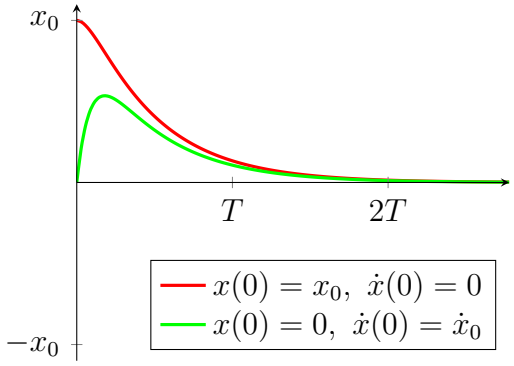


2. Two real roots ($c^2 - 4mk > 0$): "Overdamped"

In this scenario, the trolley damping constant is so high that the motion of the trolley is just a battle between the spring and the damper. No oscillation occurs, just an exponential drift to $x = 0$. The general solution for this scenario is

$$x = Ae^{\lambda_1 t} + Be^{\lambda_2 t}$$

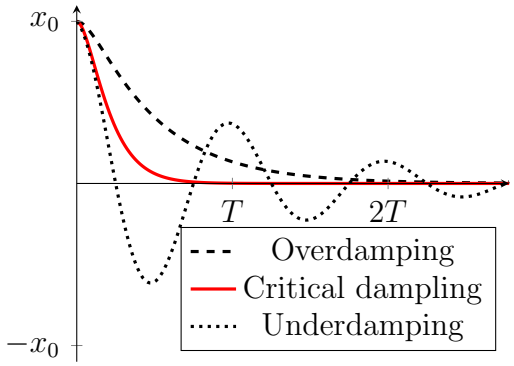
where λ_1 and λ_2 are the roots of the characteristic equation (which will always be negative...).



3. One real root ($c^2 - 4mk = 0$): "Critical damping"

In this final case, the damping is tuned such that the trolley returns to the middle as fast as is possible *without* oscillation. This can be thought of as the scenario in which energy is lost from the system fastest. It is also when the damping constant equals the undamped natural frequency $\gamma = \omega_0$, *i.e.*, $\omega_d = 0$. It has the general solution

$$x = e^{-\gamma t}(A + Bt)$$



Mechanical systems, such as car suspension or nice kitchen draws which close slowly, often use grease with a carefully tuned viscosity to ensure that they are critically damped, in order to dissipate energy from the system as quickly as possible, but avoiding oscillation.

9.4.4 Other physical systems

We can write the general form of the spring-mass-damper problem so that it includes a force as a function of time, $F(t)$, which would make it into a heterogeneous second-order, linear ODE. You'll learn methods for solving these later in the course.

$$m\ddot{x} + c\dot{x} + kx = F(t)$$

Is it also possible to describe other physical systems using simple second-order, linear ODEs. The behaviour of a **series electrical circuit** consisting of an inductor, a resistor and a capacitor (the order is irrelevant) can be described using the following expression,

$$L\frac{\mathrm{d}I}{\mathrm{d}t} + RI + \frac{q}{C} = V(t)$$

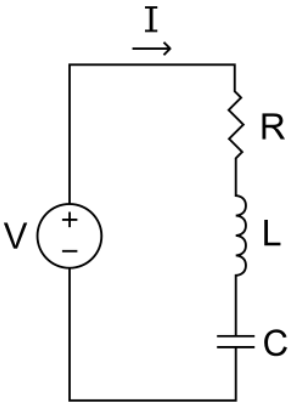
where L is the inductance C is the capacitance, I is the current, R is the resistance, q is the net charge through the system, V is the voltage and t is time. At first this may appear significantly different from the trolley's equation; however, once you've realised that $I = \frac{\mathrm{d}q}{\mathrm{d}t}$, the similarity should be clearer.

$$L\ddot{q} + R\dot{q} + C^{-1}q = V(t)$$

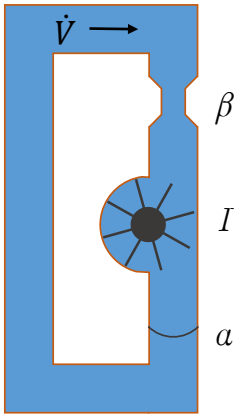
voltage and force, as well as charge and displacement. This in turn implies conceptual translations from inductance to mass; resistance to damping; and inverse capacitance to spring stiffness. And it turns out the the maths really is identical.

Similarly, **idealised hydraulic systems** containing a flywheel (mass or inductor), a constriction (damper or resistor) and a membrane (spring or capacitor), can be modelled using the following expression

$$I\ddot{V} + \beta\dot{V} + \alpha V = P(t)$$

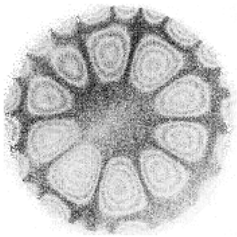


where I is the flow specific moment of inertia of the fly wheel; β is a viscous constriction factor; α is a volume specific membrane stiffness; v is the volume of fluid displaced; P is the pressure drop across the system; and t is time... but even if you didn't know what these terms meant, you could still solve this ODE and discover things about this system. It's important to remember that all of the physical systems discussed in this chapter are only *approximately* described using our simple ODE and lots of assumptions need to be valid for this approximation to be useful... but useful they are!



9.5 ODEs summary

Although we only explored one tiny region of the differential equations universe (linear, homogeneous ODEs), this has already allowed us to build approximate descriptions of several interesting physical systems. We also spent longer than a typical undergraduate course making sure that the foundations of our understanding were secure and I hope this means that when you come to tackle more complex problems, you'll be starting from a solid base.

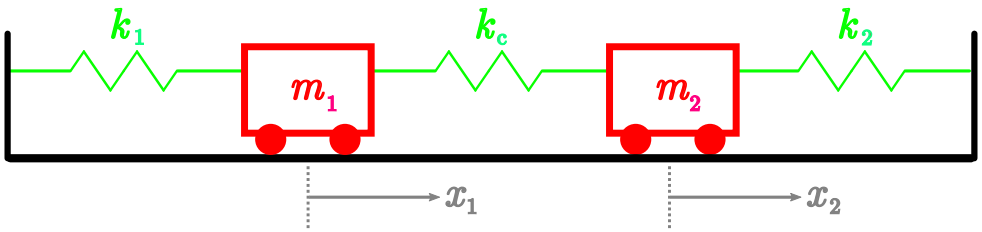


Chapter 10

Coupled Oscillators

This chapter will bring together many of the topics we've covered so far in the course: graph sketching, vectors, matrices, eigenanalysis, complex numbers and ODEs... you're going to love it. Furthermore, it will allow you to start developing insight into how maths can help us model more complex physical systems, with many components coupled (*i.e.*, linked) together.

Take the spring-mass system from the previous chapter (forget the damper for now) and then connect this mass, via another spring, to a second mass, which is itself connected to a wall by a third spring (see figure below).



We can use a similar analysis approach to that laid out in the previous chapter, where we related forces to accelerations, but we will have to consider each of the two masses separately.

As such, we will be considering a separate position variable for each mass, both of which are defined as pointing in the same direction and equal to zero when the system is in static equilibrium (*i.e.*, when x and all its derivative equal zero). For the following analysis, let's also assume that all three spring are neither extended or compressed when the system is in static equilibrium.

10.1 Sum of forces

So, for the left hand mass, m_1, we must consider the effect of both the left hand spring, k_1, and the central connecting spring, k_c. If we move m_1 to the right (*i.e.*, a position $x_1 > 0$), then spring k_1 will act to pull it back towards $x_1 = 0$, just as in the previous chapter. But what about spring k_c?

Well, whether spring k_c is stretched or compressed depends not just on the position of m_1, but also on the position of m_2. More specifically the extension of spring k_c, depends on the relative

location of m_1 to m_2, which we can write as $(x_1 - x_2)$, *i.e.*, if $(x_1 - x_2) > 0$, then spring k_c must be compressed, which act to push m_1 in the negative x_1 direction. Summing the forces acting on m_1 resulting from springs k_1 and k_c, we can say

$$\Sigma F_{m_1} = -k_1 x_1 - k_c(x_1 - x_2) = m_1 \ddot{x}_1$$

Applying the same logic to the second mass, clearly as it moves in the positive x_2 direction, this will compress spring k_2, acting to push it back towards $x_2 = 0$. Similarly, the action of spring k_c will depend on the relative location of the two masses, leading to the equation

$$\Sigma F_{m_2} = -k_2 x_2 - k_c(x_2 - x_1) = m_2 \ddot{x}_2$$

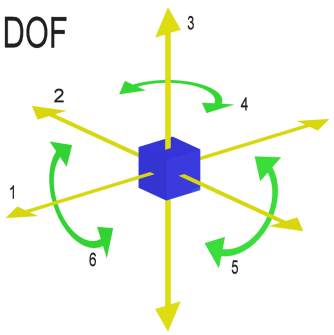
We can take these two equations and write them as a pair of second order differential equations, factorised for the x terms rather than the k terms.

$$\begin{aligned} m_1\ddot{x}_1 + (k_1 + k_c)x_1 - k_c x_2 &= 0 \\ m_2\ddot{x}_2 + (k_2 + k_c)x_2 - k_c x_1 &= 0 \end{aligned}$$

However, based on our work in the previous chapter it's still not clear how to "solve" this system, where, presumably a solution will take the form of a pair of equations, giving the explicit location of m_1 and m_2 (*i.e.*, x_1 and x_2) as a function of time, t.

10.2 Natural frequencies and Eigenmodes

We saw in the ODEs chapter that spring-mass systems have a "natural frequency" at which the system will oscillate if "perturbed" (*i.e.*, , set in motion). This concept extends to systems of coupled masses, but instead of having a single natural frequency, they have one natural frequency per "degree of freedom" (DoF). The number of DoF of a system is the number of variables required to fully describe its independent displacement/rotation. The adjacent diagram shows the six DoF of a body in 3D space, 3 translational and 3 rotational. So, for our two masses moving in a one dimensional model, we have two degrees of freedom only, described by x_1 and x_2.



Let's imagine a scenario where the two masses are vibrating at the same frequency, ω. This will of course have been the result of a certain set of initial conditions, but we're not interested in these for now.

If the masses are oscillating at the same frequency, then their general solutions can be described with the following pair of equations (which we can then differentiate twice to describe their acceleration, as below).

$$\begin{aligned} x_1(t) =& A_1 \sin(\omega t) + B_1 \cos(\omega t) \\ x_2(t) =& A_2 \sin(\omega t) + B_2 \cos(\omega t) \end{aligned}$$

Where the amplitudes, A and B, are constants. If we differentiate these two functions with respect to t, we find the acceleration function, which is simply the negative of the position functions, multiplied by ω^2.

$$\begin{aligned} \ddot{x}_1(t) =& -\omega^2 A_1 \sin(\omega t) - \omega^2 B_1 \cos(\omega t) = -\omega^2 x_1(t) \\ \ddot{x}_2(t) =& -\omega^2 A_2 \sin(\omega t) - \omega^2 B_2 \cos(\omega t) = -\omega^2 x_2(t) \end{aligned}$$

With this in mind, we can go back to the pair of second order differentiation equations that we initially constructed to describe our system and then re-arrange them leaving only the acceleration terms on the right hand side.

$$\begin{aligned} -\frac{(k_1 + k_c)}{m_1} x_1 + \frac{k_c}{m_1} x_2 =& \ddot{x}_1 \\ \frac{k_c}{m_2} x_1 - \frac{(k_2 + k_c)}{m_2} x_2 =& \ddot{x}_2 \end{aligned}$$

Next, we re-express these equations using matrix notation

$$\begin{bmatrix} \frac{-(k_1+k_c)}{m_1} & \frac{k_c}{m_1} \\ \frac{k_c}{m_2} & \frac{-(k_2+k_c)}{m_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix}$$

Pulling in the $\ddot{x} = -\omega^2 x$ result that we found above by investigating the trial general solutions, we can now say

$$\begin{bmatrix} \frac{-(k_1+k_c)}{m_1} & \frac{k_c}{m_1} \\ \frac{k_c}{m_2} & \frac{-(k_2+k_c)}{m_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = -\omega^2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

which should, I hope, remind you of something... what do we call a vector which, when a matrix is applied to it, yields the same result as multiplying by a scalar? An eigenvector of course!

Thinking back to earlier in the course, we can say that to find an eigenvector, we start from the general expression $Ax = \lambda x$ and rearrange to $(A - \lambda I)x = 0$, which has solutions at $\det(A - \lambda I) = 0$. In this case, we've given λ the meaning "$-\omega^2$". So, we need to find

$$\det \begin{bmatrix} \frac{-(k_1+k_c)}{m_1} - \lambda & \frac{k_c}{m_1} \\ \frac{k_c}{m_2} & \frac{-(k_2+k_c)}{m_2} - \lambda \end{bmatrix} = 0$$

At this point, we could work through and find the general algebraic solution for the eigenvalues (*i.e.*, values of "$-\omega^2$") and their corresponding eigenvectors, but the algebra is both messy and unhelpful in terms of intuition building for what the results would mean. So, instead let's put some numbers in using an example...

10.3 Example system

A pair of masses, m_1 and m_2, are coupled to each other via a spring, k_c, and to rigid walls either side of them by two further springs, k_1 and k_2. The positions of the two masses are defined by the variables x_1 and x_2, which are both at zero when the system is at rest and at equilibrium (the same as the first figure in this chapter).

Considering that $m_2 = 3m_1 = 3$ kg and $k_\mathrm{c} = 2k_1 = 3k_2 = 6$ N/m, we can now write down the two governing equations

$$\begin{aligned} m_1\ddot{x}_1 + (k_1 + k_\mathrm{c})x_1 - k_\mathrm{c}x_2 &= 0 \quad \xrightarrow{\text{sub in values}} \quad \ddot{x}_1 + 9x_1 - 6x_2 = 0 \\ m_2\ddot{x}_2 + (k_2 + k_\mathrm{c})x_2 - k_\mathrm{c}x_1 &= 0 \quad \xrightarrow{\text{sub in values}} \quad 3\ddot{x}_2 + 8x_2 - 6x_1 = 0 \end{aligned}$$

and, skipping through the various rearrangements, the matrix for finding eigenvalues becomes

$$\det\begin{bmatrix} \frac{-(k_1+k_\mathrm{c})}{m_1} - \lambda & \frac{k_\mathrm{c}}{m_1} \\ \frac{k_\mathrm{c}}{m_2} & \frac{-(k_2+k_\mathrm{c})}{m_2} - \lambda \end{bmatrix} = 0 \quad \xrightarrow{\text{sub in values}} \quad \det\begin{bmatrix} -9-\lambda & 6 \\ 2 & \frac{-8}{3} - \lambda \end{bmatrix} = 0$$

This returns the two eigenvalues $\lambda_1 = -10.53...$ and $\lambda_2 = -1.14...$, which we can interpret as our two natural frequencies $\omega_1 = \sqrt{-\lambda_1} = 3.24$ s^{-1} and $\omega_2 = \sqrt{-\lambda_2} = 1.07$ s^{-1} (ignoring the negative frequencies resulting from the square roots).

Now that we've got our eigenvalues (telling us our natural frequencies), what can we learn from their corresponding eigenvectors?

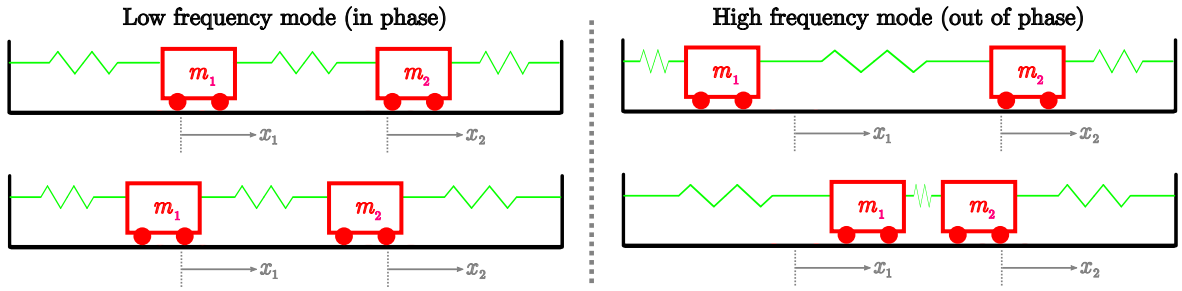
Following our standard eigenvector finding method, when we substitute in the first eigenvalue, we get

$$\begin{bmatrix} -9 - -10.53 & 6 \\ 2 & \frac{-8}{3} - -10.53 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.53 & 6 \\ 2 & 7.86 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

which gives the eigenvector $\mathbf{x}_1 = \begin{bmatrix} -3.93 \\ 1 \end{bmatrix}$. Similarly, for the second eigenvalue, we get the eigenvector $\mathbf{x}_2 = \begin{bmatrix} 0.76 \\ 1 \end{bmatrix}$, but what do this vectors mean?

Delightfully, it tells us amplitudes of oscillation of the two masses at that frequency! So, at $\omega = 3.24$ s^{-1}, the first mass will not only be moving more than the second mass, but also in the opposite direction (*i.e.*, 180° out of phase); and then at $\omega = 1.07$ s^{-1}, the first mass will have a smaller amplitude than the second, but this time they will be moving in the same direction (*i.e.*, in phase with each other).

The figure below illustrates the mode shapes at these two characteristic frequencies, showing to snapshots at the point of maximum displacement of each mass.

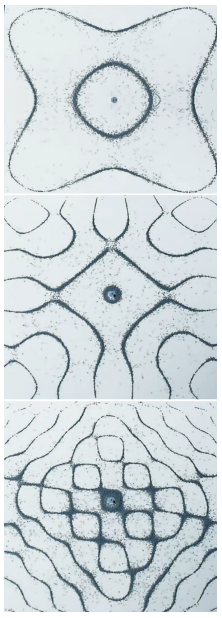


10.4 Generalising

Even for this two mass system, working through the algebra can be quite arduous, but at least it represents something physical, so you should get some clues if you go too wrong (no complex eigenvalues for simple linear systems, for example!). What about systems with more degrees of freedom?

Well, I hope you'd agree that following through the "sum of forces" to the matrix representation wouldn't be too much harder for a 3 mass system, just more messing around with algebra. You'd end up with (3×3) matrix describing the system, which would have 3 real eigenvalues and three eigenvectors. What's more, the same pattern would emerge: at low frequencies, all the masses would be in phase and moving together and then as you go to higher frequency they would increasingly vibrate alternately in opposite directions. So, at the highest natural frequency of a multi-mass system, each mass would be moving in the opposite direction to its neighbours.

10.5 Mind blown

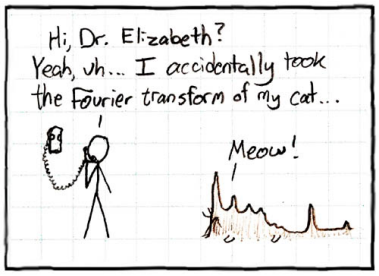


Let's round off this chapter and this term by blowing your minds...

In your mind's eye, zoom in to the atoms of a metal bar or plate. You can think of each nucleus as a little mass, connected to it's neighbours be electromagnetic forces, which you can think of as little springs...

This means that when you inject some energy in to the system (hit it), it should have some preferred frequencies at which it will vibrate, as well as "modes shapes" corresponding to each of those frequencies. These modes shapes are complicated patterns and depend on the exact size and shape of the object, what it's made of and how it's being held. They are also quite beautiful!

Term 2



Chapter 11

The Laplace Transform

11.1 Origins of the Laplace transform

To understand what Laplace transforms are and where they come from, it's useful to think back to power series. In the following equation, we simply state the generic power series formulation, which is the infinite sum of terms containing powers of x^n, each with it's own corresponding coefficient a_n. This polynomial will return us a number if we feed it a value of x, so we can say that it is equal to some function of x, which we will call $A(x)$.

$$a_0 + a_1x + a_2x^2 + a_3x^3 + ... \equiv \sum_0^\infty a_nx^n = A(x)$$

We can now make a small modification to the notation and replace the subscript notation, a_n, with a totally equivalent functional notation, where $a(n)$ is a function which takes a natural number $n = 0, 1, 2, 3...$ and returns another number $a(n)$. So we can now write the near identical expression:

$$\sum_0^\infty a(n)x^n = A(x)$$

However, this expression allows us to see the power series in a new light, where we are really just associating (wiggly arrow) the two functions $a(n)$ and $A(x)$, via the power series.

$$a(n) \rightsquigarrow A(x)$$

So, if you are given a function $a(n)$, which returns coefficients as a function of n, what you get back through this power series association is a function $A(x)$ describing the sum of that power series as a function of x. For example, consider the simple function $a(n) = 1$, which is just the case where for *any* value of n, $a(n)$ will always equal 1. Our power series simply becomes

$$1 + x + x^2 + x^3 + ... \equiv \sum_0^\infty x^n = A(x)$$

Thinking back to our chapter on sequence and series, the above expression is simply a geometric series, where the first term is 1 and the common ratio is x, so substituting this into our expression for the summation of a geometric series to m terms, we get,

$$S_m = a_1\frac{1-r^m}{1-r} \qquad \xrightarrow{\text{substitute}} \qquad A(x) = \lim_{n\to\infty}\left(\frac{1-x^n}{1-x}\right)$$

By analysing the convergence of the above expression, we can see that this summation can only be evaluated when $|x| < 1$ (in which case $\lim_{n\to\infty}(x^n) = 0$). So we can now say that

$$\sum_0^\infty x^n = \frac{1}{1-x} \ , \qquad |x| < 1$$

Bringing all of the above together, we have just found our first example of an association, where in general $a(n) \rightsquigarrow A(x)$ and in our specific case described above $1 \rightsquigarrow \frac{1}{1-x} \ , \quad |x| < 1$.

Let's try another example, think of the case where $a(n) = \frac{1}{n!}$.

$$1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + ... \equiv \sum_0^\infty \frac{1}{n!}x^n = A(x)$$

We saw from our power series chapter that this series converges to e^x for all x, so we've now found our second association.

$$\frac{1}{n!} \rightsquigarrow e^x$$

11.1.1 Discrete to continuous

This is now where the magic happens... Our function $a(n)$ took a natural number as its input, but what if instead we used a function $a(t)$ that would take any positive real number $0 \leq t < \infty$ (*i.e.*, we are going from discrete to continuous). We can now no longer use the discrete sum operation, but instead we must use its continuous analogue, integration.

$$\sum_0^\infty a_n x^n = A(x) \ , \quad |x| < 1 \qquad \xrightarrow{\text{discrete to continuous}} \qquad \int_0^\infty a(t)x^t \, \mathrm{d}t = A(x) \ , \quad 0 < x < 1$$

Similar to the summation case, for this integral to stand any chance of converging, we will need $x < 1$ so that evaluating the limit $t = \infty$ will not cause it to explode. Also, to ensure that the answer is real (as opposed to complex), we must now also keep x positive, so $0 < x < 1$.

Although the integral expression above is a valid form of the variable transformation we are looking for, very often it's more convenient to integrate "e to the power of something" rather than "x to the power of something". So, we simply use our rule of logs and remember that $x = e^{\ln(x)}$ and therefore $x^t = e^{t\ln(x)}$.

$$\int_0^\infty a(t)x^t \, \mathrm{d}t = A(x) \ , \quad 0 < x < 1 \quad \xrightarrow{\text{rearrange}} \quad \int_0^\infty a(t)e^{t\ln(x)} \, \mathrm{d}t = A(x) \ , \quad \ln(x) < 0$$

At this point we are going to make one more adjustment, just to clean everything up. Rather than having this annoying "$\ln(x)$" term (because of the range of x, we also know "$\ln(x)$" is going to be negative), we are simply going to make a substitution. We will introduce a new variable $s = -\ln(x)$

$$\int_0^\infty a(t)e^{t\ln(x)} \, \mathrm{d}t = A(x) \ , \quad \ln(x) < 0 \qquad \xrightarrow{\text{substitute } s} \qquad \int_0^\infty a(t)e^{-st} \, \mathrm{d}t = A(s) \ , \quad s > 0$$

The final thing left for us to do is just to make the expression look more familiar. Currently, we have the function $a(t)$, but more typically, we call functions $f(t)$ (not that this makes any difference, but this is how you will see it written elsewhere). We now have the final expression.

Laplace Transform: $$\int_0^\infty f(t)e^{-st} \, \mathrm{d}t = F(s) \ , \quad s > 0$$

11.1.2 Definitions and details

It's worth noting here that you can now see for yourself the difference between a transform and an operator. The majority of the maths you'll have met so far deals with operators (*e.g.* differentiation), transforms change the variable. It's typical, when transforming a function, $f(t)$, to use the upper case letter F to represent the equivalent transformed function, in order to highlight the link between the two. You will also sometimes see $\tilde{f}(s)$, instead of $F(s)$, as the upper case may already have been used for something else.

$$f(t) \xrightarrow{\text{Transform}} F(s) \qquad\qquad f(t) \xrightarrow{\text{Operator}} g(t)$$

Other notation conventions include the use of the swirly '$\mathcal{L}$' symbol to denote a Laplace transform. So we can write $\mathcal{L}\{f(t)\} = F(s)$, often (but not always) using "braces" {} instead of brackets (). Finally, we can also still use the wiggly arrow for this purpose, writing $f(t) \rightsquigarrow \tilde{f}(s)$.

Another detail that we have not discussed yet is the improper integral $\int_0^\infty$, which needs special treatment. The integral of a function over an infinite interval is the limit of the integral over a finite interval as the bound on the interval tends to infinity. In symbols:

$$\int_0^\infty f(t)\,\mathrm{d}t = \lim_{k\to\infty} \int_0^k f(t)\,\mathrm{d}t$$

So, first solve the finite form of the integral then find the limiting value as we let k tend to ∞.

Finally, the Laplace transform is a *linear* transform, which means that it obeys our convenient rules of linearity. So,

$$\mathcal{L}\{f(t) + g(t)\} = \mathcal{L}\{f(t)\} + \mathcal{L}\{g(t)\}$$
$$\mathcal{L}\{cf(t)\} = c\mathcal{L}\{f(t)\}$$

This will be very handy when we want to transform long expressions, as we can now break them up into parts.

11.2 But what does it mean?

You know where Laplace transforms come from, how to calculate one and how to write them down, but what do they mean? It's easy enough to formally state what a Laplace transform is, but less easy to explain exactly what it does. One clue is that the reason for using t as the variable in our original function $f(t)$ is that we often apply Laplace transforms to functions of time, which gives a small hint as to what s might be.

Formally, we can state that: "Given f, a function of time, with value $f(t)$ at time t, the Laplace transform of f gives us an average value of f taken over all positive values of t such that the value $\tilde{f}(s)$ represents an average of f taken over all possible time intervals of length s." ... Not hugely illuminating.

In fact Laplace transforms are strongly motivated by real engineering problems, where typically we encounter models for the dynamics of phenomena which depend on rates of change of functions, *e.g.* velocities and accelerations of particles or points on rigid bodies, prompting the

use of ordinary differential equations (ODEs). We can use ordinary calculus to solve ODEs, provided that the functions are nicely behaved, which means continuous and with continuous derivatives. Unfortunately, there is much interest in engineering dynamical problems involving functions that input step change or spike impulses to systems (*e.g.* collisions of snooker balls). Now, there is an easy way to smooth out discontinuities in functions of time: simply take an average value over all time. But an ordinary average will replace the function by a constant, so we use a kind of moving average which takes continuous averages over all possible intervals of t. This very neatly deals with the discontinuities by encoding them as a smooth function with interval length s.

The amazing thing about using Laplace Transforms is that we can convert a whole ODE initial value problem into a Laplace transformed version as functions of s, simplify the algebra, find the transformed solution $\tilde{f}(s)$ and then finally undo the transform to get back to the required solution f as a function of t.

Interestingly, it turns out that the transform of a derivative of a function is a simple combination of the transform of the function and its initial value. So a calculus problem can be converted into an algebraic problem involving polynomial functions, which is typically easier to solve.

In this course we find some Laplace Transforms from first principles (*i.e.*, from the definition equation), describe some theorems that help finding more transforms, then use Laplace Transforms to solve problems involving ODEs.

11.3 Finding Laplace Transforms

We have three methods to find $\tilde{f}(s)$ for a given $f(t)$:

1. From the definition: $\mathcal{L}\{f(t)\} = \int_0^\infty f(t)e^{-st}\,\mathrm{d}t = F(s)\ , \quad s > 0$

For $f(t) = 0$, $\mathcal{L}\{0\} = \displaystyle\int_0^\infty 0\,\mathrm{d}t = 0$

For $f(t) = 1$, $\mathcal{L}\{1\} = \displaystyle\int_0^\infty e^{-st}\,\mathrm{d}t = \left[-\frac{1}{s}e^{-st}\right]_0^\infty = \frac{1}{s}$

For $f(t) = t$, $\mathcal{L}\{t\} = \displaystyle\int_0^\infty te^{-st}\,\mathrm{d}t = \left[-\frac{t}{s}e^{-st}\right]_0^\infty + \frac{1}{s}\int_0^\infty e^{-st} = \frac{1}{s^2}$

For $f(t) = \dfrac{\mathrm{d}y}{\mathrm{d}t}$, $\mathcal{L}\left\{\dfrac{\mathrm{d}y}{\mathrm{d}t}\right\} = \displaystyle\int_0^\infty e^{-st}\frac{\mathrm{d}y}{\mathrm{d}t}\,\mathrm{d}t = \left[e^{-st}y\right]_0^\infty + s\int_0^\infty e^{-st}y\,\mathrm{d}t = s\tilde{y}(s) - y(0)$

For $f(t) = \dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}$, $\mathcal{L}\left\{\dfrac{\mathrm{d}^2y}{\mathrm{d}t^2}\right\} = \displaystyle\int_0^\infty e^{-st}\frac{\mathrm{d}^2y}{\mathrm{d}t^2}\,\mathrm{d}t = \text{int. by parts} \times 2 = s^2\tilde{y}(s) - sy(0) - y'(0)$

For $f(t) = e^{at}$, $(a = \text{constant})$

$\mathcal{L}\{e^{at}\} = \displaystyle\int_0^\infty e^{-st}e^{at}\,\mathrm{d}t = \int_0^\infty e^{-(s-a)t}\,\mathrm{d}t = \left[-\frac{1}{s-a}e^{-(s-a)t}\right]_0^\infty = \frac{1}{s-a}, \quad s > a$

2. From a property: There are a number of powerful theorems about the properties of transforms: *e.g.*

$$\mathcal{L}\{af + bg\} = a\mathcal{L}\{f\} + b\mathcal{L}\{g\} \qquad \xrightarrow{\text{for example}} \qquad \mathcal{L}\{3t + 4\} = 3\frac{1}{s^2} + 4\frac{1}{s}$$

Also, as per De Moivre's theorem (*i.e.*, $\cos(at) + i\sin(at) = e^{iat}$):

$$\mathcal{L}\{e^{iat}\} = \frac{1}{s - ia} \qquad \xrightarrow{\text{realise denominator}} \qquad \mathcal{L}\{e^{iat}\} = \frac{s}{s^2 + a^2} + \frac{ia}{s^2 + a^2}$$

Hence, equating real and imaginary parts and using linearity,

$$\mathcal{L}\{\cos(at)\} = \frac{s}{s^2 + a^2} \qquad \& \qquad \mathcal{L}\{\sin(at)\} = \frac{a}{s^2 + a^2}$$

3. From a list: Computer algebra packages like Mathematica, Matlab and Maple know Laplace Transforms of all the functions you are likely to encounter, so you have access to these online, and the packages have also an inversion routine to find a function $f(t)$ from a given $\tilde{f}(s)$. There are books with long lists of transforms of known functions and compositions of functions; *e.g.* some that are harder to calculate:

$$\mathcal{L}\{t^n\} = \frac{n!}{s^{n+1}}, \quad n = 0, 1, 2, \ldots,$$
$$\mathcal{L}\{t^{1/2}\} = \frac{1}{2}\left(\frac{\pi}{s^3}\right)^{1/2},$$
$$\mathcal{L}\{t^{-1/2}\} = \left(\frac{\pi}{s}\right)^{1/2}$$
$$\mathcal{L}\left\{\frac{\mathrm{d}^n y}{\mathrm{d}t^n}\right\} = s^n\tilde{y} - \sum_{k=1}^{n} s^{k-1}y^{(n-k)}(0)$$

11.3.1 Finding inverse transforms using partial fractions

Given a function f, of t, we denote its Laplace Transform by $\mathcal{L}\{f\} = \tilde{f}$; the inverse process is written:

$$\mathcal{L}^{-1}\left\{\tilde{f(s)}\right\} = f(t)$$

A common situation is when $\tilde{f}(s)$ is a polynomial in s, or more generally, a ratio of polynomials; then we use partial fractions to simplify the expressions. Given an expression for a Laplace transform of the form N/D where the numerator, N, and denominator, D, are both polynomials of s; use partial fractions:

1. if N has degree equal to or higher than D, divide N by D until the remainder is of lower degree than D
2. For every linear factor like $(as + b)$ in D, write a partial fraction of the form $A/(as + b)$
3. For every repeated factor like $(as + b)^2$ in D write two partial fractions of the form $A/(as + b)$ and $B/(as + b)^2$. Similarly for every repeated factor like $(as + b)^3$ in D write three partial fractions of the form $A/(as + b)$, $B/(as + b)^2$ and $C/(as + b)^3$; and so on.
4. For quadratic factor $(as^2 + bs + c)$ write a partial fraction $(As + B)/(as^2 + bs + c)$

For repeated quadratic factors write a series of partial fractions, but with numerators of the form $(As + B)$ and successive powers of the quadratic factor as the denominators.

With a little more algebra you should in this way be able to write the original expression as a sum of simpler transforms, which are found in your table. You then add their inverse transforms together, to get the inverse of the original transform.

11.4 Solving ODEs and ODE Systems

The application of Laplace transforms is particularly effective for *linear* ODEs, and for systems of such ODEs. To transform an ODE, we need the appropriate initial values of the function involved and initial values of its derivatives.

Consider the example, from our ODE chapter, of an overdamped harmonic oscillator with the equation shown below. In this case, the trolley starts with an initial displacement and an initial velocity.

$$\ddot{x} + 3\dot{x} + 2x = 0 \qquad \text{where, at} \qquad t = 0: \quad x(0) = 5 \quad \& \quad \dot{x}(0) = 7$$

We can Laplace transform this system by considering each term separately

$$\left[s^2X - sx(0) - \dot{x}(0)\right] + 3\left[sX - x(0)\right] + 2\left[X\right] = 0$$

Then we can factorise to

$$\left(s^2 + 3s + 2\right)X - (s+3)x(0) - \dot{x}(0) = 0$$

Substitute in our initial values

$$\left(s^2 + 3s + 2\right)X - (s+3)5 - 7 = 0$$

Rearranging for X

$$X = \frac{22 + 5s}{s^2 + 3s + 2} \qquad \xRightarrow{\text{partial fractions}} \qquad X = \frac{17}{s+1} - \frac{12}{s+2}$$

Finally, we can look up the inverse transform from a table and write down the solution in the time domain.

$$x = 17e^{-t} - 12e^{-2t}$$

The example we've just worked through was a homogeneous second order ODE. We can also apply Laplace to solve **heterogeneous** ODEs (*i.e.*, where the right hand side of the equation does not equal zero). Modifying our example to include a force which grows with a function of time.

$$\ddot{x} + 3\dot{x} + 2x = 4t - 6 \qquad \text{where, at} \qquad t = 0: \quad x(0) = 5 \quad \& \quad \dot{x}(0) = 7$$

We simple transform each term and grind through the algebra once again:

$$\left[s^2X - sx(0) - \dot{x}(0)\right] + 3\left[sX - x(0)\right] + 2\left[X\right] = \frac{4}{s^2} - \frac{6}{s} \qquad \xRightarrow{\text{many steps...}} \qquad x = 27e^{-t} - 16e^{-2t} + 2t - 6$$

Laplace transforms also allow us to solve systems that we would struggle with in the time domain, such as step inputs and impulse problems, where all we'd need to do is look up the transform of the relevant additional terms and work through the algebra.

Frequently used Laplace Transforms

Function	**Transformed function**
$f(t)$	$\tilde{f}(s) = \int_0^\infty e^{-st} f(t) dt$
0	0
1	$1/s$
t^n, for $n = 0, 1, 2, \ldots$	$n!/s^{n+1}$
$t^{1/2}$	$\frac{1}{2}(\pi/s^3)^{1/2}$
$t^{-1/2}$	$\left(\frac{\pi}{s}\right)^{1/2}$
e^{at}	$1/(s-a)$
$\sin(\omega t)$	$\omega/(s^2+\omega^2)$
$\cos(\omega t)$	$s/(s^2+\omega^2)$
$t\sin(\omega t)$	$2\omega s/(s^2+\omega^2)^2$
$t\cos(\omega t)$	$(s^2-\omega^2)/(s^2+\omega^2)^2$
$e^{at}t^n$	$n!/(s-a)^{n+1}$
$e^{at}\sin(\omega t)$	$w/((s-a)^2+\omega^2)$
$e^{at}\cos(\omega t)$	$(s-a)/((s-a)^2+\omega^2)$
$\sinh(\omega t)$	$\omega/(s^2-\omega^2)$
$\cosh(\omega t)$	$s/(s^2-\omega^2)$
Impulse: (Dirac δ) : $\delta(t-a)$ $\quad (\neq 0$ at $t=a$, else $=0)$	e^{-as}
Step function: $H_a(t) (= 0$ for $t < a$ and $= 1, t \geq a)$	e^{-as}/s
Delay of g: $H_a(t)g(t-a)$	$e^{-as}\tilde{g}(s)$
Shift of g: $e^{at}g(t)$	$\tilde{g}(s-a)$
Convolution: $f(t) * g(t) = \int_0^t f(t-\tau)g(\tau)d\tau$	$\tilde{g}(s)\tilde{f}(s)$
Integration: $1 * g(t) = \int_0^t g(\tau)d\tau$	$\frac{1}{s}\tilde{g}(s)$
Derivative: y'	$s\tilde{y}(s) - y(0)$
Derivative: y''	$s^2\tilde{y}(s) - sy(0) - y'(0)$

Chapter 12

Fourier Series

Fourier series is a way of representing any function in terms of other functions, with some similarities to Taylor series, but also some key differences. Fourier uses a series of "harmonic" functions (*i.e.*, sine and cosine) as its basis and, like the Taylor series, the accuracy of the representation typically depends on the number of terms used. In the Taylor series, each term requires the calculation of a new coefficient in front of a higher power of x (*i.e.* $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4...$), whereas for the Fourier series, each term requires the calculation of new coefficients in front of higher frequency harmonics (*i.e.* $f(x) = a_0 + a_1\cos(x) + b_1\sin(x) + a_2\cos(2x) + b_2\sin(2x) + a_3\cos(3x) + b_3\sin(3x) + ...$).

In Taylor series, a point of interest, c, is chosen, around which the series is expanded and adding more terms allows the approximation to be useful increasingly far from c. With the Fourier series, an interval of interest, $[x_0 - L, x_0 + L]$ is selected and unlike the Taylor series, each additional term makes this approximation a little better (on average) everywhere in the interval simultaneously. Regions outside of the interval are totally ignored by this method, so unless the function is itself periodic (also with period $2L$), the approximation will not converge to it outside of this interval.

There are many alternative ways to express the Fourier series, but the one we will be using on this course is perhaps the most common due to its simplicity of interpretation.

$$g(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right)$$

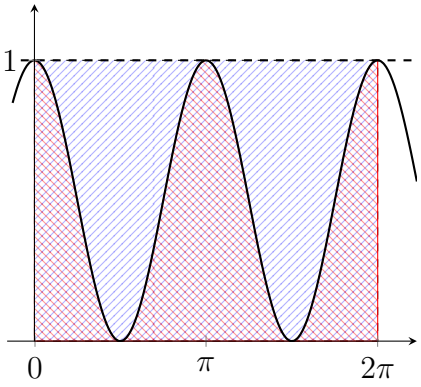
where,

$$a_0 = \frac{1}{L}\int_{-L}^{L} f(x)\,\mathrm{d}x$$
$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{n\pi x}{L}\right)\mathrm{d}x$$
$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{n\pi x}{L}\right)\mathrm{d}x$$

where $f(x)$ is the function you are trying to model in the interval $-L$ to L. If $f(x)$ is a periodic and "integrable" (*i.e.*, "can be integrated") function, with a period of $2L$, then the Fourier series can be used to perfectly recreate it. In some cases it is more convenient to perform the integrals from 0 to $2L$ (rather than from $-L$ to L), which is an equally acceptable definition.

Although these equations might look a little intimidating, its essentially just a set of instructions for calculating the coefficients a_n and b_n that go before each of the harmonic terms and it's up to you to decided how many terms you need, which will depend on the problem you are solving.

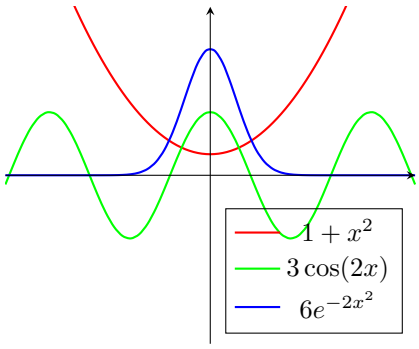
Two questions might immediately spring to mind: Firstly, why is there an a_0, but not a b_0? And secondly, why is a_0 divided by 2? To answer the first question, simply set $n = 0$ in the equation for b_n above. Since $\sin(0) = 0$, so b_0 will always equal zero for any value of $f(x)$ (therefore calculating it would be pointless). The answer to the second question is a little bit more involved, but starts by picturing the function $f(x) = \cos^2(nx)$ (black line in the adjacent figure). Notice that for $n = 1$, the red shaded area under the curve (*i.e.*, the integral) is exactly half the blue area between 0 and 1 in the vertical axis. Furthermore *any* integer value of n will have an integral of π in this range *except* $n = 0$, which has an integral of 2π (see the area under the dashed line in the figure). Hence, we need to renormalise the a_n term to make it resemble the rest of the series. A simpler way to think of the first term is that it's just the mean of the function in the interval...



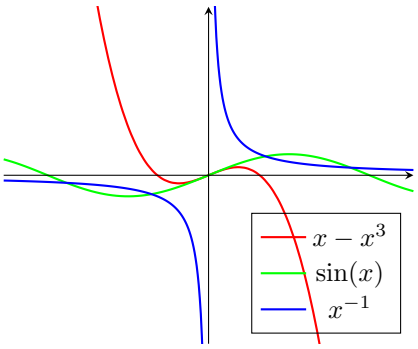
Plot showing $f(x) = \cos^2(nx)$ for $n = 0$ and $n = 1$.

12.1 Symmetry of functions

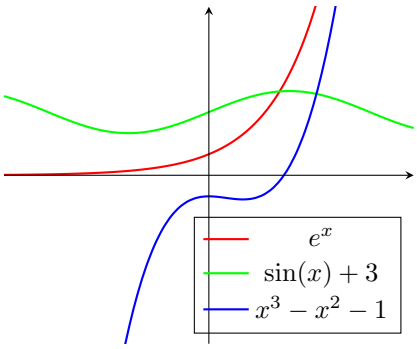
If a function is a symmetrical reflection of itself across the vertical axis, it is referred to as an "**even function**". If rotating a function by 180° around the origin leaves it appearing unchanged, it is referred to as an "**odd function**". If neither of the above criteria are met, then a function is **neither** even nor odd.



Even functions
$f(x) = f(-x)$



Odd functions
$f(x) = -f(-x)$



"Neither" functions

Notice that translating a curve up or down the vertical axis would not affect the *evenness* of an even function, but it would stop the *oddness* of an odd function, making it *neither*. Another way to thinking about this is that even functions are functions whose Taylor series would only use even powers of x (zero is even!) and odd functions are functions whose Taylor series would only use odd powers of x. Therefore, if a function uses both even and odd powers of x in its Taylor series expansion, it is neither even nor odd.

Crucially, if a function is even, then all $b_n = 0$, and if a function is odd, then all $a_n = 0$ (including a_0). Identifying this early on can save a lot of calculation time!

It turns out the functions can be decomposed into their respective even and odd constituents. Where

$$f_{\text{even}}(x) = \frac{f(x) + f(-x))}{2} \qquad f_{\text{odd}}(x) = \frac{f(x) - f(-x))}{2}$$

such that $f(x) = f_{\text{even}}(x) + f_{\text{odd}}(x)$.

Examples:

For the *even* function $g(x) = x^2 + 3$, then $g_{\text{even}}(x) = \frac{(x^2+3)+((-x)^2+3)}{2} = x^2 + 3 = f(x)$ and $g_{\text{odd}}(x) = \frac{(x^2+3)-((-x)^2+3))}{2} = 0$.

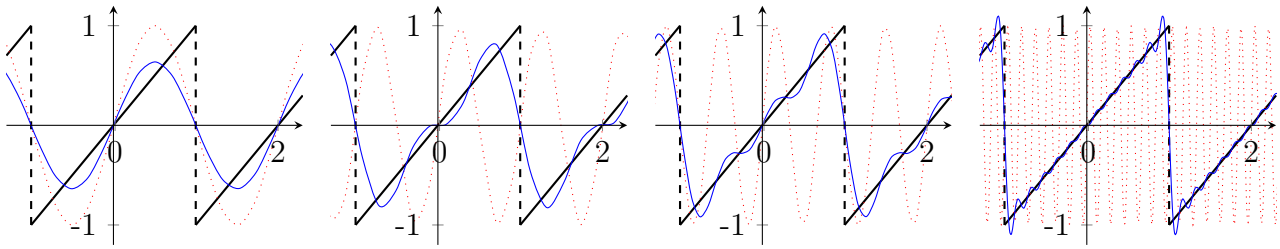
For the *odd* function $h(x) = x^{-1}$, then $h_{\text{even}}(x) = \frac{x^{-1}+(-x)^{-1}}{2} = 0$ and $h_{\text{odd}}(x) = \frac{x^{-1}-(-x)^{-1}}{2} = x^{-1} = h(x)$.

For the *neither* function $p(x) = \sin(x) + 3$, then $p_{\text{even}}(x) = \frac{(\sin(x)+3)+(\sin(-x)+3)}{2} = 3$ and $p_{\text{odd}}(x) = \frac{(\sin(x)+3)-(\sin(-x)+3)}{2} = \sin(x)$.

What about the function $f(x) = e^x$ (consider it's Taylor series... what's the pattern of even and odd terms)?

12.2 Periodic functions

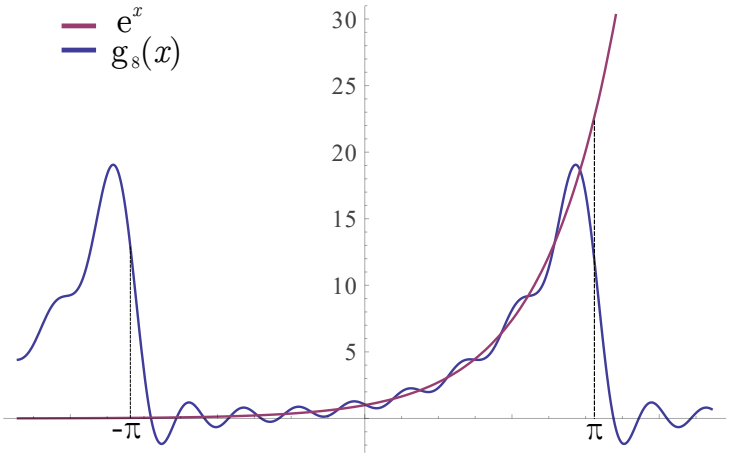
Periodic functions are composed of an infinite sequence of identical repeat units, examples include the harmonic functions (*i.e.*, sine, cosine and tangent) and various discontinuous functions such as the "saw-tooth" and "square wave" functions. The Fourier series is well suited to approximating periodic functions, even if they contain discontinuities.



First Second Third Twelfth

The four graphs above show the odd "saw-tooth" function (black), as well as the first, second, third and twelfth order Fourier approximations (blue). Also shown (red dots) are the profiles of the highest frequency sine waves used in each case. Notice how, unlike the Taylor series, the Fourier approximations improve the function (on average) at all points simultaneously. Also notice the high frequency wiggles close to the discontinuity, which are referred to as *Gibbs ringing*.

The figure on the right shows the Fourier expansion (up to the 8^{th} order) of the function $f(x) = e^x$ in the interval $-\pi$ to π. Clearly, $f(x) = e^x$ is not a periodic function, so although the approximation appears to be reasonable in the interval considered, it is totally useless outside of this domain. However, as use of the domain $[-\pi, \pi]$ was arbitrary, what's to stop us from making this interval much larger? As we shall see later on in this chapter, we can even investigate the case where the interval $[-\infty, \infty]$ is used, which yields a very powerful result!



12.3 Complex exponential representation

Previously in the course we saw that the trigonometric functions could be written as complex exponentials and vice versa, i.e.,

$$e^{ix} = \cos(x) + i\sin(x)$$

$$\Rightarrow \qquad \cos(x) = \frac{e^{ix} + e^{-ix}}{2} \qquad \& \qquad \sin(x) = \frac{e^{ix} - e^{-ix}}{2i}$$

Using these we can reformulate the Fourier Series in terms of complex exponentials, rather than sines and cosines. This simplifies the representation to,

$$g(x) = \sum_{n=-\infty}^{\infty} C_n \exp\left(\frac{in\pi x}{L}\right) ,$$

with,

$$C_n = \frac{1}{2L}\int_{-L}^{L} f(x)\exp\left(-\frac{in\pi x}{L}\right) \mathrm{d}x \, .$$

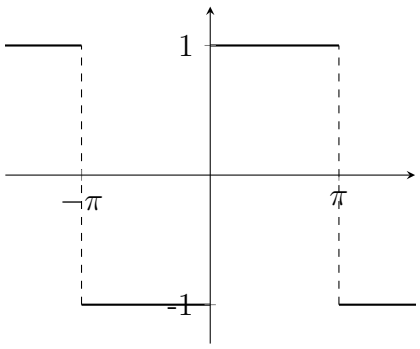
Note here, that this is a sum over both positive and negative values of n, and that the *zeroth* term is included in the sum, rather than being separate in the cosine case. The complex exponential case is completely equivalent to the sine and cosine representation; A truncated Fourier series with exponential terms from $-N$ to N will give exactly the same function as sine and cosine terms from $n = 0$ to N. The sinusoidal and complex exponential coefficients are related as,

$$a_n = C_n + C_{-n} \qquad \& \qquad b_n = i(C_n - C_{-n}) \, .$$

This description has the advantage of being simpler, there is only one type of term and the coefficients all have the same form, at the price of introducing complex numbers to describe an entirely real function (For real functions, the coefficients are constrained such that $C_{-n} = C_n^*$).

12.3.1 Example - Square wave

We will now walk through the full calculations for the example of a square wave, as shown in the adjacent figure. The function has a period of 2π and is discontinuous and odd.



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

Firstly, as we have noticed that $f(x)$ is an *odd* function, we can be sure that all the a_n terms are zero (including a_0). Also, as the period is 2π, we can replace L in our standard equation with π. This means that our Fourier series will be of the form

$$g(x) = \sum_{n=1}^{\infty} b_n \sin(nx), \quad \text{where} \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx$$

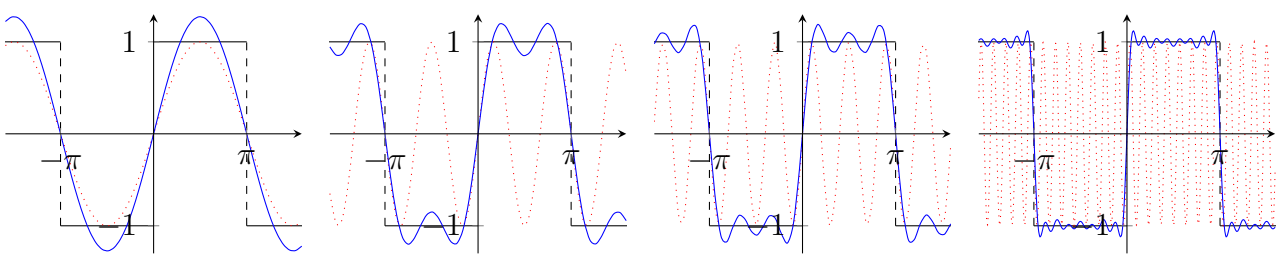
To evaluate b_n at each value of n, we must consider the two regions ([$-\pi$,0] and [0,π]) of this discontinuous function separately. To do this, we simply integrate first from $-\pi$ to 0 where $f(x) = -1$ and from 0 to $+\pi$ where $f(x) = +1$; our constant is the sum of these two integrals.

$$\begin{aligned} b_1 &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(x)dx \\ &= \frac{1}{\pi}\left(\int_{-\pi}^{0}(-1)\sin(x)dx + \int_{0}^{\pi}(1)\sin(x)dx\right) \\ &= \frac{1}{\pi}\left([\cos(x)]_{-\pi}^{0} + [-\cos(x)]_{0}^{\pi}\right) \\ &= \frac{1}{\pi}(2+2) = \frac{4}{\pi} \end{aligned}$$

$$\begin{aligned} b_2 &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(2x)dx \\ &= \frac{1}{\pi}\left(\int_{-\pi}^{0}(-1)\sin(2x)dx + \int_{0}^{\pi}(1)\sin(2x)dx\right) \\ &= \frac{1}{\pi}\left(\left[\frac{1}{2}\cos(2x)\right]_{-\pi}^{0} + \left[-\frac{1}{2}\cos(2x)\right]_{0}^{\pi}\right) \\ &= \frac{1}{\pi}(0+0) = 0 \end{aligned}$$

As is often the case with Fourier series, we can now save ourselves a lot of work by looking for patterns in the behaviour of the coefficients. From studying the calculation for b_2, it should be clear that for *any* even value of n, the result will always be zero. Similarly, for odd values of n, the calculation for b_1 shows us that the only difference between each term will be a factor of $\frac{1}{n}$. We can now write down our Fourier series approximation for this function and plot the effect of truncation. Blue is the approximation and red is the highest frequency included.

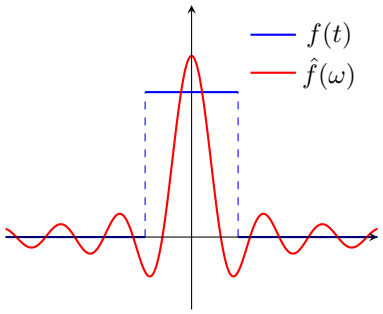
$$g(x) = \frac{4}{\pi}\left[\sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + ...\right] = \frac{4}{\pi}\sum_{n=1,3,5,...}^{\infty}\frac{1}{n}\sin(nx)$$



$n = 1$ $n = 3$ $n = 5$ $n = 15$

12.4 Fourier transform

Although we won't be studying the transform in much detail, take some time to look back at the graph of the function $f(x) = e^x$ and its Fourier series approximation. Although it seems clear that this approach has some limitations for non-periodic functions, consider what happens when we imagine the period to be infinite? This means that the problem we have with the "fake" periodic discontinuities should disappear, which is exactly what the Fourier *transform* allows us to do. This means that if we take a signal in the time domain, $f(t)$, we should now be able to write a new expression for this in the *frequency domain*, $\hat{f}(\omega)$.



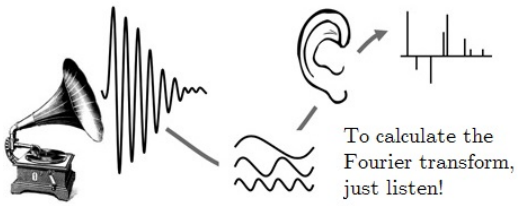
The integral below does just that, where the sine and cosine functions that we used in the series expansion have been replaced with a complex exponential (*N.B.* $(e^{iz})^n = \cos(nz) + i\sin(nz)$). Furthermore, if we have started with frequency data, there is also an inverse Fourier transform for recovering the time signal. The figure above shows a function and its transpose together.

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i t\omega}dt \qquad f(t) = \int_{-\infty}^{\infty} \hat{f}(\omega)e^{2\pi i t\omega}dw$$

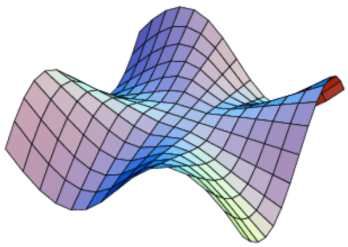
The relationship between time and frequency is not the only useful pairing and the same concept relates position and momentum. Fourier transforms are also very useful in calculus, where a differentiation in the time domain can be as simple as a multiplication in frequency space.

12.5 Mind blown

Pretty much everything about Joseph Fourier's (1768-1830) work is mind blowing. Fourier analysis is at the heart of so much of modern technology, from the compression of images (this is how a `jpeg` file is so small, but also why sharp edges get blurred) to the transmission of sound messages by your phone. What is perhaps even more astonishing, although it shouldn't be surprising, is that nature also makes use of this concept.



Inside your ear, there is a tube full of tiny hairs and when a sound wave enters your ear, it causes some of these hairs to resonate. Low frequencies stimulate the soft region near the entrance, whereas high frequency activate the narrow end of the tube. Real sounds are rarely perfect sine waves, but rather a superposition of many waves all with different phases and frequencies. If your brain simply used a pressure sensor to measure incoming sound, it would need to perform some very complicated numerical analysis in order to reconstruct the incoming waves. Instead, most of this work is done by the mechanical Fourier transform performed by the hairs, which simply tell the brain which frequencies are incoming and how loud they are.



Chapter 13

Multivariate Calculus

13.1 Functions of multiple variables

So far we have been dealing with functions that take a number, e.g. x, as input and return a number, $f(x)$, as output. We can upgrade to functions of multiple variables, or *multivariate functions*. i.e.,

$$f(x, y) = 4x \sin(y) + 5e^{\frac{y}{x}} .$$

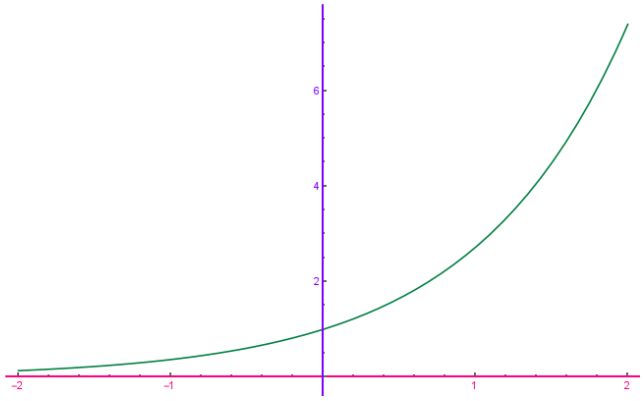
This function takes two inputs, x and y and returns a single number $f(x, y)$ as the result. In principle, our multivariate functions could also return a vector. We will cover this later in the chapter, but for now we'll concentrate on the case where a function takes multiple inputs and returns one output as there's lots to say about this case before vectorising everything!

The first thing to consider is how to actually visualise these functions. 2D functions aren't generally a problem since the 2-input and 1-output adds upto a 3D space that we can visualise. The common representations include surface plots, contour plots, and density plots.

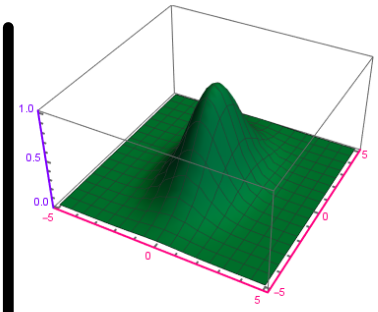
An example of some of these is included on the next page - here input dimensions are drawn in pink, and output dimensions in purple.

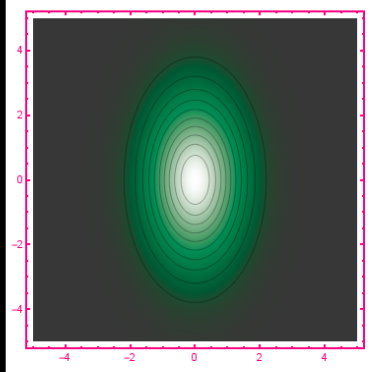
For higher dimensions, we'll have to rely on the intuition built up in the 2D case, as we'll no longer be able to plot the function in its entirety.

Takes 1 input; Returns 1 output

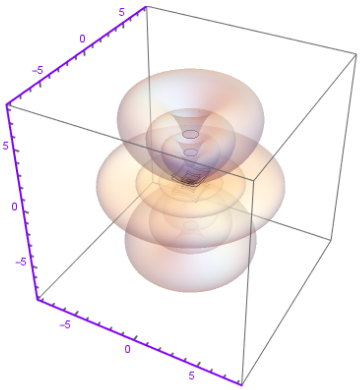


Takes 2 inputs; Returns 1 output

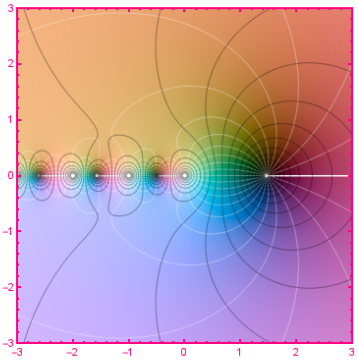


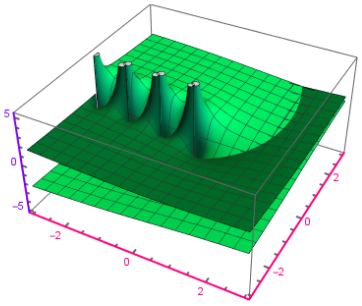


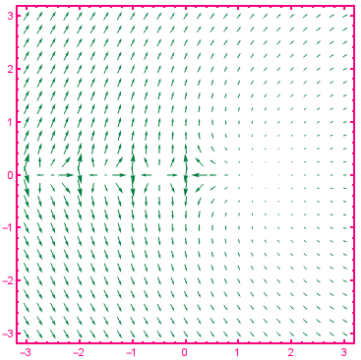
Takes 3 inputs; Returns 1 output



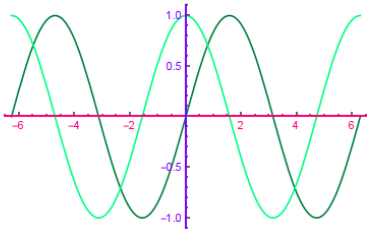
Takes 2 inputs, Returns 2 outputs

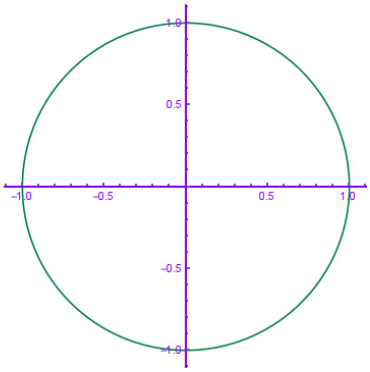


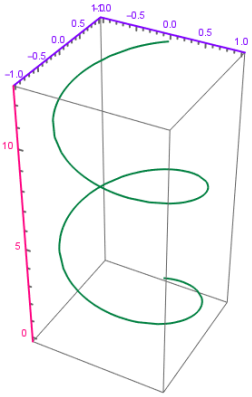




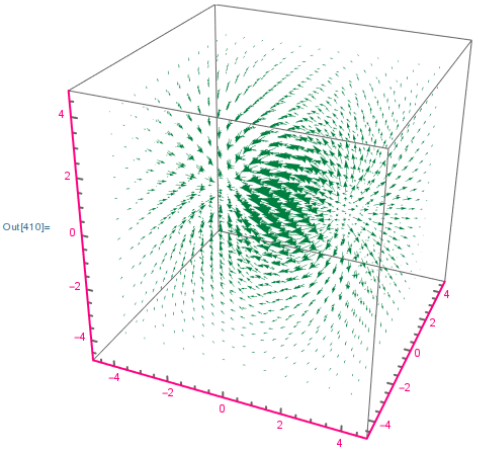
Takes 1 input, Returns 2 outputs







Takes 3 inputs, Returns 3 outputs



13.2 Partial derivatives

Now we have multivariate functions, can we do calculus with them? The answer is of-course YES! but we'll need to upgrade our machinery.

For a function $f(x, y, z)$, we can now differentiate with respect to either x, y, or z. We'll define the *partial derivative* as differentiating with respect to one variable, *whilst keeping the others constant*. e.g.,

$$\left(\frac{\partial f}{\partial x}\right)_{yz} ,$$

means taking the derivative of $f(x, y)$ with respect to x whilst holding y and z constant. Notice the curly-d, ∂, this is pronounced *'partial'*. Sometimes the little subscript indicating which variable is being held constant is omitted when it's otherwise clear what's going on.

There are other notations you might see in the wild,

$$\left(\frac{\partial f(x,y)}{\partial x}\right)_{yz} = \frac{\partial f}{\partial x} = \partial_x f = f_x ,$$

Let's see a couple of examples; consider,

$$f(x, y) = 3x^2y + 2y + xy^2 ,$$

differentiating with respect to x, whilst treating y as if it was just a constant, gives,

$$\begin{aligned}\left(\frac{\partial f}{\partial x}\right)_y &= 6xy + 0 + y^2 \\ &= 6xy + y^2 .\end{aligned}$$

We can do the same differentiating with respect to y,

$$\left(\frac{\partial f}{\partial y}\right)_x = 3x^2 + 2 + 2xy .$$

Let's go again with a more complicated example,

$$\begin{aligned}f(x, y) &= 4x\sin(y) + 5e^{\frac{y}{x}} \\ \frac{\partial f}{\partial x} &= 4\sin(y) - \frac{5y}{x^2}e^{\frac{y}{x}} \\ \frac{\partial f}{\partial y} &= 4x\cos(y) + \frac{5}{x}e^{\frac{y}{x}} .\end{aligned}$$

Make sure you see how these results are obtained by holding one varible constant. (It may help you to replace the constant variable with another letter, like a, to give the impression that these really are just constants.)

This gives us all we need to upgrade most of our 4 differentiation rules.

Sum rule:

$$\frac{\partial}{\partial x}\left[f(x, y) + g(x, y)\right] = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x}$$

Power rule:

$$\frac{\partial}{\partial x}\left[x^n f(y)\right] = nx^{n-1} f(y)$$

Product rule:

$$\frac{\partial}{\partial x}\left[f(x,y)\, g(x,y)\right] = \frac{\partial f(x,y)}{\partial x}\, g(x,y) + f(x,y)\, \frac{\partial g(x,y)}{\partial x}$$

Chain rule (part 1):

$$\frac{\partial}{\partial x} f(g(x,y)) = \frac{\mathrm{d}f}{\mathrm{d}g}\frac{\partial g}{\partial x}$$

Notice here that f is a function of one variable, when we differentiate it we use ordinary upright ds, as $\frac{\mathrm{d}f}{\mathrm{d}g}$ is just the derivative of f with respect to its single argument.

We'll have to wait until later in the chapter to see what happens if f was also a function of multiple variables.

13.2.1 Higher order derivatives

We can differentiate more than once to give us higher order derivatives. There are now more options as to how to do this. e.g. for,

$$\begin{aligned} f(x,y) &= 2ye^{3x} - x^3y^2 + y^5 \\ \frac{\partial f}{\partial x} &= 6ye^{3x} - 3x^2y^2 \\ \frac{\partial^2 f}{\partial x^2} &= 18ye^{3x} - 6xy^2 \ . \end{aligned}$$

So far so good, but we could have decided to differentiate with respect to y after the first x derivative to form a mixed derivative,

$$\frac{\partial}{\partial y}\frac{\partial f}{\partial x} = \frac{\partial^2 f}{\partial y \partial x} = 6e^{3x} - 6x^2y$$

Let's start by differentiating with respect to y,

$$\begin{aligned} f(x,y) &= 2ye^{3x} - x^3y^2 + y^5 \\ \frac{\partial f}{\partial y} &= 2e^{3x} - 2x^3y + 5y^4 \\ \frac{\partial^2 f}{\partial y^2} &= -2x^3 + 20y^3 \\ \frac{\partial^2 f}{\partial x \partial y} &= 6e^{3x} - 6x^2y \ . \end{aligned}$$

Note how $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$. In general it doesn't matter the order your differentiate by.

13.3 Stationary points

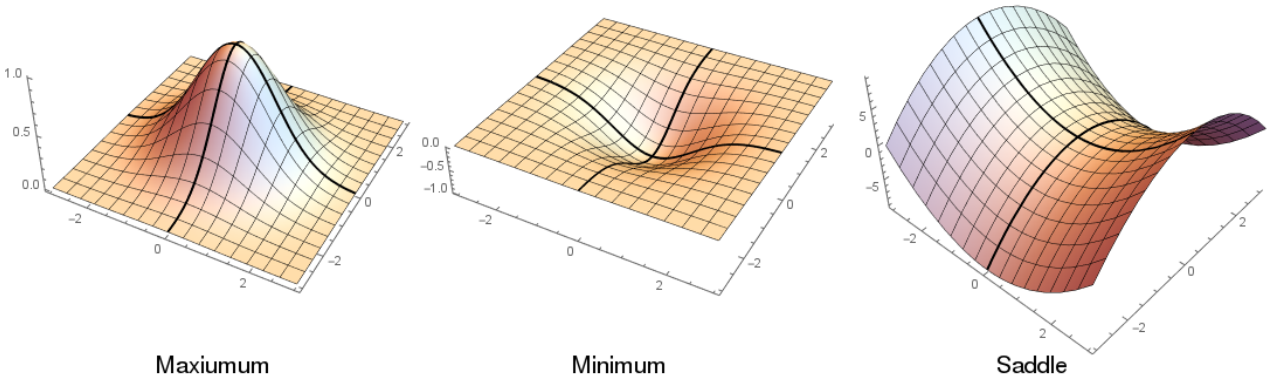
Let's look into stationary points (turning points) for multivariate functions. Here for a turning point, all partial derivatives need to be zero. i.e. $f_x = 0$ and $f_y = 0$. Consider the function,

$$f(x, y) = x^2 + xy + y^2 - 5x - y + 7 \ ,$$

the derivatives are,

$$\begin{aligned} f_x &= 2x + y - 5 \\ f_y &= x + 2y - 1 \ , \end{aligned}$$

Setting these equal to zero, we have a pair of simultaneous equations that have the solution, $x = 3, y = -1$. This indicates that there's a stationary point at $(3, -1)$, but what is it's character? a maximum, minumum, or something else? We'll cover later in the module how you can tell, it will involve the second derivative (and the cross terms, f_{xy}) In addition to maxima, minima, and inflection points, there is another point of interest that appears in multi-dimensional systems, a saddle point, which is a special kind of inflection point. This is where a stationary point appears like a maximum from one direction and a minimum from another.



13.4 Total differentials and derivatives

We have so far looked at the derivatives along x and y, but we may want to take the derivative along other directions, or a curved path within the x, y space; for this we can use the total derivative. Let's consider the change in the function, $f(x, y)$ as we take a small step away, i.e.,

$$\Delta f = f(x + \Delta x, y + \Delta y) - f(x, y)$$

Then let's do a bit of mathematical trickery, including adding zero and multiplying by one.

$$\Delta f = f(x + \Delta x, y + \Delta y) \underbrace{- f(x, y + \Delta y) + f(x, y + \Delta y)}_{=0} - f(x, y)$$

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

In the limits as we take Δx and Δy to infinitesimals we can see the rise over run definition of the partial derivatives. So we get the following expression for the total derivative,

$$\mathrm{d}f = \left(\frac{\partial f}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial f}{\partial y}\right)_x \mathrm{d}y$$

(or in more than two dimensions, this generalises as you'd expect)

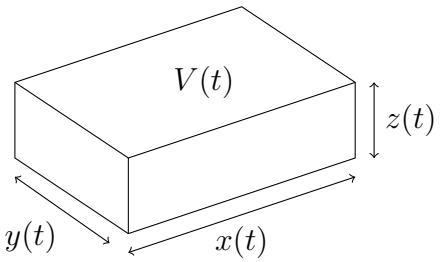
$$\mathrm{d}f = \left(\frac{\partial f}{\partial x}\right)_{yz\ldots} \mathrm{d}x + \left(\frac{\partial f}{\partial y}\right)_{xz\ldots} \mathrm{d}y + \left(\frac{\partial f}{\partial z}\right)_{xy\ldots} \mathrm{d}z + \ldots$$

Let's take an example to illustrate how the a total differential can be used.

Example: "A box has sides x, y, z changing in length over time t. Find the rate of change in volume."

The volume is $V(x, y, z) = xyz$. And x, y and z are all changing functions of time t:

$$x = x(t) \quad y = y(t) \quad z = z(t)$$



Let's set up the total differential for this problem,

$$\mathrm{d}V = \left(\frac{\partial V}{\partial x}\right)_{yz} \mathrm{d}x + \left(\frac{\partial V}{\partial y}\right)_{xz} \mathrm{d}y + \left(\frac{\partial V}{\partial z}\right)_{xy} \mathrm{d}z$$

In this example, if we calculate the partial derivatives, $\frac{\partial V}{\partial x} = yz$, $\frac{\partial V}{\partial y} = xz$, and $\frac{\partial V}{\partial z} = xy$.

The derivatives $\mathrm{d}x \,/\, \mathrm{d}t$, $\mathrm{d}y \,/\, \mathrm{d}t$ and $\mathrm{d}z \,/\, \mathrm{d}t$ are either given as functions of t or simply as values (*e.g.* 'x is decreasing at 0.2 metres per second' means $\mathrm{d}x \,/\, \mathrm{d}t = -0.2$ if SI units were assumed.)

In general, given a function $f(x, y, z \ldots)$, where the variables $x, y, z, \ldots$ are each functions of another variable t, then the **total derivative** is given by:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = f_x \frac{\mathrm{d}x}{\mathrm{d}t} + f_y \frac{\mathrm{d}y}{\mathrm{d}t} + f_z \frac{\mathrm{d}z}{\mathrm{d}t} + \ldots$$

In some cases, you might have a function of some variables which are related to each other through a single underlying variable, *as well as* the underlying variable itself.

Example: Consider the function $f(x, y, z, t)$ then the total derivative with respect to t can be expressed as:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t}\frac{\mathrm{d}t}{\mathrm{d}t} + \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial f}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t}$$

If the velocity field is defined as $u = \mathrm{d}x \,/\, \mathrm{d}t$, $v = \mathrm{d}y \,/\, \mathrm{d}t$ and $w = \mathrm{d}z \,/\, \mathrm{d}t$ then:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + u\frac{\partial f}{\partial x} + v\frac{\partial f}{\partial y} + w\frac{\partial f}{\partial z}$$

which is sometimes also referred to as the "material derivative" and written $\frac{\mathrm{D}f}{\mathrm{D}t}$.

13.4.1 Chain rule

With the total differential, we can complete our upgrade to the chain rule. i.e. what happens when we have, f a function of u and v, which are each functions of both x and y,

$$\frac{\partial}{\partial y} f(u(x,y), v(x,y)) \ .$$

We can solve this by writing the total derivative,

$$\mathrm{d}f = \left(\frac{\partial f}{\partial u}\right)_v \mathrm{d}u + \left(\frac{\partial f}{\partial v}\right)_u \mathrm{d}v \ ,$$

from which we can derive,

$$\left(\frac{\partial f}{\partial y}\right)_x = \left(\frac{\partial f}{\partial u}\right)_v \left(\frac{\partial u}{\partial y}\right)_x + \left(\frac{\partial f}{\partial v}\right)_u \left(\frac{\partial v}{\partial y}\right)_x \ ,$$

which is the fully upgraded chain rule.

Let's test this with an example, let $f(u,v) = \sqrt{u^2 - v^2}$, $u(x,y) = y/x$, $v(x,y) = x + y$, Then,

$$\begin{aligned}
\left(\frac{\partial f}{\partial u}\right)_v &= \frac{u}{\sqrt{u^2 - v^2}} \\
\left(\frac{\partial f}{\partial v}\right)_u &= \frac{-v}{\sqrt{u^2 - v^2}} \\
\left(\frac{\partial u}{\partial y}\right)_x &= \frac{1}{x} \\
\left(\frac{\partial v}{\partial y}\right)_x &= 1 \ .
\end{aligned}$$

Putting these together,

$$\left(\frac{\partial f}{\partial y}\right)_x = \frac{u}{\sqrt{u^2 - v^2}} \frac{1}{x} + \frac{-v}{\sqrt{u^2 - v^2}} (\times 1) \ ,$$

and replacing u and v, and rearranging,

$$\left(\frac{\partial f}{\partial y}\right)_x = \frac{y/x^2 - x - y}{\sqrt{(y/x)^2 - (x+y)^2}} \ ,$$

Let's see if we can get the same answer by direct substitution.

$$\begin{aligned}
f(u(x,y), v(x,y)) &= \sqrt{(y/x)^2 - (x+y)^2} \\
\frac{\partial f}{\partial y} &= \frac{1}{2} \frac{1}{\sqrt{(y/x)^2 - (x+y)^2}} \cdot \left(\frac{2y}{x^2} - 2(x+y)\right) \\
&= \frac{y/x^2 - x - y}{\sqrt{(y/x)^2 - (x+y)^2}} \ .
\end{aligned}$$

As we would hope, this comes out the same as previously. Often in particularly complicated examples, the full chain rule can save time and effort.

13.5 Vector calculus

13.5.1 Vector functions

It can be of benefit to vectorise our functions, i.e.

$$f(x, y, z) \to f(\mathbf{x})$$

This is especially the case if you have varibles that naturally bundle together, like physical spatial dimensions, or if you have a large number of variables to keep track of, i.e. in data science.

In addition to functions that take a vector as input, there are functions that return a vector as output, and functions that do both.

Functions that take values at every point in space sometimes get called fields, i.e. the temperature at any point, $T(\mathbf{x})$, in a room is a scalar field, and the velocity of the airflow at any point in a room, $\mathbf{v}(\mathbf{x})$, is a vector field.

differentiating a vector function with respect to a scalar is easy, it differentiates component wise, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{F}(t) = \frac{\mathrm{d}F_x(t)}{\mathrm{d}t}\mathbf{i} + \frac{\mathrm{d}F_y(t)}{\mathrm{d}t}\mathbf{j} + \frac{\mathrm{d}F_z(t)}{\mathrm{d}t}\mathbf{k}\,.$$

Identities like the chain rule, work as you might expect,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(a(t)\mathbf{b}(t)\right) &= \frac{\mathrm{d}a(t)}{\mathrm{d}t}\mathbf{b}(t) + a(t)\frac{\mathrm{d}\mathbf{b}(t)}{\mathrm{d}t} \\
\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{a}(t)\cdot\mathbf{b}(t)\right) &= \frac{\mathrm{d}\mathbf{a}(t)}{\mathrm{d}t}\cdot\mathbf{b}(t) + \mathbf{a}(t)\cdot\frac{\mathrm{d}\mathbf{b}(t)}{\mathrm{d}t} \\
\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{a}(t)\times\mathbf{b}(t)\right) &= \frac{\mathrm{d}\mathbf{a}(t)}{\mathrm{d}t}\times\mathbf{b}(t) + \mathbf{a}(t)\times\frac{\mathrm{d}\mathbf{b}(t)}{\mathrm{d}t}\,.
\end{aligned}$$

13.5.2 The del operator

When the argument is a vector, we can introduce the vector differential operator ∇, pronounced *del* (the symbol itself is called *nabla*. The operator can also be written as $\frac{\partial}{\partial\mathbf{x}}$), which behaves like a vector with the components,

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z} + \dots\,,$$

there are a number of useful ways to apply this, depending what kind of quantity we have.

where we have coordinate vector $\mathbf{x} = (x, y, z)^{\mathrm{T}}$, $f(\mathbf{x})$ is a scalar field, and $\mathbf{u}(\mathbf{x}) = (u(\mathbf{x}), v(\mathbf{x}), w(\mathbf{x}))^{\mathrm{T}}$ is a vector field. (Don't forget the A^{T} notation allows us to write a column vector as a transposed row vector, to save space!) The properties translate as you'd expect in dimensions other than 3d (except curl which doesn't exist apart from in 3d!). The rules as to what is a valid input or output, are the same as for the scalar, dot, and cross product, so no extra learning of cases required.

Name	Example	Input	Output
Divergence (div)	$\nabla \cdot \mathbf{u}(\mathbf{x}) = \frac{\partial}{\partial x}u(\mathbf{x}) + \frac{\partial}{\partial y}v(\mathbf{x}) + \frac{\partial}{\partial z}w(\mathbf{x})$	Vector	Scalar
Gradient (grad)	$\nabla f(\mathbf{x}) = \mathbf{i}\frac{\partial}{\partial x}f(\mathbf{x}) + \mathbf{j}\frac{\partial}{\partial y}f(\mathbf{x}) + \mathbf{k}\frac{\partial}{\partial z}f(\mathbf{x})$	Scalar	Vector
Laplacian (Del-squared)	$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) f(\mathbf{x})$ $\nabla^2 \mathbf{u}(\mathbf{x}) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) \mathbf{u}(\mathbf{x})$	Scalar Vector	Scalar Vector
Curl (curl)	$\nabla \times \mathbf{u}(\mathbf{x}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u(\mathbf{x}) & v(\mathbf{x}) & w(\mathbf{x}) \end{vmatrix} = \begin{bmatrix} w_y - v_z \\ u_z - w_x \\ v_x - u_y \end{bmatrix}$	Vector	Vector

The divergence measures if in a vector field, the vectors nearby a point flow inwards or outwards towards the point.

The gradient measures for a scalar field, which direction is the steepest i.e. which direction changes the value of the function quickest, and by how much.

The Laplacian, as you will see, it comes up a lot in partial differential equations. It can be applied to scalars or vectors (∇^2 is a scalar operator).

The curl will measure by how much a small object placed in a vector field would rotate if pushed by those vectors.

There are plenty of vector calculus identities that exist, that we won't go into detail here. There's a comprehensive list on Wikipedia: Vector calculus identities, should you need to look them up.

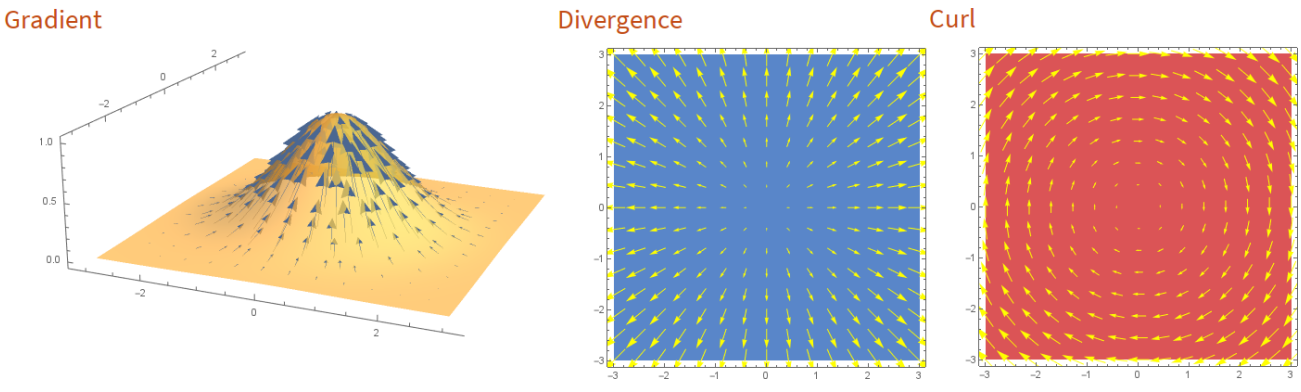


Figure 13.1: a) Gradient of a scalar field. b) Vector field with a constant divergence. c) Vector field with a constant curl.

Examples

Find the **divergence** of the vector function, $\begin{bmatrix} xyz \\ 3y^2x \\ x^2 + y^2 \end{bmatrix}$.

Let's start by writing out the div operator in full,

$$\begin{aligned}\nabla \cdot \begin{bmatrix} xyz \\ 3y^2x \\ x^2+y^2 \end{bmatrix} &= \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} \cdot \begin{bmatrix} xyz \\ 3y^2x \\ x^2+y^2 \end{bmatrix} \\ &= \frac{\partial}{\partial x}(xyz) + \frac{\partial}{\partial y}(3y^2x) + \frac{\partial}{\partial z}(x^2+y^2) \\ &= (yz) + (6yx) + (0) \\ &= y(z+6x)\ .\end{aligned}$$

Calculate the **gradient** of $xy - z^2$.

$$\begin{aligned}\nabla(xy-z^2) &= \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix} (xy-z^2) \\ &= \begin{bmatrix} y \\ x \\ -2z \end{bmatrix}\end{aligned}$$

Calculate the **Laplacian** of $\sin(xy)$.

$$\begin{aligned}\nabla^2 \sin(xy) &= \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\sin(xy) \\ &= -y^2\sin(xy) - x^2\sin(xy)\end{aligned}$$

Calculate the **Curl** of $y\mathbf{i} - x\mathbf{j}$.

$$\begin{aligned}\nabla \times \begin{bmatrix} y \\ -x \\ 0 \end{bmatrix} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y & -x & 0 \end{vmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ -1-1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}\end{aligned}$$

13.5.3 Gradient revisited

Previously we saw the gradient operator and how it takes a scalar field and returns a vector field.

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix}$$

We can generalise this to apply to vectors too. Since the gradient takes a scalar and *upgrades* it to a vector, taking the gradient of a vector field should upgrade it to a matrix,

$$\nabla \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{bmatrix}$$

This patern will continue giving larger objects each time. They become harder to write down because we run out of space, but the general form of scalar, vector, matrix is called a tensor (these are the same tensors that *flow* in the famous machine learning library TensorFlow).

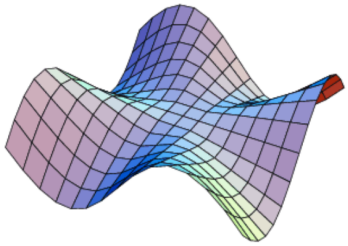
Applying the gradient once (to a scalar or a vector field etc.) has a special name, the *Jacobian*. $\mathbf{J} = \nabla f$. For a scalar function, this returns a vector - what if we apply it again? It should return a matrix, and it does!

$$\mathbf{H} = \nabla(\nabla f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix}$$

This one has the special name, the *Hessian*, do note that $\nabla(\nabla f) \neq \nabla^2 f$, as one is a matrix and the other a scalar, although the trace of the Hessian does equal the Laplacian! $\mathrm{Tr}(\nabla(\nabla f)) = \nabla^2 f$.

We'll revisit the Hessian and Jacobian when we look at optimisation later in the module.

Do note that computers are quite good at doing linear algebra (this is actually the primary feature of programs like matlab and numpy in python). What is important is that you get a feel for how the tools work so that you can interpret the meaning of results.



Chapter 14

Partial Differential Equations

14.1 Recap

As you will remember, an ordinary differential equation (ODE) is an equation with at least one ordinary derivative. The definition for partial differential equations is as expected, an equation with at least one partial derivative. Let's recap ODEs, and see what insights we can gain as we upgrade to PDEs. Take the following ODE,

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = x^2 \,,$$

We can find a *solution* to this equation, which is to say a function of x that satisfies the ODE. i.e.,

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

This simple expression represents a family of solutions to the above equation, parametrised by the constant c at the end. If we were given further information, we could find the particular solution. Say we know the value of the function at $x = 0$, we could replace c with a specific value,

$$f(x) = \frac{1}{3}x^3 + f(0) \,.$$

So far so good, what if we had the very similar PDE,

$$\frac{\partial f(x,y)}{\partial x} = x^2 \,,$$

Not a lot has changed, so the solution should look very much like the ODE case,

$$f(x,y) = \frac{1}{3}x^3 + c(y) \,.$$

The only difference is in the *'constant'* term; the only requirement is that is constant when varying x, that's to say nothing of varying y. In fact this term is upgraded to a *function* of y. This subtle change has implications for how we specify a particular solution. Previously we could lock down our solution by specifying information at specific points on the function, e.g. at $x = 0$, now we have to specify the function for all values of y.

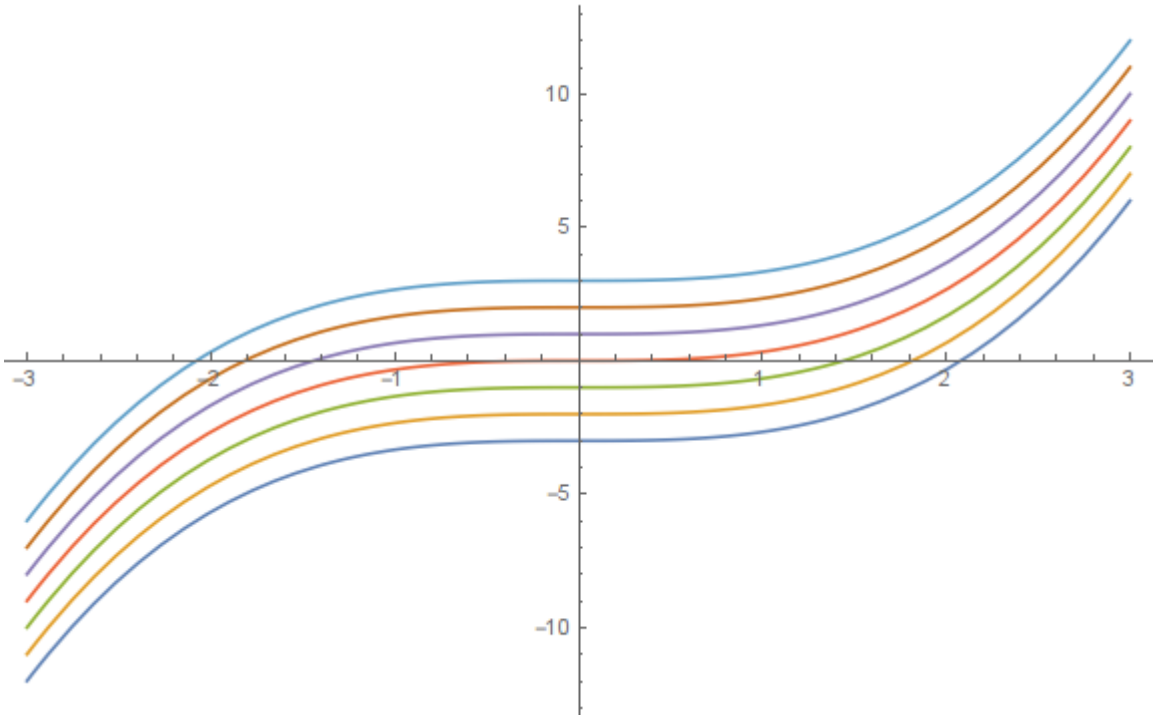


Figure 14.1: Solutions to the ODE $\frac{\mathrm{d}f(x)}{\mathrm{d}x} = x^2$.

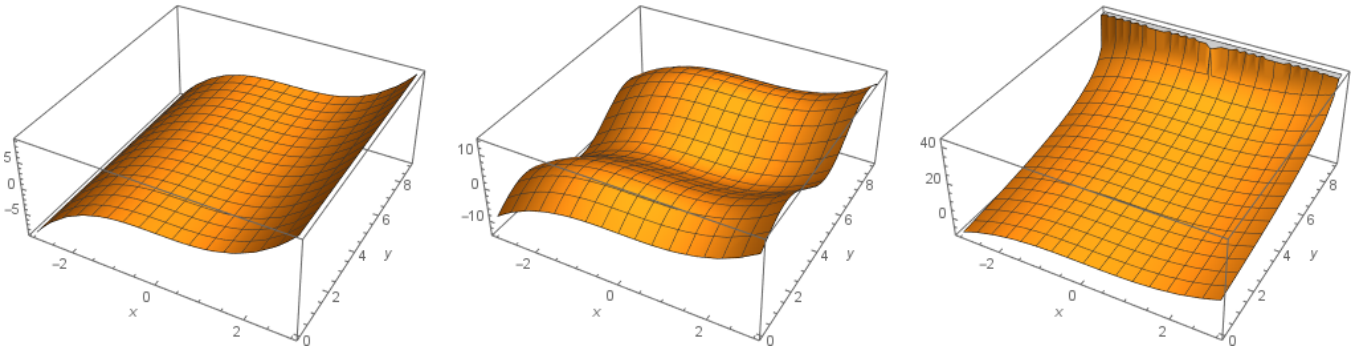


Figure 14.2: Solutions to the PDE $\frac{\partial f(x,y)}{\partial x} = x^2$.

Example: Find $c(y)$ for the particular solution of the above PDE where $f(x, y)$ takes the values $f(y) = e^y$ when $x = 0$.

$$\begin{aligned} f(x,y) &= \frac{1}{3}x^3 + c(y) \\ f(y) = f(0,y) &= \frac{1}{3}0^3 + c(y) \\ c(y) &= e^y \\ \Rightarrow f(x,y) &= \frac{1}{3}x^3 + e^y \end{aligned}$$

Example: Find $c(y)$ for the particular solution of the above PDE where $f(x, y)$ takes the values $f(y) = 2\sin(\pi y)$ along the curve $x = \sqrt{y}$.

This example is more difficult, it illustrates how we can specify boundary conditions over a

curve in x, y space, rather than a specific straight line.

$$\begin{aligned} f(x,y) &= \frac{1}{3}x^3 + c(y) \\ f(y) = f(\sqrt{y}, y) &= \frac{1}{3}(y^{1/2})^3 + c(y) \\ 2\sin(\pi y) &= \frac{1}{3}(y^{1/2})^3 + c(y) \\ c(y) &= 2\sin(\pi y) - \frac{1}{3}y^{3/2} \\ \Rightarrow f(x,y) &= \frac{1}{3}x^3 + 2\sin(\pi y) - \frac{1}{3}y^{3/2} \end{aligned}$$

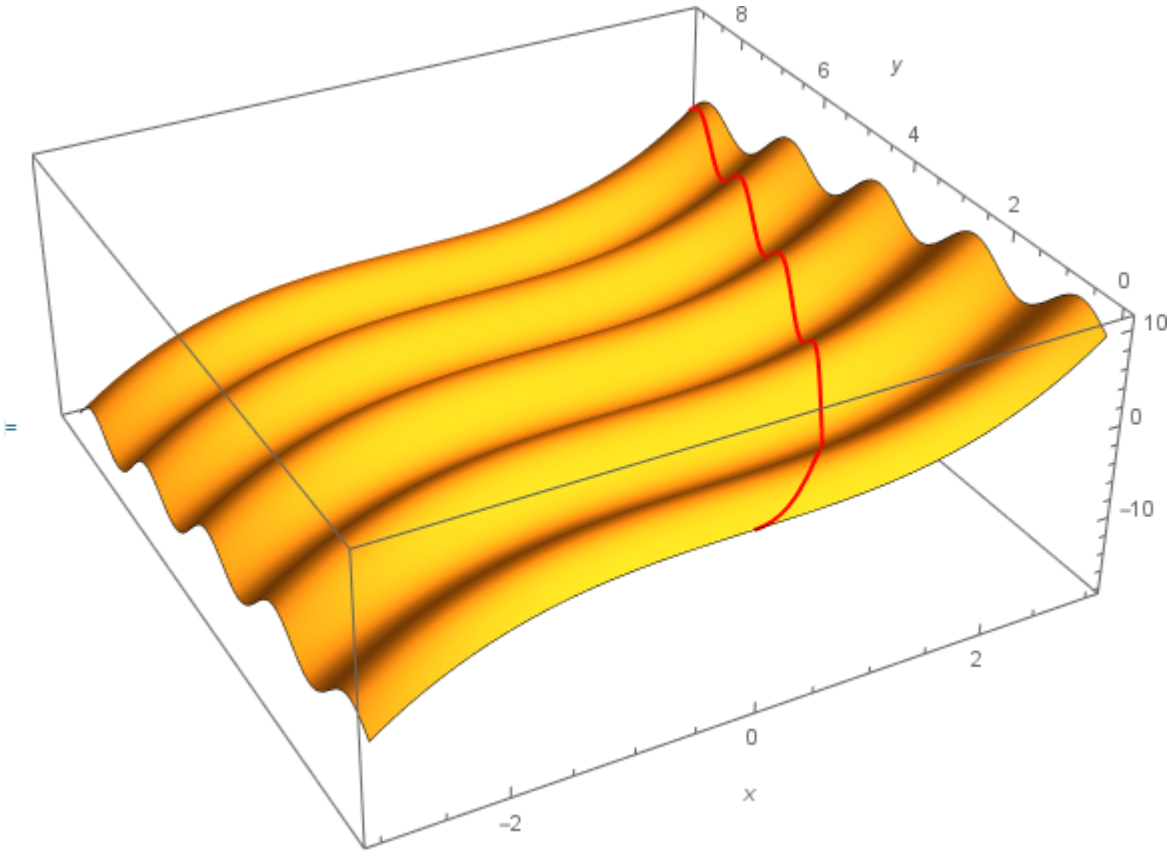


Figure 14.3: Particular solution to the PDE $\frac{\partial f(x,y)}{\partial x} = x^2$ with $f(x,y) = 2\sin(\pi y)$ along $x = \sqrt{y}$.

These examples have set the scope for more complicated partial differential equations. Our PDEs are equations in more than one dimension, whose general solutions are functions of multiple variables, and the particular solution to these is specified by a information given at curves within the space.

14.2 PDE strategies

Often for engineering purposes, the goal isn't to find a new solution to a differential equation, but to manipulate known solutions to the situation at hand. The PDE we saw last time was simplistic, it almost looked like the simplest ODE. This was to get a feel for what kind of objects

we're looking at. Now let's consider a more realistic PDE, one that has derivatives in more than one variable, i.e. the wave equation,

$$\frac{\partial^2 f(x,t)}{\partial t^2} = c^2 \frac{\partial^2 f(x,t)}{\partial x^2} \ .$$

In principle we don't know how to solve this yet. An effective technique can be to use a trial solution, and test whether it does indeed solve the equation, perhaps generating a further condition on when it is valid. Perhaps frustratingly, there's not often a good way of selecting trial solutions other than already knowing that they are likely to work. I might call this, in jest, the Wolfram Alpha approach.

For the wave equation, let's test the hypothesis that solutions take the form,

$$f(x,t) = f(x \mp ct) \ ,$$

which represents wavepackets with a shape $f(x)$ that moves to the right ($-$ sign), or to the left ($+$ sign), with a speed c. To do this, we must find the partial derivatives of the test solution and see if they match the PDE,

$$\begin{aligned}
\frac{\partial f(x \mp ct)}{\partial t} &= \mp c f'(x \mp ct) \\
\frac{\partial^2 f(x \mp ct)}{\partial t^2} &= c^2 f''(x \mp ct) \\
\frac{\partial f(x \mp ct)}{\partial x} &= f'(x \mp ct) \\
\frac{\partial^2 f(x \mp ct)}{\partial x^2} &= f''(x \mp ct)
\end{aligned}$$

Therefore, inserting these into the PDE,

$$c^2 f''(x \mp ct) = c^2 f''(x \mp ct)$$

which as the LHS and RHS are equal, is always true.

Since the wave equation is a *linear* PDE, then any sum of solutions is also a solution. Therefore, the general solution to the wave equation is,

$$f(x,t) = f_+(x - ct) + f_-(x + ct) \ .$$

For a particular solution, it is not enough just to know the value of the function at a boundary, i.e. at $t = 0$, since we would have one equation and two unknowns,

$$f(x,0) = f_+(x) + f_-(x),$$

If we had a second piece of information, such as the value of the function at another time, or the time derivative, then we could solve for f_+ and f_-. e.g.,

$$\left.\frac{\partial f(x,t)}{\partial t}\right|_{t=0} = -c f'_+(x) + c f'_-(x),$$

Therefore, since,

$$\frac{\partial f(x,0)}{\partial x} = f'_+(x) + f'_-(x)$$

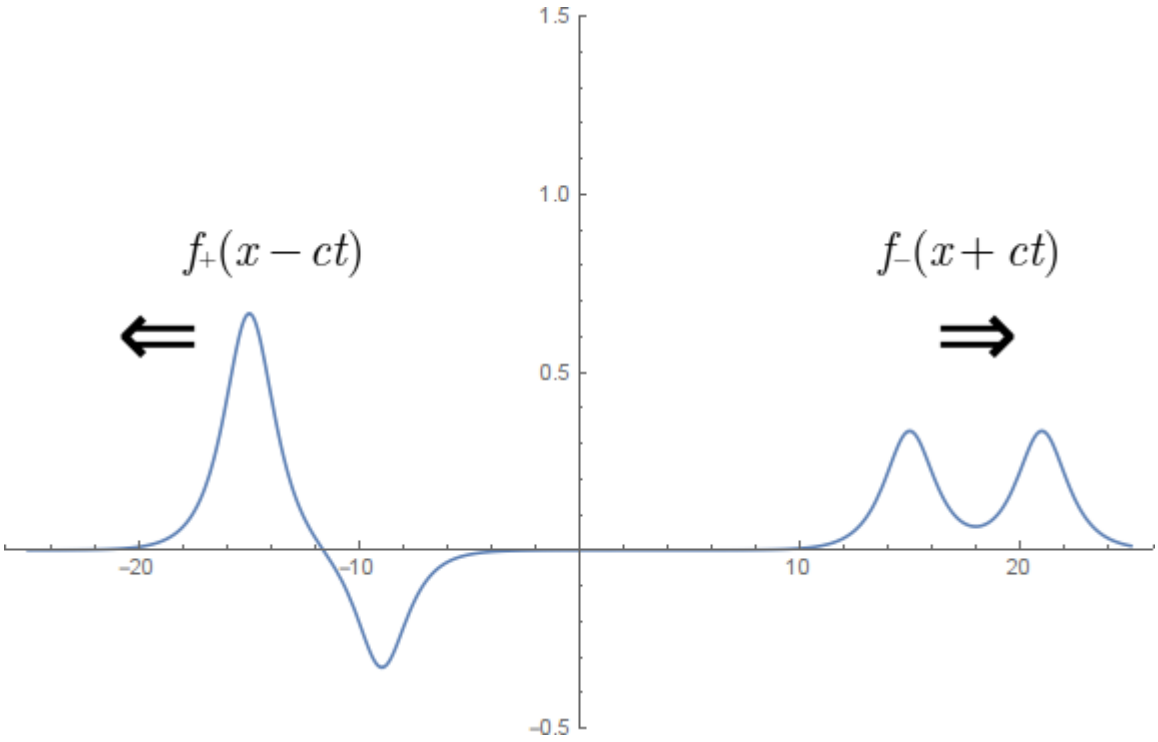


Figure 14.4: Left and right traveling wave solutions to the wave equation.

then,

$$f'_+(x) = \frac{1}{2}\frac{\partial f(x,0)}{\partial x} - \frac{1}{2c}\left.\frac{\partial f(x,t)}{\partial t}\right|_{t=0}$$
$$f'_-(x) = \frac{1}{2}\frac{\partial f(x,0)}{\partial x} + \frac{1}{2c}\left.\frac{\partial f(x,t)}{\partial t}\right|_{t=0},$$

Then $f_\mp$ can be found by integrating these.

14.2.1 Separation of variables

The previous example relied on us knowing something specific about the wave equation. We won't always have this knowledge, so let's explore a more general technique called *separation of variables*. This technique assumes that we can write a solution that is the product of functions of one variable. i.e.,

$$f(x,t) = X(x)T(t)\ .$$

Here you will note that $X(x)$ is only a function of x and $T(t)$ is only a function of t. If we plug this into the wave equation, we get,

$$X(x)T''(t) = c^2X''(x)T(t)\ .$$

Now, dividing through by $X(x)T(t)$, will give us,

$$\frac{T''(t)}{T(t)} = c^2\frac{X''(x)}{X(x)}\ .$$

Here we see that the LHS is only a function of t and the RHS is only a function of x. This implies that each must be constant, since it can't be a function of x, the LHS doesn't depend

on x, nor can it be a function of t since the RHS doesn't depend on t. This allows us to split the equation out into two ODEs,

$$\begin{aligned}\frac{T''(t)}{T(t)} &= c^2\frac{X''(x)}{X(x)} = -\omega^2\\ T''(t) &= -\omega^2 T(t)\\ X''(x) &= -\frac{\omega^2}{c^2}X(x)\ ,\end{aligned}$$

where $-\omega^2$ has been introduced here as the constant term. (I've chosen this form specifically because I know what comes next, but there's nothing stopping me saying the constant was just A for example). We are able to solve these ODEs, they give sinusoidal solutions,

$$\begin{aligned}T(t) &= A_+e^{i\omega t} + A_-e^{-i\omega t}\\ X(x) &= B_+e^{i\omega x/c} + B_-e^{-i\omega x/c}\end{aligned}$$

These are then combined into one equation (with some rearranging),

$$f(x,t) = C_+(\omega)e^{i\frac{\omega}{c}(x-ct)} + C_-(\omega)e^{i\frac{\omega}{c}(x+ct)}\ .$$

You'll notice how the $x \mp ct$ behaviour gets recovered here. The constants have been given an explicit ω dependency, this is to indicate that the general solution is a sum over all possible values for ω (which is any real number),

$$f(x,t) = \int_{-\infty}^{\infty}\mathrm{d}\omega\ C_+(\omega)e^{i\frac{\omega}{c}(x-ct)} + C_-(\omega)e^{i\frac{\omega}{c}(x+ct)}\ .$$

Dealing with expressions like this is the subject of *Fourier analysis*.

14.2.2 Example - Application to PDEs

You may wonder why we spend lots of time on the separation of variables technique in PDEs. The answer is, often (i.e. for the wave equation, Laplace' equation and diffusion equation) we separate out ODEs that permit sinusoidal solutions. These sinusoids can be combined into a Fourier series in one of the variables, which then can tell us how the function behaves in the other variables.

E.g. for the diffusion equation, $\frac{\partial f}{\partial t} - \alpha\frac{\partial^2 f}{\partial x^2} = 0$, we get ODEs,

$$\begin{aligned}X''(x) &= -k^2X(x)\\ T'(t) &= -\gamma T(t)\\ \text{with}\ \ \gamma &= \alpha k^2\ ,\end{aligned}$$

with solutions,

$$f(x,t) = ae^{-\alpha k^2 t}\cos(kx) + be^{-\alpha k^2 t}\sin(kx)\ .$$

Remember, that this is true for any arbitrary k. We could set $k = \frac{n\pi}{L}$, i.e.,

$$f(x,t) = ae^{-\frac{\alpha n^2\pi^2}{L^2}t}\cos\Big(\frac{n\pi x}{L}\Big) + be^{-\frac{\alpha n^2\pi^2}{L^2}t}\sin\Big(\frac{n\pi x}{L}\Big)\ .$$

Remember that sum of solutions to a linear PDE or ODE is also a solution, so,

$$g(x,t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n e^{-\frac{\alpha n^2 \pi^2}{L^2} t} \cos\left(\frac{n\pi x}{L}\right) + b_n e^{-\frac{\alpha n^2 \pi^2}{L^2} t} \sin\left(\frac{n\pi x}{L}\right) \right) ,$$

is a solution too.

Now, here's where it gets interesting. If we set $t = 0$, then this expression turns into the Fourier series.

$$g(x,0) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right) ,$$

So, if we work out the Fourier coefficients a_n, b_n for an initial condition, $f(x, t = 0)$, then we can know how that function evolves in time. By grouping together the fourier coefficients at the exponential, we can get time-varying coefficients,

$$a_n(t) = a_n e^{-\frac{\alpha n^2 \pi^2}{L^2} t}$$
$$b_n(t) = b_n e^{-\frac{\alpha n^2 \pi^2}{L^2} t}$$

which allow us to reconstruct a new Fourier series at any moment in time, just by knowing the $t = 0$ state.

Let's apply this to the previous square wave solution,

$$g(x,0) = \frac{4}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{1}{n} \sin(nx)$$

therefore, by pattern matching to the general separable solution,

$$g(x,t) = \frac{4}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{1}{n} \sin(nx) e^{-\alpha n^2 t}$$

is the full particular solution, that solves the diffusion equation. It looks like this:

Diffusion equation

Let's apply our technique to another useful PDE, the diffusion equation,

$$\frac{\partial f(x,t)}{\partial t} = \alpha \frac{\partial^2 f(x,t)}{\partial x^2} .$$

This equation models how a localised concentration of a quantity spreads out, or diffuses, over time. The constant α is the diffusivity, which measures how readily a concentration diffuses away. A mental image of this could be how a blob of honey might spread out if you dropped it on a table (assuming no surface tension). The expression is found in an engineering context to model how concentrations of ions, or gasses pass from an area of high concentration to an area of low concentration, or indeed how heat is conducted in a solid object (the diffusion equation also gets called the heat equation).

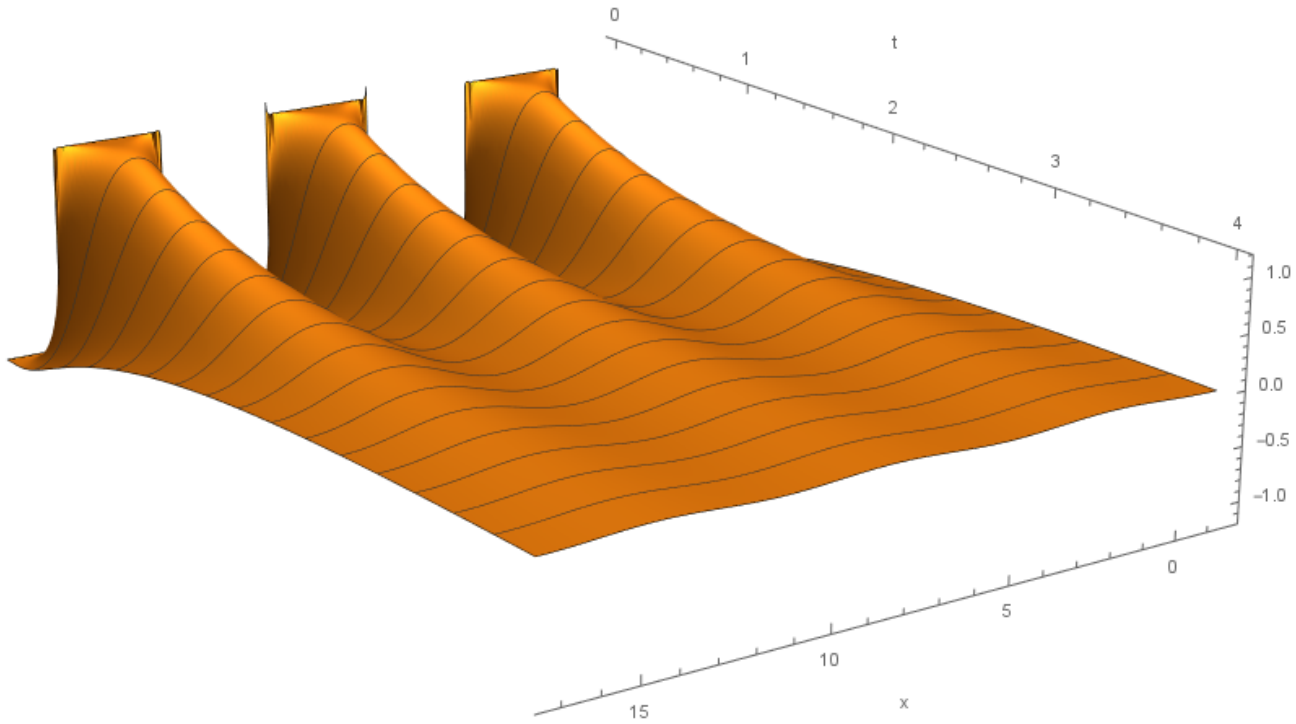


Figure 14.5: A square wave evolving in time under the diffusion equation, calculated using Fourier series.

Let's attempt to gain insights about the diffusion equation by applying separation of variables. Again, let $f(x,t) = X(x)T(t)$, then,

$$\begin{aligned}\frac{\partial f(x,t)}{\partial t} &= \alpha \frac{\partial^2 f(x,t)}{\partial x^2} \\ X(x)T'(t) &= \alpha X''(x)T(t) \\ \frac{T'(t)}{T(t)} &= \alpha \frac{X''(x)}{X(x)} = -\gamma \ .\end{aligned}$$

Giving us two ODEs, with solutions,

$$\begin{aligned}T'(t) &= -\gamma T(t) \\ X''(x) &= -\frac{\gamma}{\alpha} X(x) \\ T(t) &= Ae^{-\gamma t} \\ X(x) &= B \sin\left(\sqrt{\frac{\gamma}{\alpha}} x\right) + C \cos\left(\sqrt{\frac{\gamma}{\alpha}} x\right) \\ \Rightarrow f(x,t) &= B' \sin\left(\sqrt{\frac{\gamma}{\alpha}} x\right) e^{-\gamma t} + C' \cos\left(\sqrt{\frac{\gamma}{\alpha}} x\right) e^{-\gamma t}\end{aligned}$$

What we see here is that fine features that vary spatially over short distances tend to die out quite quickly - they have a large γ parameter in the sin and cos, but equally that large parameter is in the $e^{-\gamma t}$, which means this feature decays fast. Conversely, coarse features that vary over longer distances tend to remain for longer.

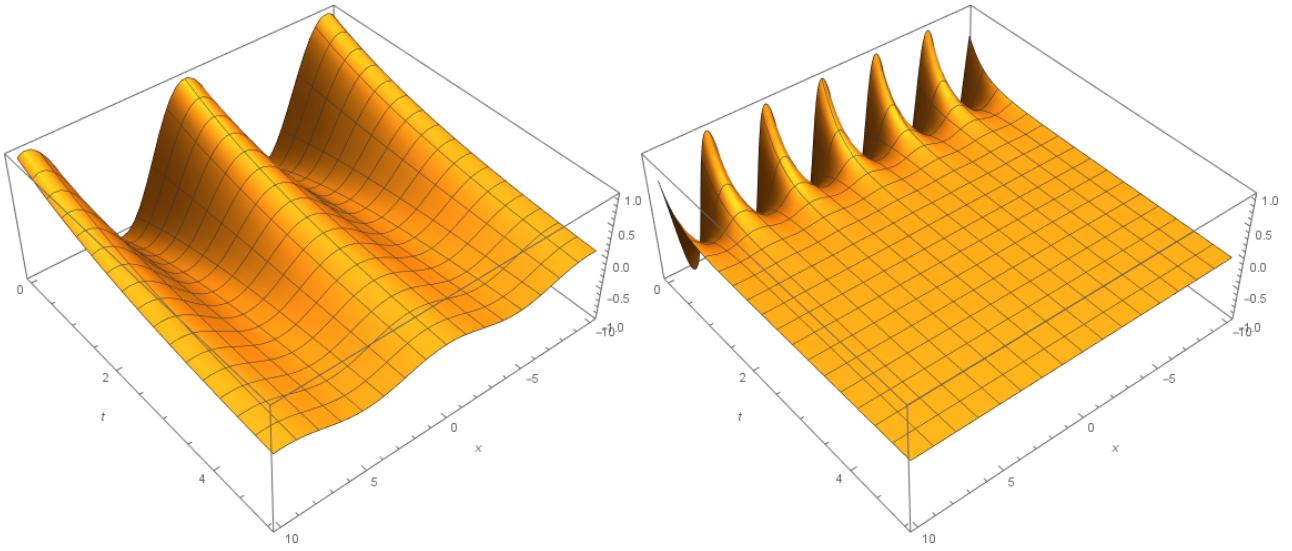


Figure 14.6: Sinusoidal solutions to the diffusion equation.

14.2.3 Fundamental solution

Let's look at the diffusion equation from a different angle, to see if we can gain any more insight. If we assume a concentration as having a Gaussian profile at an initial time,

$$f(x,0) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \ ,$$

where here σ is the standard deviation (or characteristic width) of the concentration, it would be reasonable to assume the width gets bigger over time. We could guess that the profile stays Gaussian over time too - this is a hypothesis, we'll need to confirm that it is indeed true.

First we need a general piece of information from the diffusion equation, namely, does the area of the concentration curve remain constant for all times? To answer this, we can integrate the diffusion equation itself,

$$\int_{-\infty}^{\infty} \mathrm{d}x \ \frac{\partial f(x,t)}{\partial t} = \int_{-\infty}^{\infty} \mathrm{d}x \ \alpha \frac{\partial^2 f(x,t)}{\partial x^2}$$

On the left hand side, we can reverse the order of differentiation and integration, and on the right, we can directly integrate,

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} \mathrm{d}x \ f(x,t) = \alpha \left.\frac{\partial f(x,t)}{\partial x}\right|_{x\to\infty} - \alpha \left.\frac{\partial f(x,t)}{\partial x}\right|_{x\to-\infty} \ ,$$

if we assume the concentration goes to zero sufficiently fast towards infinity, then the derivatives towards infinity will also go to zero, giving,

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} \mathrm{d}x \ f(x,t) = 0 \ .$$

This confirms that the area underneath the concentration curve does indeed remain constant over time. This area would represent things like the total number of particles (molecules, ions, etc.); this remaining constant seems reasonable.

With this in mind, let's rewrite our Gaussian to have a constant unit area,

$$f(x,0) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \ .$$

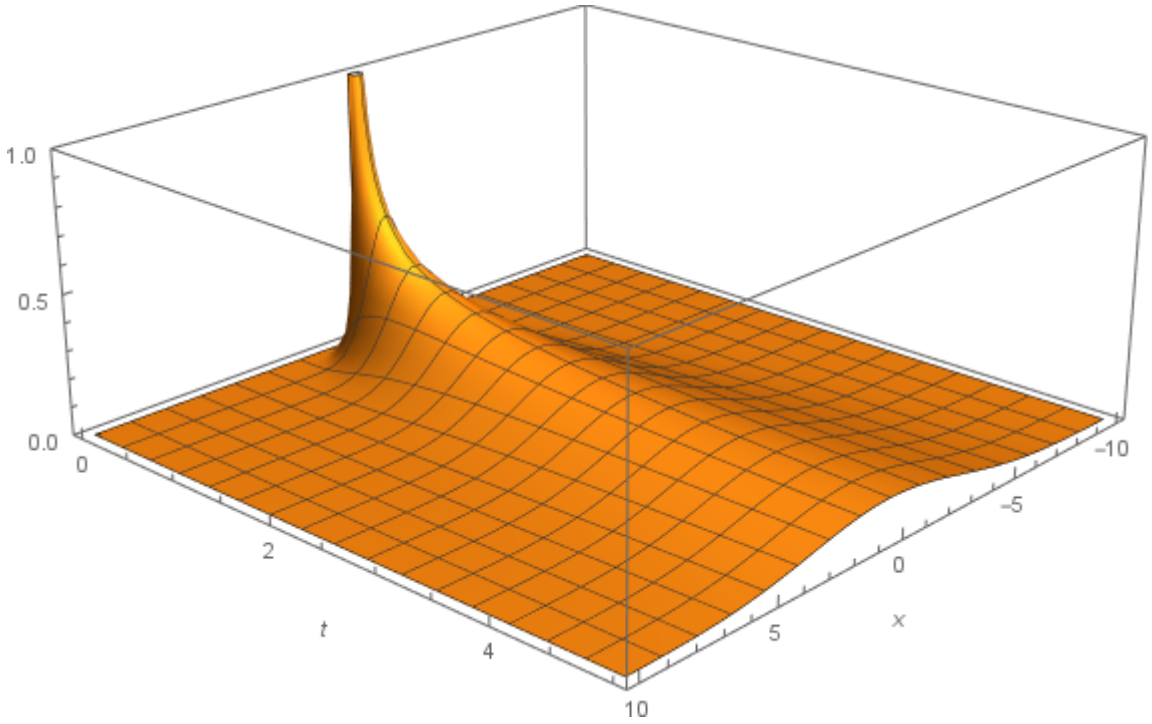


Figure 14.7: Gaussian solution to the diffusion equation.

Let's follow our hypothesis that the width changes as a function of time - although as an unknown function for now,

$$f(x,t) = \frac{1}{\sigma(t)\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma(t)^2}\right) .$$

We'll need to plug this into the PDE, so let's first calculate the relevant partial derivatives:

$$\begin{aligned}
\frac{\partial f(x,t)}{\partial t} &= \frac{1}{\sqrt{2\pi}}\left[-\frac{1}{\sigma(t)^2} + \frac{x^2}{\sigma(t)^4}\right]\sigma'(t)\exp\left(-\frac{x^2}{2\sigma(t)^2}\right) \\
\frac{\partial f(x,t)}{\partial x} &= \frac{1}{\sqrt{2\pi}}\left[-\frac{x}{\sigma(t)^3}\right]\exp\left(-\frac{x^2}{2\sigma(t)^2}\right) \\
\frac{\partial^2 f(x,t)}{\partial x^2} &= \frac{1}{\sqrt{2\pi}}\left[-\frac{1}{\sigma(t)^3} + \frac{x^2}{\sigma(t)^5}\right]\exp\left(-\frac{x^2}{2\sigma(t)^2}\right)
\end{aligned}$$

Inserting into the PDE,

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}}\left[-\frac{1}{\sigma(t)^2} + \frac{x^2}{\sigma(t)^4}\right]\sigma'(t)\exp\left(-\frac{x^2}{2\sigma(t)^2}\right) &= \alpha\frac{1}{\sqrt{2\pi}}\left[-\frac{1}{\sigma(t)^3} + \frac{x^2}{\sigma(t)^5}\right]\exp\left(-\frac{x^2}{2\sigma(t)^2}\right) \\
\Rightarrow \sigma'(t) &= \frac{\alpha}{\sigma(t)}
\end{aligned}$$

Now, this is a non-linear ODE, but it can be solved fairly easily.

$$\begin{aligned}
\frac{\mathrm{d}\sigma(t)}{\mathrm{d}t}\sigma(t) &= \alpha \\
\int \mathrm{d}\sigma(t)\ \sigma(t) &= \int \mathrm{d}t\ \alpha \\
\frac{\sigma(t)^2}{2} &= \alpha t + c \\
\sigma(t) &= \sqrt{2\alpha t + 2c}
\end{aligned}$$

or by fixing the constant term,

$$\sigma(t) = \sqrt{\sigma(0)^2 + 2\alpha t}$$

Let's pause and have a look at what we've just uncovered. Our guess that the concentration shape would remain Gaussian as it diffuses was correct, and the condition for this to be true, is the function form of $\sigma(t)$ just derived. Our derived solution is therefore,

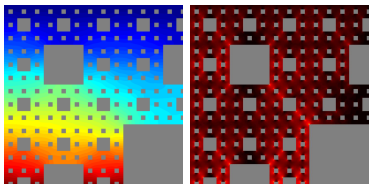
$$f(x,t) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma(0)^2 + 2\alpha t}} \exp\left(-\frac{x^2}{2\sigma(0)^2 + 4\alpha t}\right) .$$

The width of the concentration expands proportionally to the square root of the diffusivity times time, i.e. relatively slowly, and slowing as it widens out. Note that this expression gives imaginary widths for times $t < -\sigma(0)/(2\alpha)$, which is nonsense. Therefore, this solution has a bounded domain, where it only predicts behaviour in the range $t \in \left(-\frac{\sigma(0)}{2\alpha}, \infty\right)$. In the limit as the width goes to zero, we have a solution with finite area (i.e. particle number) yet localised exactly to a single location at a single instant in time, and spreads out from there.

$$f(x,t) = \frac{1}{\sqrt{4\pi\alpha t}} \exp\left(-\frac{x^2}{4\alpha t}\right) .$$

This form is of particular use to construct the time-evolution of an arbitrary starting concentration, earning it the rather grandiose name of the fundamental solution.

In this chapter we have looked at partial differential equations, seeing how they generalise from ODEs and that they need to be specified over curves rather than single points for a particular solution. We've explored some ways of finding solutions to PDEs, either by constructing them out of known pieces, or techniques to reduce the PDEs to a set of related ODEs. What we've not covered is non-linear PDEs (a whole course on it's own) or inhomogeneous PDEs, where there is a driving term for our waves, or heat/particle sources and sinks in our diffusion equation. Often these PDEs can be solved to required precision numerically, and we'll explore this later in the module. Quite often, we have equations that are in more spatial dimensions, and where the parameters (like wave speed and diffusivity) are able to change as a function of space. In these cases, stitching together solutions from simpler cases, or numerically solving are often the only way of tackling, but the intuition build here should give you insight when faced with that task.



Chapter 15

Finite Differences

15.1 Introduction

You will have seen that in some simple scenarios it is possible to find analytical solutions to differential equations; however, even very minor increases in complexity usually make this approach impractical, if not impossible!

Finite difference methods are a numerical approach to solving differential equations by approximating derivatives with "difference quotients". First, the simulation domain (typically time and/or space) is discretised (chopped up into chunks), where the properties of the system are stored at discrete nodes. Then a Taylor series expansion is used to describe the relation of each node to the others in the system. This method is exclusively used with the help of computers, as it involves many simple steps being repeated zillions of times. However, if the equation for every node really did include all the others in the system, this would be too much even for computers, so a truncated series is used, which typically only involves each node's nearest neighbours.

15.1.1 Taylor series again

As we learnt previously, the Taylor series assumes complete knowledge of a function at a single point (the value and all its derivatives) and uses this information to recreate the whole function with a power series. For all smooth, continuous functions, there will exist an exact polynomial description if the series is expanded to enough terms. In some cases, such as trigonometric functions, infinitely many terms are required. If we assume that we know every thing about the function $f(x)$ at the point c, then we can approximate $f(x)$ at any other point by using:

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \frac{f^{(3)}(c)}{3!}(x-c)^3 + ... + \frac{f^{(n)}(c)}{n!}(x-c)^n$$

Next, we make the substitution $x = c + \Delta x$ (we use the notation Δx to imply a very small, but non-zero step size. The motivation for this will become clear after the next step...)

$$f(c+\Delta x) = f(c) + f'(c)(\Delta x) + \frac{f''(c)}{2!}(\Delta x)^2 + \frac{f^{(3)}(c)}{3!}(\Delta x)^3 + ... + \frac{f^{(n)}(c)}{n!}(\Delta x)^n$$

our expression now says that if we know everything about $f(x)$ at point c, we can also find $f(x)$ at some point Δx away from c.

Many mathematical descriptions of physical phenomena, such as diffusion, involve the relation of differentials. The expression in the previous equation can be rearranged to make the first differential the subject, as shown in eq. 15.1 (be sure to have a go at this rearrangement yourself!).

$$f'(c) = \frac{f(c+\Delta x) - f(c)}{\Delta x} - \left[\frac{\Delta x}{2} f''(c) + \frac{\Delta x^2}{6} f'''(c) + \frac{\Delta x^3}{24} f^{(4)}(c) + ...\right] \tag{15.1}$$

So we now have an equation for the first derivative at position c in terms of the values of the function at $f(c)$ and $f(c+\Delta x)$, as well as a bunch of more complicated derivative terms (which have been put into square brackets...). What we can now say is that *if* Δx is very small, then all the terms inside the big square brackets must be *really small* compared to the first terms.

In fact, we will intentionally leave out the square bracket terms and use only the first part of the expression to form our approximation. We can now say that by truncating the Taylor series expansion before the second derivative term, we will expect to get an error on the order of Δx, which is fine if Δx is small enough! We often write this error as $e = \mathcal{O}(\Delta x)$, where the symbol $\mathcal{O}$ means "on the order of". So the approximation of the function $f(x)$ at point c becomes

$$f'(c) \approx \frac{f(c+\Delta x) - f(c)}{\Delta x} \tag{15.2}$$

This formulation is referred to as the *Forward Euler* approach and was first described in 1768 (way before computers!) by Leonhard Euler. If you think back to your study of linear functions for graph plotting, the equation above looks suspiciously like "gradient=rise/run", which I hope does not surprise you too much! In fact, this *linearisation* is exactly what a first order finite differencing scheme does (*i.e.* it takes any function and describes it as loads of tiny line segments - the smaller these lines are, the better the approximation will be!).

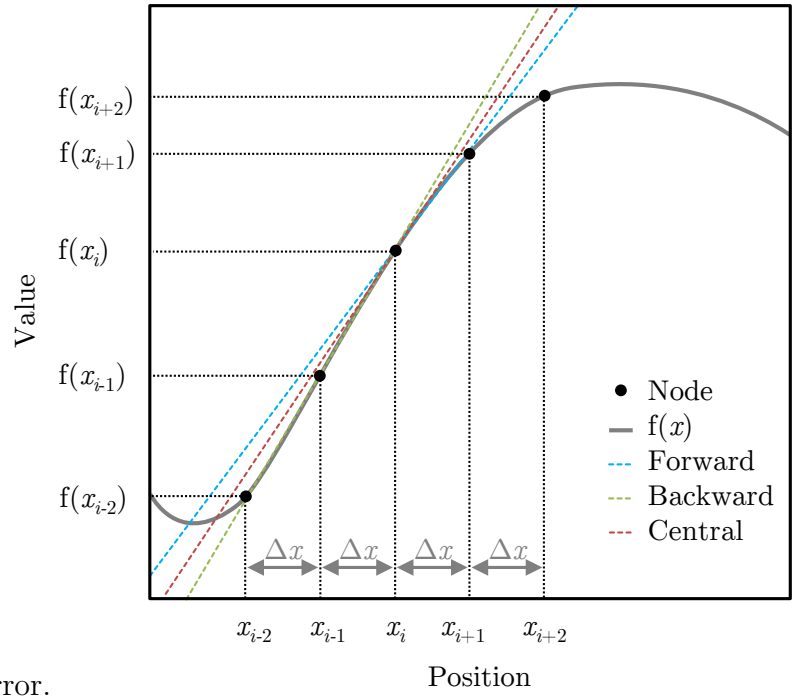
The *Backward Euler* can be similarly constructed by stepping Δx away from c in the negative direction (eq. 15.3) and has the same magnitude of error.

$$f'(c) = \frac{f(c) - f(c-\Delta x)}{\Delta x} + \left[\frac{\Delta x}{2} f''(c) - \frac{\Delta x^2}{6} f'''(c) + \frac{\Delta x^3}{24} f^{(4)}(c) - ...\right] \tag{15.3}$$

An improved approximation can be found by taking the average of the *Forward Euler* and *Backward Euler*. Equation 15.4 shows the resulting expression, referred to as the *central difference*, which now has an improved $\mathcal{O}(\Delta x^2)$ truncation error.

$$f'(c) = \frac{f(c+\Delta x) - f(c-\Delta x)}{2\Delta x} - \left[\frac{\Delta x^2}{6} f'''(c) + \frac{\Delta x^4}{120} f^{(5)}(c) + ...\right] \tag{15.4}$$

These three methods are shown graphically in the adjacent plot, where the difference between the outputs has been highlighted by extending their tangent approximations for the point x_i.



Similarly, we can also construct an approximation for the second derivative by taking the difference between the *Forward Euler* and *Backward Euler*. This should also not come as a surprise when we consider that the second derivative is just the gradient of the gradient!

As with the central difference, this approximation to the second derivative also has an $\mathcal{O}(\Delta x^2)$ truncation error.

$$f''(c) = \frac{f(c+\Delta x) - 2f(c) + f(c-\Delta x)}{\Delta x^2} - \left(\frac{\Delta x^2}{12} f^{(4)}(c) + ...\right) \tag{15.5}$$

15.2 Application example - Numerical diffusion

Now that we have suitable approximations to the first and second derivatives of a function, we can use them to approximate the solution to a useful equation. Diffusive processes are very common in engineering and, in 1 dimension, are described by the equation

$$\frac{\partial C(t,x)}{\partial t} = D\frac{\partial^2 C(t,x)}{\partial x^2}$$

where C represents the concentration of a diffusion species, t is time, x is distance and D is some kind of diffusivity coefficient (i.e. how easy is it for this thing to move around). The ∂ symbol is used to signify that a partial derivative is being evaluated as C is a function of both t and x. Firstly, using the forward difference approximation in eq. 15.1, we can approximate the time derivative to be

$$\frac{\partial C(t,x)}{\partial t} \approx \frac{C(t+\Delta t,x) - C(t,x)}{\Delta t}$$

Then, using the second derivative approximation in eq. 15.5 we can approximate the second spatial derivative

$$\frac{\partial^2 C(t,x)}{\partial x^2} \approx \frac{C(t,x+\Delta x) - 2C(t,x) + C(t,x-\Delta x)}{\Delta x^2}$$

Substituting these two approximations back into the diffusion equation, we get

$$\frac{C(t+\Delta t,x)-C(t,x)}{\Delta t} \approx D\frac{C(t,x+\Delta x)-2C(t,x)+C(t,x-\Delta x)}{\Delta x^2}$$

Looking at the expression above, we can assume that we know all the values of C at time t (i.e. now) and are therefore using this expression iteratively to work out the values at time $t+\Delta t$ (i.e. the next point in the future). Therefore, we can rearrange the equation above to make our unknown the subject

$$C(t+\Delta t,x) \approx D\frac{\Delta t}{\Delta x^2}\left[C(t,x+\Delta x)-2C(t,x)+C(t,x-\Delta x)\right]+C(t,x)$$

Finally, to make this more convenient for a computer to calculate many times, we will first make the substitution $\sigma = D\frac{\Delta t}{\Delta x^2}$ and then expand the brackets such that each value of C is only called once.

$$C(t+\Delta t,x) \approx \sigma C(t,x+\Delta x)+\sigma C(t,x-\Delta x)+(1-2\sigma)C(t,x)$$

We now have an expression ready to be input into a simulation which models time dependant diffusion. This approach is referred to as a *Forward-Time Central-Space* (FTCS) model because of how the approximations were derived. There are many subtle tweaks that we can implement in order to speed up or improve the accuracy of this simulation, but many real world codes today would simply use the expression we've found above. Interestingly, the FTSC method becomes susceptible to instability and oscillation if $\sigma > 0.5$, so we must be careful to avoid this. Detailed explanation of why this is is beyond the scope of this course, but use the MatLab code shared below to see for yourself!

15.3 Systems of equations and conditions

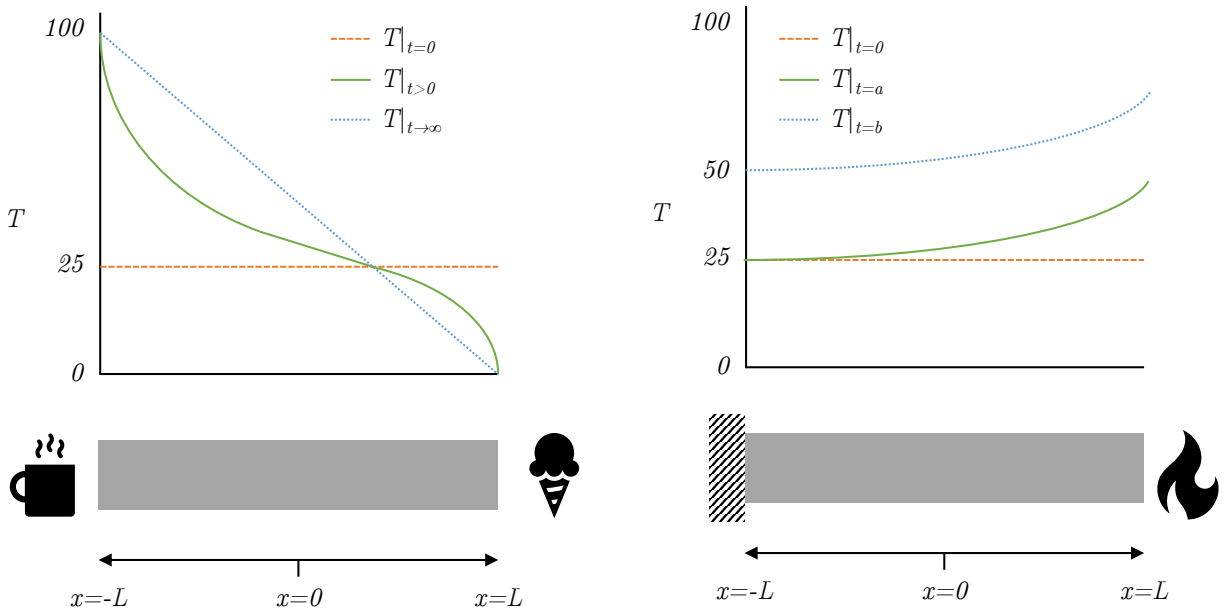
To describe a system that we wish to simulate, it's not enough just to give the governing equation (e.g. the diffusion equation of the wave equation). In addition, you will need to know things about what happens at the edges of the system (e.g. is there an insulator blocking heat transfer or maybe a blow torch adding more heat!), as well as the state of the system at some point in time (usually the initial condition). It is often convenient to write all of this information in a little cluster using the following format which we call a system of equations. A solution is something that satisfies all of these equations at the same time.

$$\begin{cases} \partial_t T = \alpha \partial_{xx} T & \text{on } (0,\infty)\times[-L,L], \\ T|_{t=0} = 25 & \forall x, \\ T|_{x=-L} = 100 & \forall t>0, \\ T|_{x=L} = 0 & \forall t>0. \end{cases} \tag{15.6}$$

The system above describes a one dimensional heat diffusion problem, with initial and boundary conditions. The first line is called the governing equation, which in this case is diffusion equation applied to the temperature, T. The parameter α is just a coefficient mediating the process, which in this case can be interpreted as the thermal diffusivity. The text to the right of this equation tells us where/when this equation applies, which in this case is all time from now, $0 < t < \infty$, and a region of space $2L$ wide, $-L \leq x \leq L$ (notice I've used the same rule for bracket selection as described in Chapter 1).

The second line of the equation contains two new symbols: a vertical line symbol "|" which can be read as *such that*, or just *at*; and an upside-down capital A symbol "∀" which should be read as *for all*. So the line reads "The temperature at time equals zero is equal to 25 for all x", i.e. initially the temperature is 25 everywhere.

Following the same logic, the third line reads "The temperature at position $x = -L$ is equal to 100 for all time greater than zero" i.e. the temperature on the left hand side of the system equals 100 from now on. Similarly, the final line of the system says "The temperature at position $x = L$ is equal to zero for all time greater than zero" i.e. the temperature on the right hand side of the system equals 0 from now on. So, although the system is initially room temperature everywhere, as soon as the clock begins, the temperatures at the two edges snap to new values, as if one end is touching boiling water and the other end is touching an ice cube.



This scenario is represented in the figure above, which shows what the distribution of temperature is initially, some time later, and eventually converging to a steady state scenario with a constant temperature gradient across the sample from hot to cold. Fixing the value of a system at any particular location is called a Dirichlet boundary condition; however, we could equally well fix the gradient instead (called a Neumann boundary condition), as in the following example system.

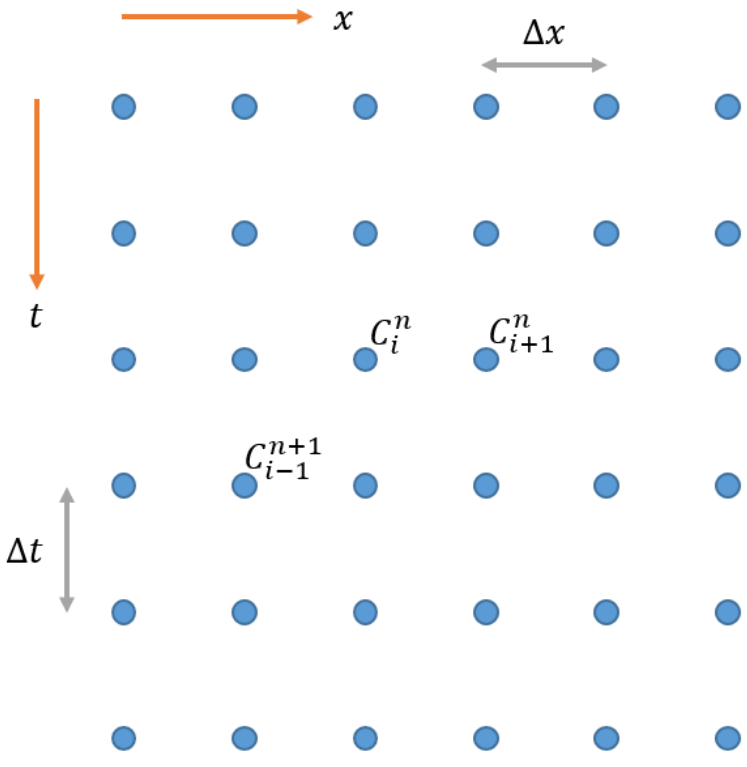
$$\begin{cases} \partial_t T = \alpha \partial_{xx} T & \text{on } (0,\infty)\times[-L,L], \\ T|_{t=0} = 25 & \forall x, \\ \partial_x T|_{x=-L} = 0 & \forall t > 0, \\ \partial_x T|_{x=-L} = 10 & \forall t > 0. \end{cases} \tag{15.7}$$

This time the left hand side of the system has a fixed gradient of zero (which we can think of as an insulating barrier, letting no heat in or out) and the right hand side has a constant gradient of 10 degrees per meter (we can think of this as a controlled heat source, like a laser). The figure shows the distribution of temperature initially, at time $t = a$ which is when $T|_{x=-L}$ first increases about 25 degrees, and at time $t = b$ which is when $T|_{x=-L} = 50$. Notice that the gradient of the temperature at $x = -L$ and $x = L$ are the same at $t = b$ as they are at $t = a$. These gradients are defined in the system of equations. Since this system is gaining heat at one end and not losing any heat at the other, it will just keep getting hotter and hotter, although the shape of the temperature profile will stay the same for $t \geq a$.

15.4 Notation

The functional notation used above can be a bit of a hassle to write, as you probably noticed in the above. In some cases it can be convenient to instead use subscripts and superscripts to communicate the value of t and x, either in the continuous form, or using indexes to refer to locations in the discrete form.

Continuous form: $C(t+\Delta t, x+\Delta x) \equiv C^{t+\Delta t}_{x+\Delta x}$ Discrete form: $C(n+1, i+1) \equiv C^{n+1}_{i+1}$



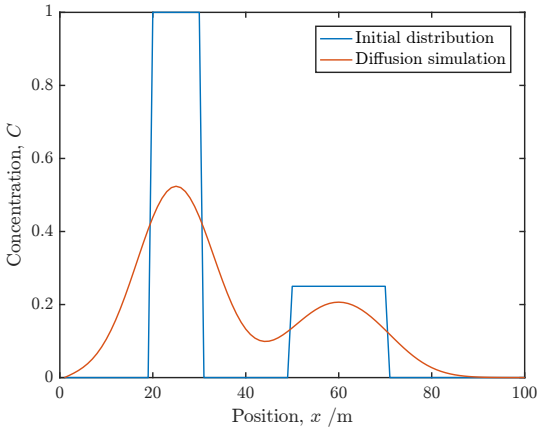
15.5 Code

```
%% Begin Function
% 1 Dimensional diffusion
D=1; % Define D [m^2/s]
delta_t=0.1; % Define the time step [s]
delta_x=1; % Define the spatial step [m]
sigma=D*delta_t/delta_x^2 %Calculate sigma
steps_t=300; % Define number of time steps
steps_x=100; % Define number of spatial nodes
C=zeros(steps_t,steps_x); % Define concentration matrix

% Initalise concentration at t=0
C(1,20:30)=1;
C(1,50:70)=0.25;

for t=1:steps_t-1 % Iterate through time steps
        % Use finite difference to calcuate concentration at next time step
        % Assume spatial end values are always zero
        C(t+1,2:end-1)=sigma*C(t,1:end-2)+sigma*C(t,3:end)+(1-2*sigma)*C(t,2:end-1);
        % Plot results
        plot(1:steps_x,C(1,:),1:steps_x,C(t+1,:));
        drawnow
end
legend('Initial distribution','Diffusion simulation');
xlabel('Position, x /m');
ylabel('Concentration, C');
%% End Function
```

Copy and paste the code above into a Matlab file (save as `Diffusion1D.m`). When you run this script, it should generate a plot after each time step similar to that shown below. When reading it, remember that all the green text after the `%` symbols are called comments and are included just to help you understand the code (*i.e.*, they are totally ignored by the computer).



This simulation could equally represent heat transfer, mass transport or even the movement of bacteria. Notice the smoothing effect that diffusion has, turning an initially very sharp distribution in to two overlapping curves similar to Gaussians.

You should work through the code to make sure you understand each line and then try modifying it. Perhaps start by changing the initial distribution to something more unusual. Next change the value of the node at $x = 0$ to something other than zero to observe the effect (*i.e.*, `C(:,1)=1;`). What does this do?

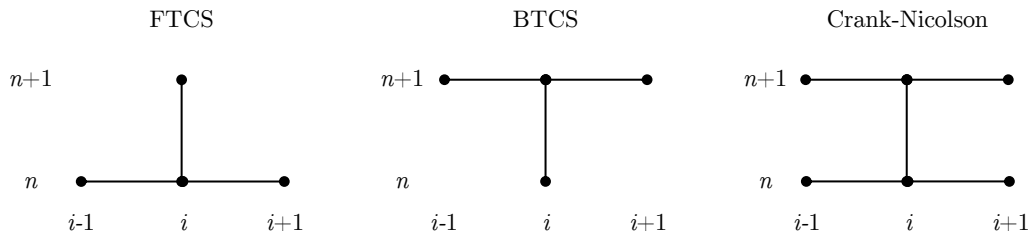
The spatial nodes at either end of our simulation are currently not being updated as we

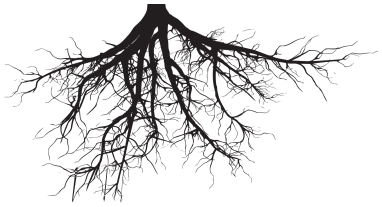
iterate. This means that their values don't change, meaning that they act like a *sink* and the mass/heat/bacteria will just flow through them as if they were falling off the edge of a table. However, clearly this is only appropriate for a certain set of scenarios. What happens at these boundaries is what we refer to as *boundary conditions*. Our current fixed value boundaries are referred to as *Dirichlet* boundaries, but several other options exist.

15.5.1 Alternative approaches

An alternative approach, called *Backward-Time Central-Space* (BTCS), can be derived in the same manner as the FTCS and has the same associated truncation error, but it does not yield an explicit solution. Instead an implicit approach must be used where the values of concentration within each iteration are calculated simultaneously by solving a system of linear equations. This matrix operation typically incurs more computational expense per iteration than FTCS; however, the BTCS approach is unconditionally stable and immune to oscillation for any value of σ.

The figure below illustrates the nodes required (or *stencil*) to update each value of C in an iteration for the FTCS and BTCS methods, as well as a third (particularly awesome) scheme, called Crank Nicolson, which you'll have to google.





Chapter 16

Root Finding

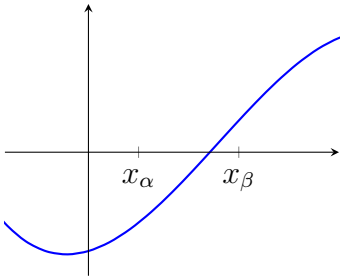
As discussed earlier in the course, it is sometimes possible to find the roots of equations simply by factoring them into linear factors or using known analytical solutions (such as the quadratic formula). However, in many cases there is no direct approach to find the roots and a numerical method must be used.

A variety of methods have been developed over the years, which vary in simplicity and efficiency. This chapter will explore the bisection method and the Newton-Raphson method.

16.1 The Bisection Method

If a function has a single root between the two x-coordinates, x_α and x_β, then $f(x_\alpha)$ and $f(x_\beta)$ should have different signs (*i.e.*, one positive and one negative).

The bisection method starts with the user specifying an interval (*i.e.*, two x-coordinates), which they believe to contain a single root. Next, this interval is bisected (*i.e.*, cut in half) to create two smaller intervals either side of our bisection point, x_1. By evaluation $f(x_1)$, we can then determine which of the two new intervals must contain the root (*i.e.*, the one with the sign change).



If we are lucky with the selection of our initial interval, our bisection point may eventually coincide with the root itself ($f(x_n) = 0$), but this will not usually be the case. As such, we must continue to iterate the bisection method until the interval containing the root is acceptably small. This method is summarised as follows:

1. For iteration n, calculate x_n, which is the midpoint of the current interval, $x_n = \frac{1}{2}(x_\alpha + x_\beta)$.
2. Calculate the function value at the midpoint, $f(x_n)$.
3. If convergence is satisfactory (that is, if the interval $\frac{1}{2}(x_\alpha - x_\beta)$ or the value $f(x_n)$ is sufficiently small), return x_n and stop iterating.
4. Check the sign of $f(x_n)$ and replace either x_α or x_β. Go to step 1 if not converged.

Example - The function $f(x) = -3x^3 + 7x^2 + 2x - 4$ has three distinct roots. We would like to find an approximation to the first positive root using the bisection method. We know that the first root lies between the points $x_\alpha = 0.5$ and $x_\beta = 1.5$ (this interval $[x_\alpha, x_\beta]$ is highlighted in fig. 16.1).

$$\begin{aligned} f(x_\alpha) &= -3x_\alpha^3 + 7x_\alpha^2 + 2x_\alpha - 4 \\ &= -3(0.5)^3 + 7(0.5)^2 + 2(0.5) - 4 \\ &= -1.625 \quad \text{(negative)} \end{aligned}$$

$$\begin{aligned} f(x_\beta) &= -3x_\beta^3 + 7x_\beta^2 + 2x_\beta - 4 \\ &= -3(1.5)^3 + 7(1.5)^2 + 2(1.5) - 4 \\ &= 4.625 \quad \text{(positive)} \end{aligned}$$

The first bisection point occurs at
$x_1 = \frac{1}{2}(x_\alpha + x_\beta) = \frac{1}{2}(0.5 + 1.5) = 1.$

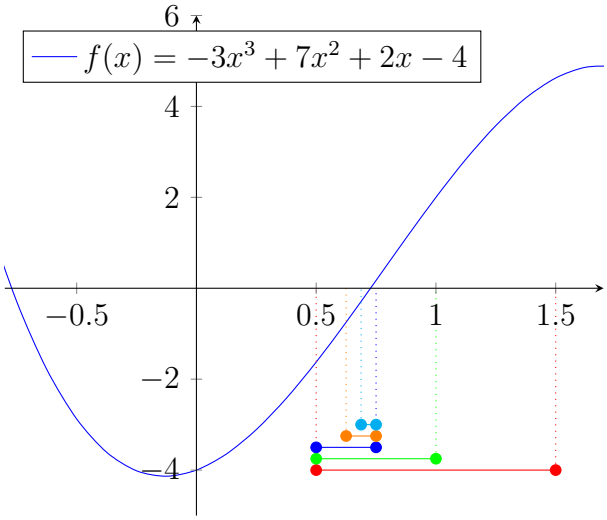


Figure 16.1: Shows successively smaller intervals resulting from each iteration of the bisection method.

$$\begin{aligned} f(x_1) &= -3x_1^3 + 7x_1^2 + 2x_1 - 4 \\ &= -3(1)^3 + 7(1)^2 + 2(1) - 4 \\ &= 2 \quad \text{(positive)} \end{aligned}$$

As the function is positive at x_1, we can exclude x_β and repeat this process in our new interval $[x_\alpha, x_1]$. The second bisection point occurs at $x_2 = \frac{1}{2}(x_\alpha + x_1) = \frac{1}{2}(0.5 + 1) = 0.75$.

$$f(x_2) = 0.172... \quad \text{(positive)}$$

As the function is positive at x_2, we can exclude x_1 and repeat this process in our new interval $[x_\alpha, x_2]$. The third bisection point occurs at $x_3 = \frac{1}{2}(x_\alpha + x_2) = \frac{1}{2}(0.5 + 0.75) = 0.625$.

$$f(x_3) = -0.748... \quad \text{(negative)}$$

As the function is negative at x_3, we can exclude x_α and repeat this process in our new interval $[x_3, x_2]$. The fourth bisection point occurs at $x_4 = \frac{1}{2}(x_2 + x_3) = \frac{1}{2}(0.625 + 0.75) = 0.6875$.

$$f(x_4) = -0.291... \quad \text{(negative)}$$

If we choose to terminate the iterations here, our approximation of $x_4 = 0.6875$ is still 5% lower than the correct value (found analytically), but this may be acceptably close for our application. We can also say with confidence that the root must be in our final range of $[x_4, x_2] = [0.6875, 0.75]$.

16.1.1 Potential Problems

The bisection method is conceptually simple, but is generally considered slow compared to other methods available. It is also very sensitive to the choice of the initial interval.

There are several cases that you should be aware of as they may cause you difficulty. The first is that if you have evaluated a function at two points and found their sign to be opposite, this does not guarantee that there is a root in this interval. Figure 16.2 shows an interval containing a discontinuity. If the bisection method was pursued it would locate the discontinuity as if it were a root.

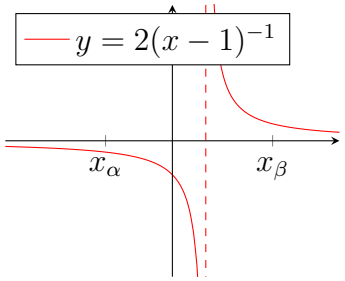


Figure 16.2: Function with discontinuity in the initial interval.

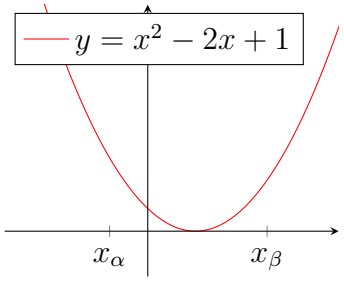


Figure 16.3: Function with coincident roots.

Another case to be aware of is if a function has multiple equal roots (*i.e.*, two roots at the same x-coordinate). The function $y = x^2 - 2x + 1$ is shown in fig. 16.3, which can be thought of as having a pair of coincident roots at $x = 1$ (factorise the expression if you don't see why!).

The final case that we will mention here is for functions that have multiple roots packed close together. Figure 16.4 shows the function $f(x) = x(x(16x-160)+529)-578$. If you did not spot that this was a cubic function, you may have presumed by looking at the graph that the interval $[x_\alpha, x_\beta]$ contained only one root. ¯_(ツ)_/¯

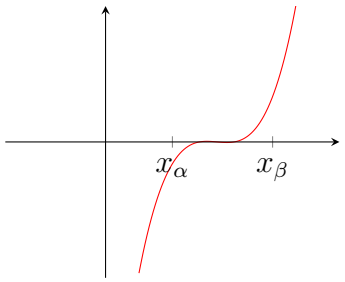


Figure 16.4: Cubic function with multiple close roots.

If you proceed with the bisection method from this starting interval, you will still end up finding a root, but you won't know which of the three you've found. It is now possible to find the remaining two roots, but it requires some careful thought. By taking the root you've just found as an approximation for, x_γ, and investigating the intervals either side of it (*i.e.*, the intervals $[x_\alpha, x_\gamma]$ and $[x_\gamma, x_\beta]$). Be aware that if x_γ was either the first or last of the three possible roots, then one of these two new intervals will lead you straight back to x_γ. Also, if you were lucky and managed to find the first root exactly (*i.e.*, $f(x_\gamma) = 0$), then clearly you cannot use this as one of your bounds as it is neither positive or negative, and will therefore have to use $x_\gamma + \delta$ instead, where δ represents a very small change in x_γ.

16.2 The Newton-Raphson Method

The Newton-Raphson (NR) method, sometimes just called Newton's method, named after Isaac Newton and Joseph Raphson, is an iterative method for approximating the roots of real-valued functions.

Starting from an initial guess for the root, x_0, the NR method requires the value of the function at this point, $f(x_0)$, as well as the function's local gradient, $f'(x_0)$. It uses these two pieces of information to construct a tangent line and then gives the x-intercept of this line as the next guess (this sounds complicated, but will become much clearer once you've seen a graph!).

To derive the NR formula, we need to be able to find the equation of a tangent. We know that all straight lines will have an equation of the form $y = mx + c$, where m is the gradient and c is the y-intercept. We also know that the gradient to our function at the point x_0 is $f'(x_0)$. So by substituting the relevant co-ordinates into our equation we get,

$$f(x_0) = f'(x_0)x_0 + c, \tag{16.1}$$

which can be rearranged to find c

$$c = f(x_0) - f'(x_0)x_0. \tag{16.2}$$

We can now write the following expression for the tangent line at x_0

$$y = f'(x_0)x + f(x_0) - f'(x_0)x_0, \tag{16.3}$$

Finally, by setting $y = 0$ and rearranging to make x the subject, we get

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}, \tag{16.4}$$

Now that we have our explicit equation for the intercept of the tangent line, we can rewrite this in the iterative notation that is the NR method.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{16.5}$$

A common use for the NR method is finding a numerical approximation for the n^{th} root of a number, as shown in the following example.

16.2.1 NR Example

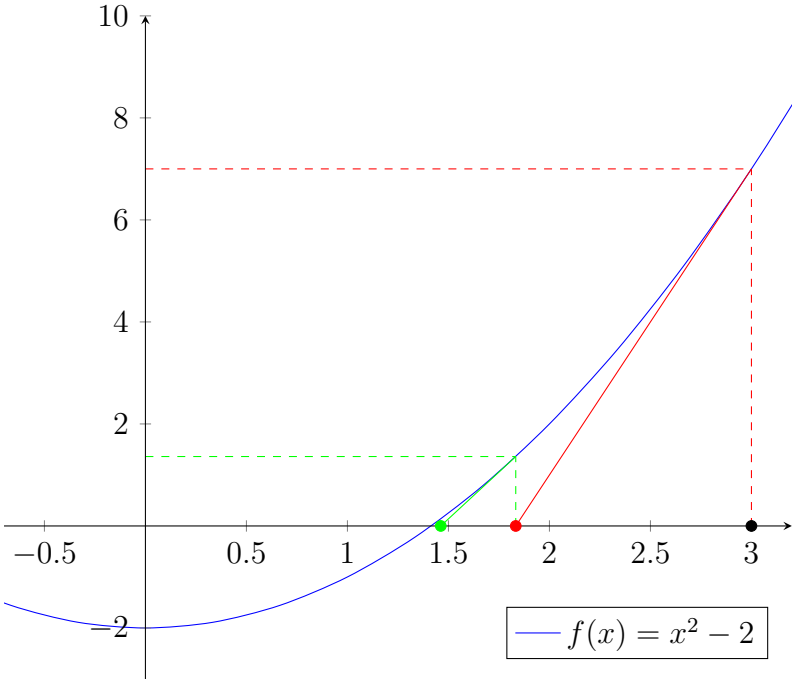
Find an approximation for the square root of 2 by using the NR method to find the root of the equation $f(x) = x^2 - 2$.

Differentiating our $f(x)$ shows the gradient function to be $f'(x) = 2x$. So we can now state the NR method in terms of our specific problem,

$$x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n} \quad ,$$

which rearranges to,

$$x_{n+1} = 0.5x_n + x_n^{-1} \quad .$$



If we take our starting guess to be $x_0 = 3$ (we could have made a better guess, but this was useful for illustration!), we can then write the following expression for our first iteration,

$$x_1 = 0.5(3) + \frac{1}{3} = \frac{11}{6} = 1.8\dot{3} \quad .$$

This gives us the improved approximation, $x_1 = 1.8\dot{3}$ (shown as the red dot on the graph). We can then iterate a second time using our new point,

$$x_2 = 0.5\left(\frac{11}{6}\right) + \frac{6}{11} = \frac{193}{132} = 1.46\dot{2}1\dot{1} \quad ,$$

which gets us to the green point, $x_2 = 1.46\dot{2}1\dot{1}$. A third time,

$$x_3 = 0.5\left(\frac{193}{132}\right) + \frac{132}{193} = \frac{72097}{50952} = 1.415... \quad ,$$

so even starting from a poor guess and with just three iterations we now have an estimate of the square root of 2 that is correct to within 0.1%.

However, if we had started from $x_0 = -3$ (or any negative number), we would have found the other root ($-\sqrt{2}$) instead.

Furthermore, if we have chosen our initial guess to be $x_0 = 0$, the NR method would not have yielded anything (try it!).

This illustrates that the selection of our initial guess is important. Typically, to find a specific root, we should aim to have no discontinuities or stationary points between the root and our initial guess. As we have just seen, in the case $f(x) = x^2 - 2$, to find the positive root, our initial guess had to be in the range $0 < x_0 < \infty$. For more complicated functions, selection of the initial guess requires careful consideration.

16.3 Secant method

The final root finding method that we'll be covering in this course is called the "secant method". You may remember form high school geometry that a secant is a line that intersects with a curve in at least two (distinct) places.

The key advantage of this approach is that although it makes use of gradients like the Newton-Raphson method, it does not require you to actually know the derivative of the function.

This method is summarised as follows:

1. Select two starting points, x_0 and x_1.
2. Use these two points to construct a secant.
3. Take the root of the secant as the next approximation, x_n.
4. Stop if $f(x_n)$ is close enough to zero.
5. Return to step 2 using the most recent two points, x_n and x_{n-1}.

As the two starting points have the coordinates $(x_0, f(x_0))$ and $(x_1, f(x_1))$, it's possible to construct a straight line of the form $y = mx + c$ that passes through both points.

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

Then, the root of this line (*i.e.*, where it crosses the horizontal axis) can be found by setting $y = 0$ and solving for x.

$$x = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)} = x_0 - f(x_0) \left(\frac{f(x_1) - f(x_0)}{x_1 - x_0} \right)^{-1}$$

Notice that in the second representation above, the secant method resembles the Newton-Raphson method, except that the gradient term has been replaced by a "rise over run" approximation of the gradient using the two points.

So, in general, for an iteration, was can write

$$x_{n+1} = \frac{x_{n-1}f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})} = x_n - f(x_n)\left(\frac{f(x_{n-1}) - f(x_n)}{x_{n-1} - x_n}\right)^{-1}$$

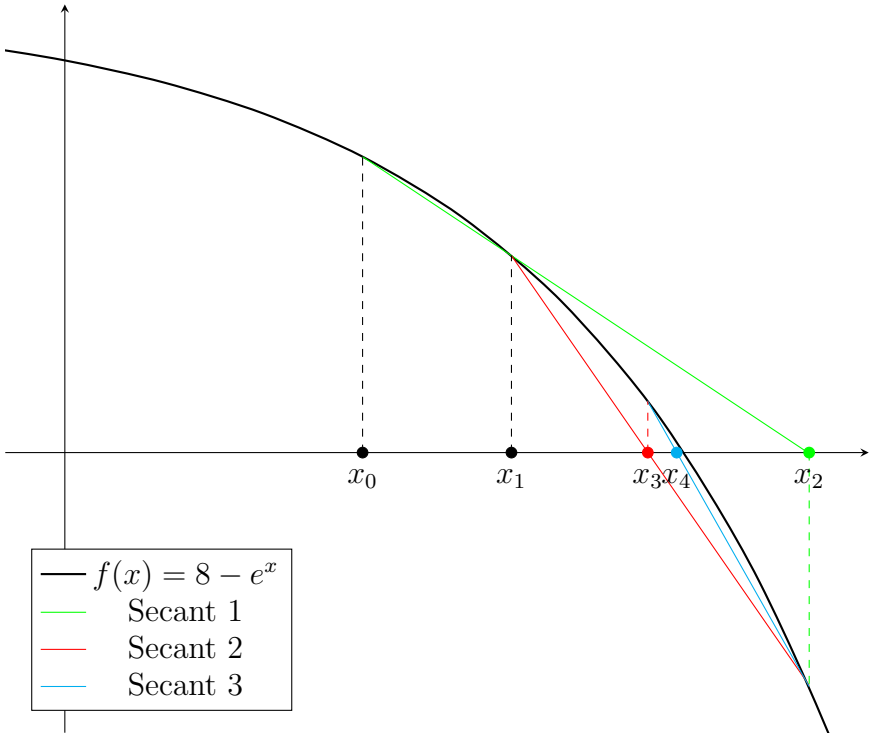
16.3.1 Example

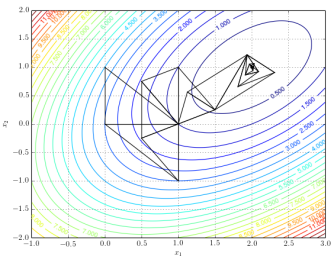
Consider the function $f(x) = 8 - e^x$, which has one real root.

Simply by substituting sequential values of x_n into the formula, we can build the following table, which converges to the root $x = 2.0794$ after five iterations (*i.e.*, it's the result of inputting the $x_4 = 2.0553$ and $x_5 = 2.0809$).

n	x_n	$f(x)$
0	1.0000	5.2817
1	1.5000	3.5183
2	2.4976	-4.1532
3	1.9575	0.9183
4	2.0553	0.1908
5	2.0809	-0.0121
6	2.0794	0.0001
7	2.0794	0.0000

The first three iterations of this process can be seen in the figure below. This example highlights the fact that the solution does not need to be within the domain specified by the first two guess points. However, it's important to also note that the secant method is not guaranteed to converge


x_0
x_1
x_3
x_4
x_2
$f(x) = 8 - e^x$
Secant 1
Secant 2
Secant 3



Chapter 17

Optimisation

Optimisation is a vast and varied topic within mathematics; it is also at the very core of engineering. Essentially, it can be thought of as the process of improving the performance of a system relative to a defined objective; or, if possible, finding the best performance.

Examples include:

- Planning the shortest path between points on a map.
- Using the minimum volume of material when manufacturing a product.
- Maximising the amount of energy produced by a wind farm.
- Finding the best parameters for recognising a face using a neural network.
- Writing the ideal headline to attract clicks on social media.
- Arranging shell companies to pay the least amount of tax.
- Choosing the optimal photos for success on a dating app.

Although this is a very diverse list of activities, notice the language "shortest", "minimum", "maximising", "best", "ideal", "least", "optimal"... they're all optimisation problems. For each one, the problem could be approached in a range of ways and a good optimisation strategy is one that finds the best results fastest.

The notion of "constraint" is very important in optimisation. Formula One racing is a great example for this as all the cars are trying to be the *fastest*, but they are constrained by rules about the allowable engine size, wheel diameter, spoiler designs etc.

The concept of a "tolerance" is also critical, where although you would like to be at the "optimum", it may be impossible to land exactly on it, so you have to decide what's *good enough* (think back to the root finding chapter).

Of course, you've interacted with simple optimisation for years already. For example, find the minimum of the simple 1D function $f(x) = ax^2 + bx + c$. To solve this, you'd just differentiate the function to $f'(x) = 2ax + b$ and set it to zero, yielding $x = -b/2a$. However, this was only possible because the function was so simple...

In many cases, the function will be multidimensional and highly complicated... maybe it's not something you can differentiate... or maybe you've not even got the function at all, but just a fluid flow simulation to design a plane that takes an hour to run each time you change the design... or a lab experiment to find a drug that takes a week to measure a single result...

Clearly, each of the problems listed above will require a wide variety of tools and new approaches are being developed all the time.

17.1 Linear regression

The word "linear" in linear regression is often mistakenly thought of as referring to fitting a straight line. In fact, linear regression describes a class of problems that are linear in the *coefficients*. So, the 1 dimension polynomial function

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

would be considered a linear regression problem, because although the function itself is not linear, the coefficients are linear. Similarly, the multidimensional function

$$g(x, y, z) = a_0 x + a_1 y + a_2 z$$

is also amenable to linear regression methods. And finally, the wild function

$$h(\mathbf{x}) = a_0 x_1 x_3^2 + a_1 \exp\left(-x_2^2\right) - a_2 \sqrt{x_1}$$

is both non-linear and multi-dimension, but it's still linear from the perspective of the coefficients and could therefore still be fit to some data using linear regression.

Unlike most of your calculus experience so far (where you differentiated functions with respect to their independent variables), for regression you will instead be differentiating with respect to the coefficients.

17.1.1 Fitting a straight line to some data

Let's start with a classic problem: Imagine you've acquired some data and you wish to find the "line of best fit". More formally, we've made n measurements of y (e.g. extension of a spring), at different values of x (e.g. force applied), and we would like to fit a line of the form $y = mx + c$ that minimises the distance between the data points and the line, but like any real data, there is some noise, so the data has some scatter.

The standard method for doing this is a least squares minimisation method, and it works as follows. We first define the "residual", r_i, for each data point (x_i, y_i);

$$r_i = y_i - (mx_i + c)$$

where r_i is the difference between the measured values of y_i and the value of the line at that point, which is just $(mx_i + c)$. However, crucially, we don't care whether the line is above or below the data, but just how far away it is, and a convenient way to "ignore" the sign is just to

square this value. Now we simply say that the best value of m and c will be those for which the **sum of the squares** of the residuals are a minimum; so we find

$$S = \sum_i r_i^2 = \sum_i (y_i - mx_i - c)^2$$

Now that we have expressed our problem mathematically, hopefully you can see that we can simply use partial differentiation to find the derivative of S with respect to each of m and c, and then set these derivatives to zero to find our minimum. We will then have 2 equations and 2 unknowns...

$$\frac{\partial S}{\partial m} = -2\sum x_i(y_i - mx_i - c) = 0$$
$$\frac{\partial S}{\partial c} = -2\sum (y_i - mx_i - c) = 0$$

Let's start by finding an expression for c. We can rearrange the expression for $\frac{\partial S}{\partial c}$ (think back to our series chapter to understand how) to give

$$cn + m\sum x_i = \sum y_i$$

and then since the average of all values of x_i is given by $\overline{x} = (\sum x_i)/n$ (and the same holds for $\overline{y}$), we can write

$$c = \overline{y} - m\overline{x}$$

We now need to build an expression for m. So, let's start by re-writing the expression for $\frac{\partial S}{\partial m}$ as

$$m\sum x_i^2 + c\sum x_i = \sum x_i y_i$$

and then we can substitute the equation we've just derived for c into the equation above, giving

$$m\sum x_i^2 + (\overline{y} - m\overline{x})\sum x_i = \sum x_i y_i$$

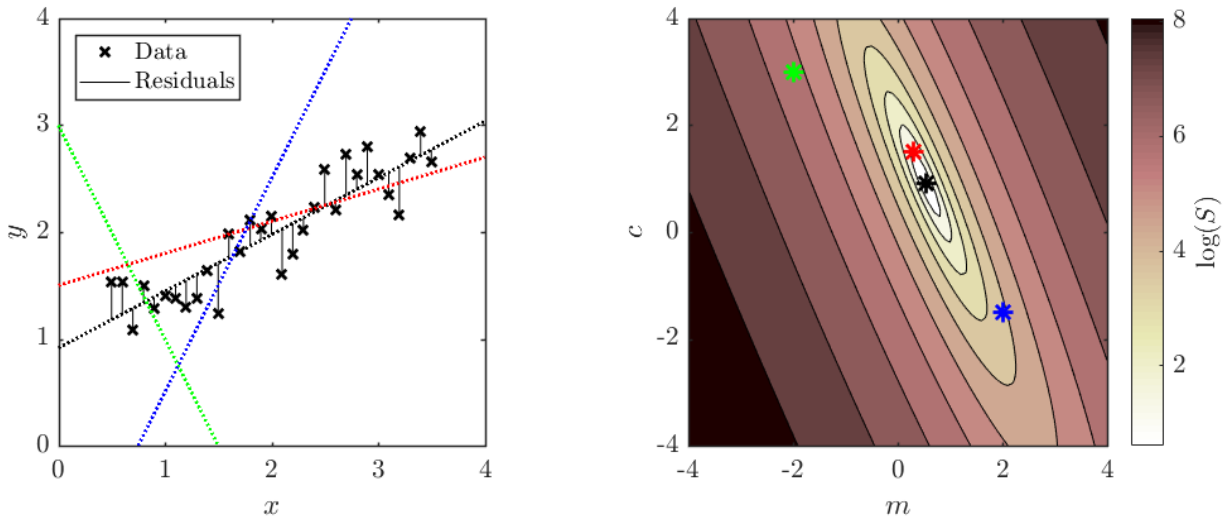
Collecting the m terms together and rearranging (remember, $(\sum x_i)/n = \overline{x}$),

$$m(\sum x_i^2 - \overline{x}^2 n) = \sum x_i y_i - \overline{x}\,\overline{y} n$$

then finally we can write an equation explicit for m and therefore also for c

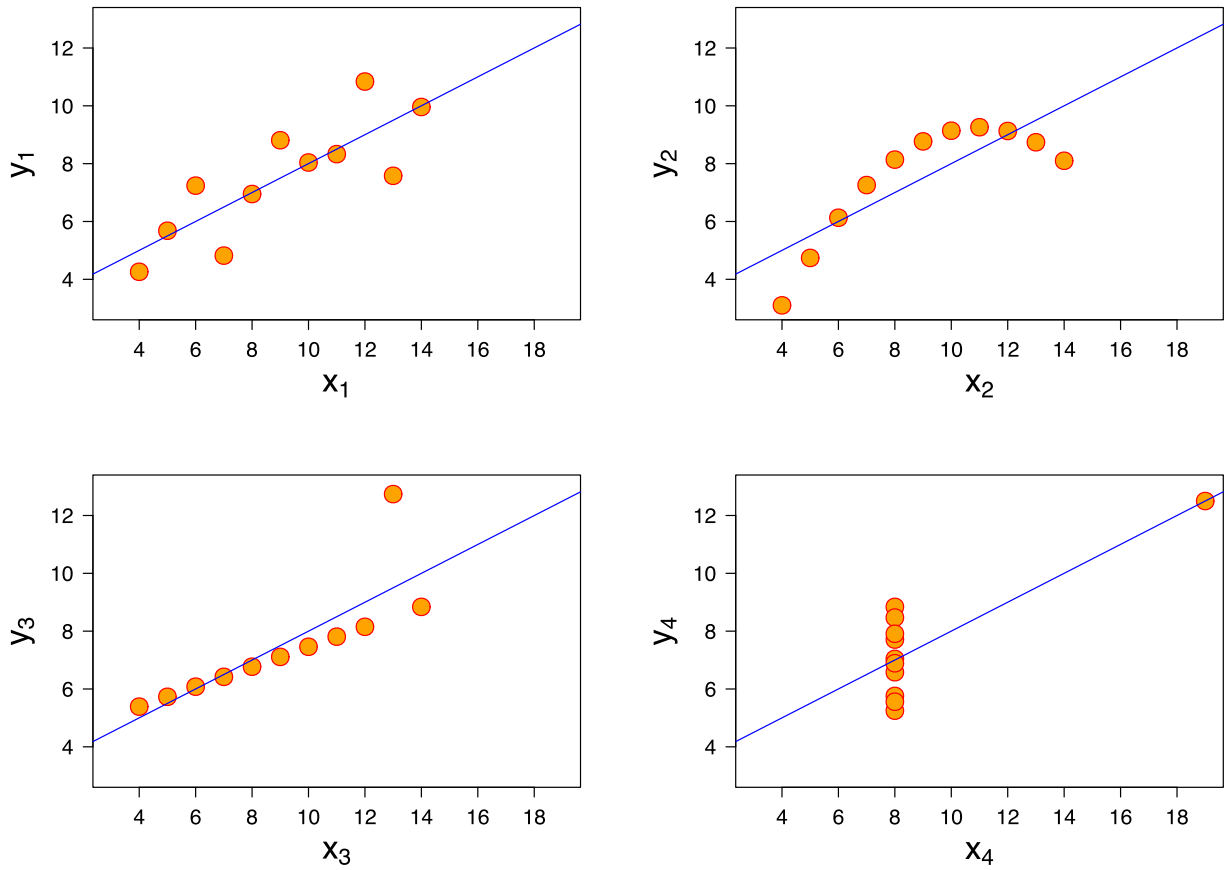
$$m = \frac{\sum x_i y_i - \overline{x}\,\overline{y} n}{\sum x_i^2 - \overline{x}^2 n} \equiv \frac{\overline{x}\,\overline{y} - \overline{xy}}{\overline{x}^2 - \overline{x^2}} \qquad \Rightarrow \qquad c = \overline{y} - \left(\frac{\overline{x}\,\overline{y} - \overline{xy}}{\overline{x}^2 - \overline{x^2}}\right)\overline{x}$$

And that's that. We now have explicit equations for m and c that allow us to find the line of best fit through an arbitrary dataset. Remember, these equations are really just finding the minima of the two partial derivatives, so this does not necessarily mean that S now equals zero (i.e. the line passes through all the points), but simply that S is the smallest it could be.



Each pair of m and c values corresponds to a value of the sum of squared residuals, S. We can plot a 2D map in (m, c)-space showing how S varies, where each point on this map corresponds to a different straight line on the data graph (matching colours), as illustrated in the figures above.

The fact that linear regression has an explicit formulation means that it has much more in common with the simple graph sketching activities



It's worth highlighting, before we finish this section, that just because you've minimised S, this does not mean that you have a "good fit". This is illustrated excellently by the above four plots, known as Anscombe's quartet. Four different data sets are shown which each have the same line of best fit AND the same value of S (and the same mean and standard deviation!).

17.2 Non-linear regression

A classic example of a non-linear regression problem is fitting a normal distribution curve to some data. Our function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where the mean, μ, standard deviation, σ, are the two parameters that we will be adjusting in order to achieve a good fit.

So, once again, we will start by writing an expression for the residuals

$$r_i = y_i - f(x_i) = y_i - \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right)$$

then our summed square error, S becomes

$$S = \sum (y_i - f(x_i))^2 = \sum \left(y_i - \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \right)^2$$

and our two monstrous derivatives are therefore

$$\frac{\partial S}{\partial \mu} = -\sum \left(\frac{(x_i - \mu)}{\sigma^3} \right) \left(\sqrt{\frac{2}{\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \left(y_i - \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \right) = 0$$

$$\frac{\partial S}{\partial \sigma} = -\sum \left(\frac{(x_i - \mu)^2 - \sigma^2}{\sigma^4} \right) \left(\sqrt{\frac{2}{\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \left(y_i - \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \right) = 0$$

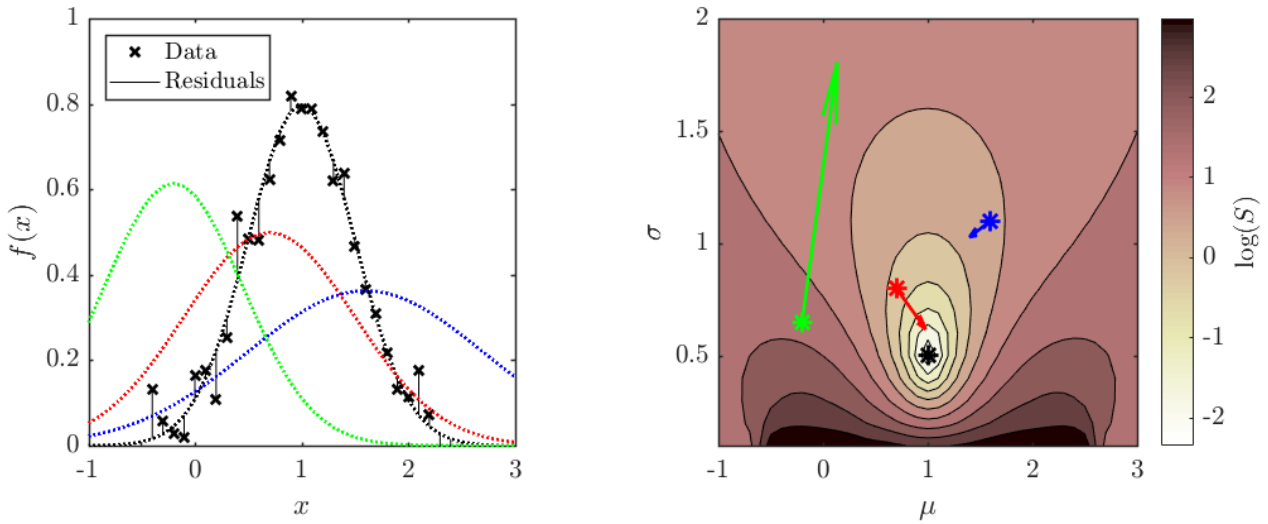
Hopefully, you've guessed that it will not be possible to rearrange the above two equations and make them explicit for μ and σ (as we did in the linear case). So what do we do?

Since we're lucky enough to even have our two derivative expressions that we really can evaluate, let's put them to use! Thinking back to our chapter on multivariate calculus, we can combine these two partial derivatives into a Jacobian vector

$$\mathrm{J}_S = \left[\frac{\partial S}{\partial \mu}, \frac{\partial S}{\partial \sigma} \right]$$

If we pick some arbitrary values of μ and σ, we can evaluate the Jacobian vector at this point. We can then think of J_S as a vector in (μ, σ)-space pointing in the direction in which S increases the most... So if we'd like to *minimise* S then our next guess for μ and σ should be in the opposite direction, i.e. $-\mathrm{J}_S$.

The left hand figure below shows some data points, as well as four normal distribution curves. The right hand figure shows a contour plot of S, parametrised by μ and σ, where the four points correspond by colour to the four adjacent curves. In addition, attached to each point is an arrow pointing in the direction $-\mathrm{J}_S$, where the length of the arrow shows the magnitude of the vector.



Essentially, we have just performed the first step of an iterative method: Starting from some arbitrary initial point, evaluate the Jacobian, move in the opposite direction to it to a new point and then evaluate the Jacobian again, and then repeat until some convergence criteria are met. It important to notice that the negative Jacobians shown in the contour plot do not point directly at the minimum (otherwise finding the solution would be easy!), but instead point in the steepest direction "downhill" locally. Also consider that although the black point doesn't appear to have a Jacobian arrow, this is simply because the local value of the partial derivative are tiny as it is close to the minimum.

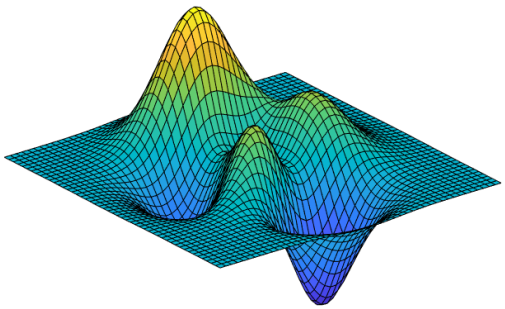
The difficulty comes in deciding *how far* to move in each step and when to stop. If you take big steps, you will move quickly down hill, but you might jump over minima and never "converge" to an acceptable solution; equally, if you take small steps, the system will converge very slowly as you make so little progress each time and it still might be a local minimum. To work out when to stop, you will need to choose some criteria, for example, picking a threshold value of S or the partial derivatives that you feel is acceptable, and then checking regularly while iterating. Also, picking a good starting point can be very important, as, in the case of the normal distribution

Imagine if our initial guess was $\mu = -100$ and $\sigma = 0.1$. This would mean that the region overlapping with the data is essentially a flat line at $f(x) \approx 0$, so the local gradients would be small as increasing or decreasing μ would only result is small changes in S. There are essentially two ways of dealing with this: The first is that you have *some* idea what your data's like and pick sensible values (i.e. correct order of magnitude at least!); the second approach is to run the fitting algorithm many times, start from a wide range of initial guesses.

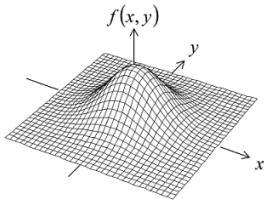
Finally, it's worth reminding you that although I show you the nice contour plot, to generate this image required me to calculate to value of S at each point. This is simple for "toy" problems like the two we've seen so far, but in general will not be possible for the kinds of optimisation problem engineers are typically faced with. Furthermore, you also might not have nice explicit expressions for the derivative and might therefore choose to use... a finite difference method to approximate them (and sometimes it is necessary to avoid gradients all together)!

17.3 Conclusion

We started this chapter off by saying that optimisation is a big topic, but even just from the two examples above you are already confronted with many of the key challenges. In both cases, I chose fitting functions with just two parameters so that I could show you nice 2D contour plots, but if there were three parameters it would already be difficult to plot and many real problems, for example training a neural networks, might require optimising thousands of parameters! So, once you've built your intuition on simple problems, and learnt strategies to deal with noisy data, local optima or complicated constraints you just have to trust the maths...



If you'd like to know more about this topic, head to our Mathematics for Machine Learning online specialisation on Coursera, which is free to Imperial students! :-)



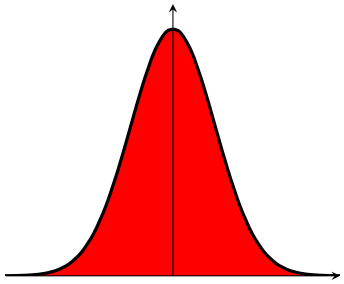
Chapter 18

The Normal Distribution

Gaussians, named after the mathematician Carl Friedrich Gauss, are functions based on the form where e is Euler's number and a, b and c are arbitrary positive constants.

$$f(x) = ae^{-b(x-c)^2} \quad ,$$

These functions have a characteristic "bell curve" shape and are asymptotic to the x-axis in both directions; however, crucially they also have a finite area. Finding the area under this curve is trick (just try integrating it from $x = -\infty$ to $x = +\infty$ directly if you're curious...), but we can find the answer by making use of a few clever substitutions (tricks!).



18.1 The Gaussian Integral

The following method (which you do not need to memorise, but should just be aware of) uses the following three tricks to find the integral:

Squared $\Rightarrow$ **Polar** $\Rightarrow$ **Substitution**

The meaning of each of the steps will hopefully become clear as we proceed. If you can remember these three words then you should be able to reproduce this derivation (I won't be asking you to) without too much difficulty!

We wish to find the area, A, under the standard Gaussian (*i.e.*, where a and b are 1 and c is 0).

$$A = \int_{-\infty}^{\infty} e^{-x^2}\, dx$$

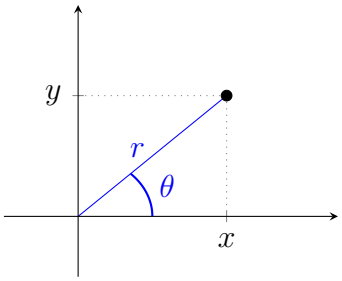
We can write an identical expression in terms of y that will have exactly the same answer (why is this useful? keep reading to find out!).

$$A = \int_{-\infty}^{\infty} e^{-y^2}\, dy$$

By multiplying these two functions together, we can form an expression for the area **squared**, A^2. As the two variables, x and y, are independent, the order of integration does not matter. Furthermore, we can rearrange the expression into a single exponent as shown.

$$\begin{aligned} A^2 &= \int_{-\infty}^{x=\infty} \int_{-\infty}^{y=\infty} e^{-y^2} e^{-x^2}\, dy\, dx \\ &= \int_{-\infty}^{x=\infty} \int_{-\infty}^{y=\infty} e^{-(x^2+y^2)}\, dy\, dx \end{aligned}$$

The variables x and y are usually, as in this case, used to describe a Cartesian coordinate system. The next step is to transform the system from Cartesian into **polar** coordinates, as illustrated in the adjacent figure.



$$\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta) \end{aligned}$$

To convert between coordinate systems, we calculate the determinant of something called the "Jacobian" matrix, which we'll be covering in more detail later in the course. This is because integration in (x, y)-space does not map directly on to (r, θ)-space (*i.e.*, $\mathrm{d}x\,\mathrm{d}y \neq \mathrm{d}r\,\mathrm{d}\theta$).

The Jacobian matrix is constructed by finding each of the four possible partial derivative combinations, as in the following table.

	$\mathrm{d}r$	$\mathrm{d}\theta$
$\mathrm{d}x$	$\frac{\mathrm{d}x}{\mathrm{d}r} = \cos(\theta)$	$\frac{\mathrm{d}x}{\mathrm{d}\theta} = -r\sin(\theta)$
$\mathrm{d}y$	$\frac{\mathrm{d}y}{\mathrm{d}r} = \sin(\theta)$	$\frac{\mathrm{d}x}{\mathrm{d}\theta} = r\cos(\theta)$

Once we have found the determinant of this matrix, we can then complete the conversion of our equation.

$$|J| = \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} = r\cos^2(\theta) + r\sin^2(\theta) = r(\cos^2(\theta) + \sin^2(\theta)) = r$$

Therefore,

$$\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$$

We can now transform our equation for A^2, which we'll do in three steps. First by using our Jacobian,

$$A^2 = \int_{-\infty}^{x=\infty} \int_{-\infty}^{y=\infty} e^{-(x^2+y^2)} r\,\mathrm{d}r\,\mathrm{d}\theta \quad .$$

Next we must transform our limits. To do this, think of $x - y$ space as a two dimensional plane. We have been asked to integrate between $-\infty$ and $+\infty$ in both directions, which can be thought of as "the entire plane". To cover the entire plane using polar coordinates, we simply need $r = 0$ to $r = +\infty$ and $\theta = 0$ to $\theta = 2\pi$.

$$A^2 = \int_{0}^{r=\infty} \int_{0}^{\theta=2\pi} e^{-(x^2+y^2)} r\,\mathrm{d}r\,\mathrm{d}\theta$$

Finally, we also know from Pythagoras that $x^2 + y^2 = r^2$, which leads to

$$A^2 = \int_0^{r=\infty} \int_0^{\theta=2\pi} e^{-r^2} r \, \mathrm{d}r \, \mathrm{d}\theta$$

In this form, we notice that the integrand (*i.e.*, the function to be integrated) does not contain θ, so we can already evaluate the θ integral, yielding

$$A^2 = 2\pi \int_0^{r=\infty} e^{-r^2} r \, \mathrm{d}r$$

The last stage is to make the **substitution** $s = -r^2$, which differentiates to

$$\mathrm{d}s = -2r \, \mathrm{d}r$$

Take care to substitute the limits correctly!

$$A^2 = -\pi \int_0^{s=-\infty} e^s \, \mathrm{d}s = -\pi \left[e^{-\infty} - e^0\right] = -\pi \left[0 - 1\right] = \pi$$

Having completed this integration, we simply take the square root of our answer to find the area, A, under our standard Gaussian... easy as $A = \sqrt{\pi}$.

18.2 The Normal Distribution

The Gaussian is also the correct shape for modelling random variables, a reasonable example of which might be the height of students in your class. The curve itself is the probability density function (PDF), so the value of the curve at a point is the *probability density* NOT the probability! It is only by finding the area under the curve between two x values that allows us to calculate a probability. For example, when we ask "how many people are 1.70 m tall?" we do not mean how many are *exactly* 1.700000... m tall, but more likely, how many are between 1.695 m and 1.705 m tall.

We have just found the area under the curve $y = e^{-x^2}$ to be $\sqrt{\pi}$, which roughly equals 1.772, but the total area under any probability curve should be 1. To understand why, consider that finding the total area under a PDF of "student height" is like asking "What is the probability that a random student in your class is *any* height?"... you can be 100% sure that they they have *a* height, so the area must be 1. In order to modify our function such that its total area is 1, we simply divide by its current area.

$$f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2} \tag{18.1}$$

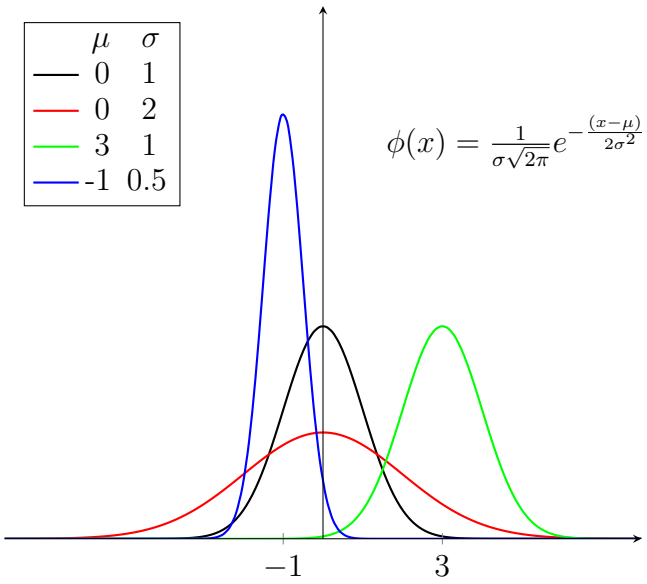
The second modification we are going to make is that we would like both the *standard deviation*, σ, and therefore also the *variance*, σ^2, to be equal to 1. The variance of eq. 18.1, which is a measure of the broadness of the bell curve, is currently equal to 0.5. Although we won't go through the derivation here, this modification simply requires dividing x^2 and the function itself both by a factor of $\sqrt{2}$, which gives what is usually referred to as the *standard normal distribution*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

We can then generalise this probability density function to its final form, which is called the *general normal distribution*, ϕ. This allows us to modify the *mean*, μ, and the standard deviation, whilst ensuring that the the total area underneath is always equal to 1, as illustrated in the figure below.

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The normal distribution is not "normal" in the sense that it is "usual" or "average", which can lead to to some confusion. One way to keep this in mind is to remember that we have just "normalised" our function by modifying the exponent and dividing it by $\sqrt{2\pi}$ (although this is not the real reason Gauss chose the name).



There any many alternative distributions that we can use to model data, but you will encounter the normal distribution frequently and need to know how to manipulate it.

Going back to our example of modelling the heights of students in a class, we now would like to be able to use the function to make predictions. Once we've got our "fitted" curve (which we call a "model"), we can use it to evaluate the probability of a randomly selected student being between two heights (*e.g.* "what is the chance a student is between 1.5 m and 1.6 m tall?").

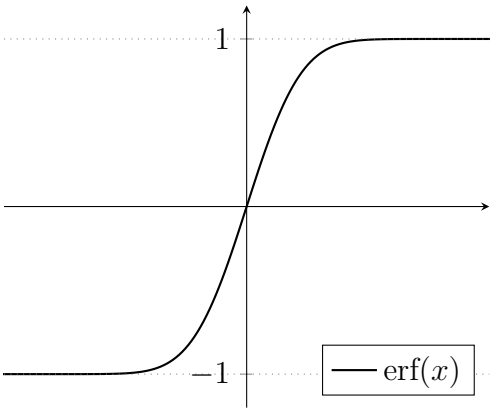
To evaluate this probability, P, we need to be able find the area under our function between two x values, x_a and x_b.

$$P(x_a < X < x_b) = \int_{x_a}^{x_b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

Although we were able to integrate the function between infinite limits, the tricks we used do not work on finite limits (try it!). You may have been taught in school to use something called a "Z-table" which contains many values of this integral so that you can look up the one you need; however, in your careers it is very unlikely that you will use this outdated approach. The next section explains how we can evaluate this integral in a way you can implement in code.

18.2.1 The error function

The Gauss error function, written $\operatorname{erf}(x)$, is a special function for evaluating the integrals of Gaussian functions. The function e^{-x^2} is *even* (*i.e.*, symmetrical about the y-axis), so the two definitions given below are equivalent, as you could either find the area under the region from $-x$ to x or just find it from 0 to x and double it.

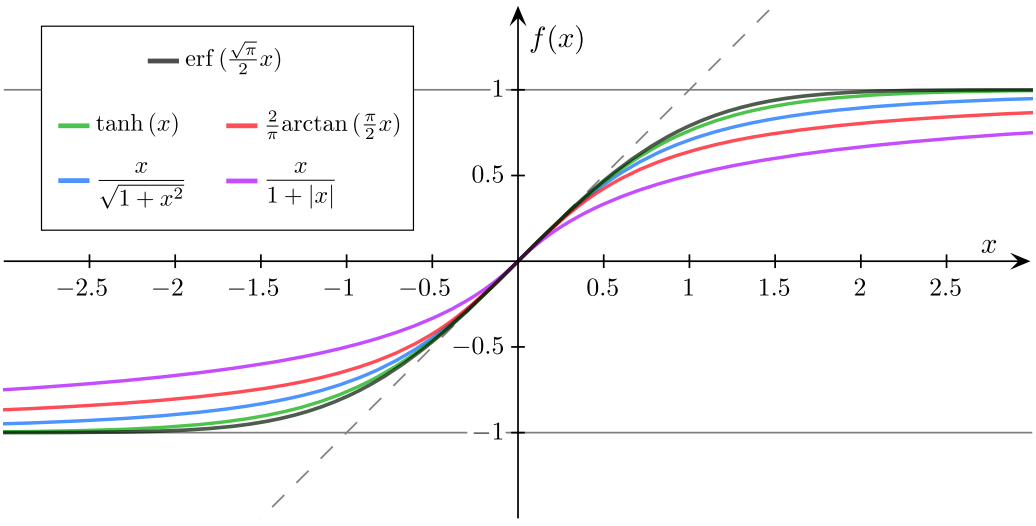


$$\begin{aligned}\operatorname{erf}(x) &= \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2}\ \mathrm{d}t \\ &= \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2}\ \mathrm{d}t\end{aligned}$$

Also, notice that the variable t is what we call a "dummy" variable, as it does not appear outside the expression and is only there to allow us to manipulate x in a certain way.

We still do not know how to evaluate $\operatorname{erf}(x)$ directly. In fact, ask yourself this: when you ask you calculator to evaluate $\sin(7)$, how does it actually do this? The reality is, there is no explicit, simple formula to exactly evaluate these functions, so various clever approximations have been developed, using approaches like the Taylor series, which we shall see later in the course. Fortunately for you, most calculators have a $\sin(x)$ evaluation button... unfortunately for you, most calculators do not have an $\operatorname{erf}(x)$ button. For the rest of your career, you'll have access to a computer with the internet that can hepl you evaluate erf; however, the fact that you will only have a calculator in the exam give us an opportunity to practice another engineering skill: approximations.

In the following figure you can see a selection of "sigmoid" (*i.e.*, "S" shaped) functions, that resemble the error function, in that they are rotationally symmetrical around the origin and have a range from -1 to 1 (N.B. Sigmoid functions are often used in neural networks as "activation" functions). The hyperbolic tangent function ("tanh") is not only a close approximation, but can also be evaluated on your calculator. We will often use the approximation $\operatorname{erf}(x) \approx \tanh(1.2x)$.



Now that we can evaluate $\text{erf}(x)$ we can use it in the following expression (called a *cumulative distribution function* or CDF) to calculate probabilities from the general normal distribution function. To find the probability being below a certain x value, we use the following expression (illustrated in fig. 18.1).

$$P(X < x_b) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x_b - \mu}{\sigma\sqrt{2}}\right)\right]$$

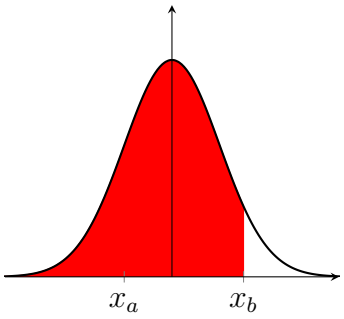


Figure 18.1: Area under a Gaussian up to x_b.

Similarly, to find the probability of something above a certain x value, we can then simply find one minus the CFD up to that point.

$$\begin{aligned} P(X > x_a) &= 1 - P(X < x_a) \\ &= 1 - \frac{1}{2}\left[1 + \text{erf}\left(\frac{x_a - \mu}{\sigma\sqrt{2}}\right)\right] \\ &= \frac{1}{2}\left[1 - \text{erf}\left(\frac{x_a - \mu}{\sigma\sqrt{2}}\right)\right] \end{aligned}$$

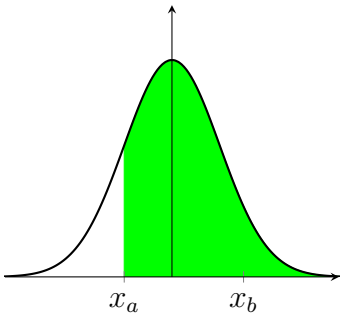


Figure 18.2: Area under a Gaussian beyond x_b.

To find the probability of something being between two specified bounds, we can then simply find the difference between two of these CDFs, as shown in fig. 18.3.

$$\begin{aligned} P(x_a < X < x_b) &= P(X < x_b) - P(X < x_a) \\ &= \frac{1}{2}\left[\text{erf}\left(\frac{x_b - \mu}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{x_a - \mu}{\sigma\sqrt{2}}\right)\right] \end{aligned}$$

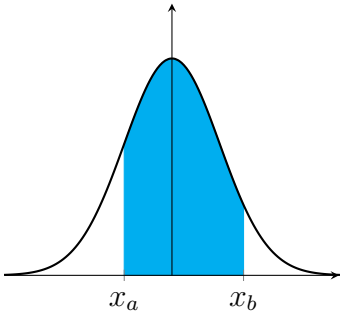


Figure 18.3: Area under a Gaussian from x_a to x_b.

The last case is when we want to find the probability of being outside a certain range, where we simply evaluate one minus the CDF of the range.

$$\begin{aligned} P(X < x_a) + P(X > x_b) &= 1 - P(x_a < X < x_a) \\ &= 1 - \frac{1}{2}\left[\text{erf}\left(\frac{x_b - \mu}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{x_a - \mu}{\sigma\sqrt{2}}\right)\right] \end{aligned}$$

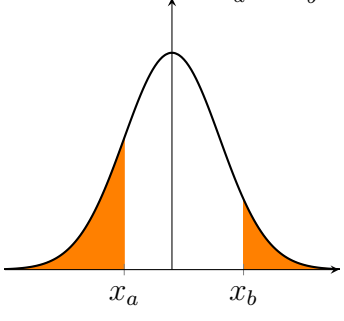


Figure 18.4: Area under a Gaussian outside x_a to x_b.

Example - In a class of 100 students, if the average height is 1.70 m and the standard deviation is 20 cm, how many students would you expect to be between 1.25 m and 1.45 m tall?

$$\begin{aligned} P(1.25 < X < 1.45) &= P(X < 1.45) - P(X < 1.25) \\ &= \frac{1}{2}\left[\operatorname{erf}\left(\frac{1.45 - 1.7}{0.2\sqrt{2}}\right) - \operatorname{erf}\left(\frac{1.25 - 1.7}{0.2\sqrt{2}}\right)\right] \\ &= \frac{1}{2}\left[(-0.7887...) - (-0.9756...)\right] \\ &= 0.093... \end{aligned}$$

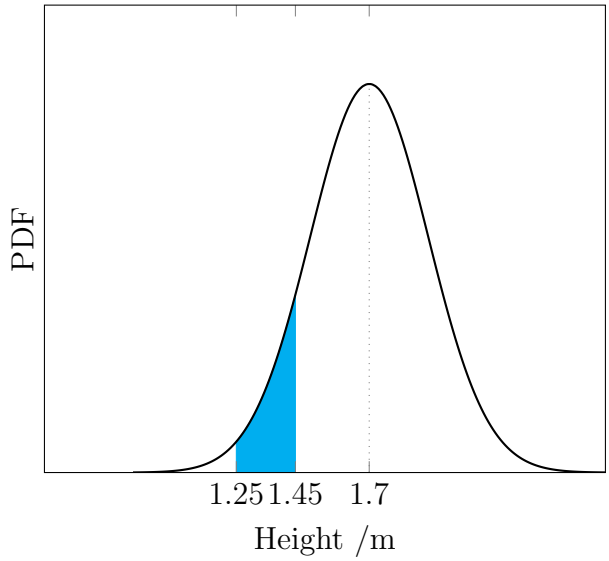


Figure 18.5: PDF of the heights of students in a class.

As the probability of any given student falling within the range is roughly 0.093, we would expect for there to be approximately 9 students between 1.25 m and 1.45 m in a class of 100.

The values of the error function were found using www.wolframalpha.com, which is a free online knowledge engine, although any other platform, such as MATLAB®, would also have a suitable function.